The Journal of Economic Perspectives

A journal of the American Economic Association

Spring 2018

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by LSC Communications, Owensville, Missouri, 65066, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Spring 2018, Vol. 32, No. 2. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. E-reader versions are free. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 32 • Number 2 • Spring 2018

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 3–24

The US Gains From Trade: Valuation Using the Demand for Foreign Factor Services

Arnaud Costinot and Andrés Rodríguez-Clare

About 8 cents out of every dollar spent in the United States—including both spending by consumers on final goods and spending by firms on intermediate goods—is spent on imports, according to the World Input–Output Database (WIOD). What if, because of a wall or some other extreme policy intervention, these goods were to remain on the other side of the US border? How much would US consumers be willing to pay to prevent this hypothetical policy change from taking place? The answer to this question represents the welfare cost from autarky or, equivalently, the welfare gains from trade.

There is little direct empirical evidence about the impact of autarky on prices and quantities. The Jeffersonian trade embargo at the beginning of the nineteenth century is one rare exception (Irwin 2005). It is not clear, however, what such historical evidence can tell us about the magnitude of US gains from trade today. In order to make progress on this important issue, we therefore propose an alternative strategy combining both theory and empirics.

Our strategy is based on two observations. First, when countries exchange goods, it is as if they were indirectly exchanging the factor services embodied in the production of these goods: unskilled labor, skilled labor, physical capital, land, and

■ *Arnaud Costinot is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Andrés Rodríguez-Clare is Professor of Economics, University of California at Berkeley, Berkeley, California. Their emails are costinot@mit.edu and andres@econ.berkeley.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.3 doi=10.1257/jep.32.2.3

so on. Second, when a country is under autarky, it is as if the prices of all foreign factor services were at least as high as their reservation values.

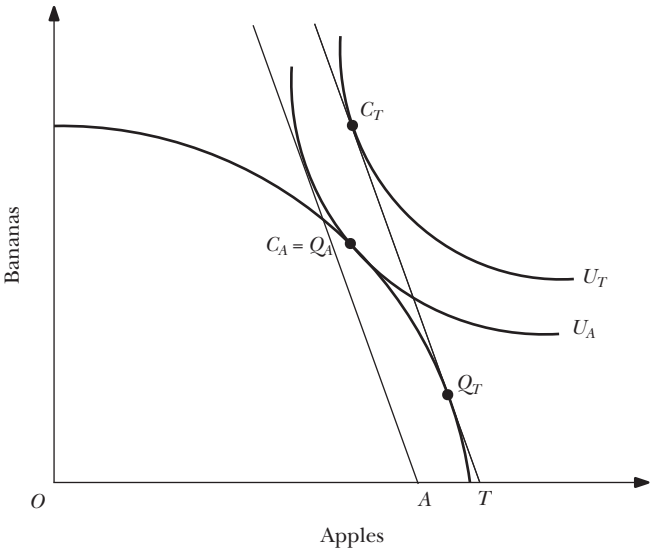
These observations suggest a parallel between the valuation of the welfare gains from trade and the "new good" problem in the field of industrial organization. When industrial organization economists want to evaluate the welfare gains from the introduction of a new product, from Apple Cinnamon Cheerios to the minivan, they estimate the demand for such products, determine the reservation price at which demand would be zero, and then measure consumer surplus by looking at the area under the (compensated) demand curve between the price at which these new products are currently being sold and their reservation price (for example, Hausman 1997; Petrin 2002; Nevo 2003). Trade economists can follow a similar strategy to measure the welfare gains from trade. In this approach, foreign factor services are just like new products that appear when trade is free but disappear under autarky.

This theoretical connection, in turn, points towards two key empirical considerations for the valuation of the US gains from trade: 1) How large are the US imports of factor services? 2) How elastic is the demand for these imported factor services? If consumers do not spend much on a new product or if this product is a close substitute to other existing products, the welfare gains to consumers from its introduction are likely to be small. Likewise, if the United States does not import much from the rest of the world or if the factor services that it imports are close substitutes to those that it would have access to under autarky, the US gains from trade are unlikely to be large.

The rest of this article is organized as follows. In the next section, we start from the textbook treatment of the welfare gains from trade, as described in Bhagwati, Panagariya, and Srinivasan (1998). Following the work of Adao, Costinot, and Donaldson (2017), we explain the advantages of measuring the gains from trade by focusing on the international exchange of factor services rather than on the specific goods and services that are imported and exported. This approach also provides an intuitive perspective on the welfare formula for the gains from trade derived in Arkolakis, Costinot, and Rodríguez-Clare (2012). Throughout this section, we restrict ourselves to a static economy with a representative agent and without distortions. This reflects both the emphasis in the existing literature and our view that this benchmark environment has worthwhile lessons to teach. We then turn to measurement, with a focus on the US economy. We describe the level of the US demand for foreign factor services and discuss the estimation of the elasticity of the US demand for foreign factor services. For a large and fairly closed economy like the United States, our analysis points towards welfare gains from trade ranging from 2 to 8 percent of GDP.

We conclude by discussing three issues set aside in our benchmark analysis: growth, distortions, and redistribution. In theory, their introduction may either increase or decrease the magnitude of the gains from trade. In practice, existing estimates incorporating such considerations do not suggest a significant and systematic bias in our benchmark estimates.

Figure 1
A First Look at the Welfare Gains from Trade



Source: Authors.
Note: In the autarky equilibrium consumption, C_A, is equal to output, Q_A. In the trade equilibrium, production of apples for export increases, and US firms produce at Q_T. After exporting apples and importing bananas, US households consume C_T. To measure the gains from trade, compare the level of income needed to achieve the post-trade utility on the higher indifference curve with the level of income (at the same relative prices) needed to achieve the utility from the autarky equilibrium. Graphically, the US gains from trade (GT) are therefore given by $GT = 1 - OA/OT$.

Gains From Trade in Theory

The Textbook Approach

As a warm-up, suppose that the United States only exports apples—the fruit, not the computer—in exchange for bananas. Since realism is clearly not the main objective at the moment, suppose further that we know everything that there is to know about the US technology and the tastes of US consumers. We have summarized this information into a production possibility frontier and a series of indifference curves in Figure 1.

In this economy, measuring the gains from trade is a fairly pedestrian affair. In the autarky equilibrium, consumption, C_A, is equal to output, Q_A. In the trade equilibrium, production of apples for export increases, and US firms produce at Q_T. After exporting apples and importing bananas, US households consume C_T. To measure the gains from trade, compare the level of income needed to achieve the post-trade utility on the higher indifference curve with the level of income (at the same relative prices) needed to achieve the utility from the autarky

equilibrium. Graphically, the US gains from trade (GT) are therefore given by $GT = 1 - OA/OT$.[1]

The actual US pattern of trade is a tad more complex. In 2005, for instance, Baldwin and Harrigan (2011) report that the United States had positive exports in about 8,500 of the product categories in the Harmonized System 10-digit classifications, ranging from "new motor vehicle engine between 1500–3000 cc, more than 6 cylinders" to "bicycles with wheels greater than 63.25 cm diameter." Product differentiation, of course, does not stop there. Within each product category, a large number of firms may themselves be exporting differentiated varieties of these products. For such an economy, how can we measure $1 - OA/OT$?

One potential strategy would be to scale up the textbook approach. Namely, one could start by estimating production sets and indifference curves for all these differentiated products around the world. With those in hand, one could then compute any counterfactual equilibrium, including the one where we send the US economy back to autarky, as well as the welfare cost associated with moving to such an equilibrium. However, the amount of actual information required to implement the textbook approach is, to put it mildly, nontrivial. On the demand side, for instance, the goal is not merely to obtain information about own-price and cross-price elasticities within any given industry. For this approach, we would need to estimate all own-price and cross-price elasticities for all goods around the world. This means estimating the cross-price elasticity between US smart phones and French red wine, Japanese hybrid cars and Costa Rican coffee, and all other possible combinations.

From Trade in Goods to Trade in Factor Services

Recently, Adao, Costinot, and Donaldson (2017) proposed an approach to reduce the dimensionality of what is required for counterfactual analysis in general, and the measurement of the gains from trade in particular.

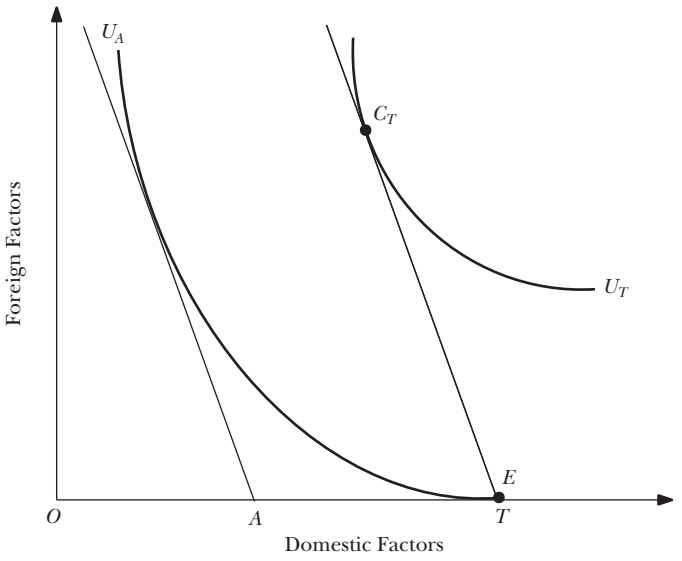
The starting point of their analysis is the equivalence between neoclassical economies and what are called "reduced exchange economies" in which countries simply trade factor services. They show that for any competitive equilibrium of a general neoclassical economy with arbitrary preferences, technologies, and sets of goods and factors, there exists an equilibrium in a reduced exchange economy that is equivalent in terms of welfare, factor prices, and the factor content of trade. Preference and technological considerations in the original neoclassical economy simply map into preferences over factor services in the reduced exchange economy. For the purposes of measuring the gains from trade, this approach implies that

[1] In formal terms, the US gains from trade correspond to the absolute value of the equivalent variation between the two equilibria. Expressed as a percentage of US initial GDP, we get

$$GT = 1 - \frac{e(p_T, U_A)}{e(p_T, U_T)},$$

where $e(p_T, U_T)$ represents the expenditure required to achieve the utility level in the trade equilibrium, which is also equal to US initial GDP, and $e(p_T, U_A)$ represents the expenditure required to achieve the autarky utility level, U_A, at the trade equilibrium prices, p_T.

Figure 2
Another Look at the Welfare Gains from Trade



Source: Authors.
Note: While the US economy is endowed with domestic factors, it has, by definition, no foreign factors. In the trade equilibrium, the former can be exchanged for the latter, leading to consumption C_T. But under autarky, consumption would have to be equal to the US endowment of domestic factors, as described by point *E*. Just like in the Figure 1, we can measure the gains from trade as $GT = 1 - OA/OT$.

instead of estimating production and demand functions around the world, we only need to estimate the reduced demand for factors, that is, the demand for factor services embodied in the goods purchased from countries around the world.[2]

Figure 2 describes this basic strategy. While the US economy is endowed with domestic factors, it has, by definition, no foreign factors. In the trade equilibrium, the former can be exchanged for the latter, leading to consumption C_T. But under autarky, consumption would have to be equal to the US endowment of domestic factors, as described by point *E*. Just like in the textbook case, we can measure the gains from trade as $GT = 1 - OA/OT$. But here, we no longer need to worry about estimating production and demand functions for goods. The gains from trade only depend on the shape of the indifference curve over domestic and foreign factor services.

[2]From an empirical standpoint, the fewer factors there are, the easier the estimation of the reduced demand for factors is. It should be clear, however, that the approach of Adao, Costinot, and Donaldson (2017) does not hinge on the assumption, common in the trade literature, that there are more goods than factors. The critical observation to reduce the dimensionality of what needs to be estimated is that knowledge of the reduced demand for factors is sufficient to measure the gains from trade; separate knowledge of demand for goods, by consumers and firms, and demand for factors, by firms, is not required.

In this situation, we can compute the gains from trade in the same way as one would compute the welfare gains from the introduction of new products. Figure 3 illustrates the approach. Under autarky, the expenditure share on foreign factors would be zero and the relative price of foreign factors would be equal to their reservation value, p_A. With trade, the expenditure share on foreign factors, λ_F, would be strictly positive and the relative price of foreign factors, p_T, would be strictly below its reservation value. We can compute (the log of) the difference between the income level required to achieve the autarky utility level, U_A, at the autarky and trade prices by integrating below the (compensated) expenditure share on foreign services between (the log of) p_A and p_T. After simple manipulations, this leads to the following general formula for the welfare gains from trade,

$$GT = 1 - \exp(-\mathcal{A}),$$

where $\mathcal{A}$ denotes the gray area in Figure 3.[3]

The ACR Formula in Perspective

The previous discussion offers an intuitive way to understand the welfare formula we derived in Arkolakis, Costinot, and Rodríguez-Clare (2012), which we will refer to as the ACR formula. In a class of commonly used trade models, including the models developed by Anderson (1979), Krugman (1980), Eaton and Kortum (2002), and various versions of Melitz (2003), welfare gains from trade can be expressed as a function of two sufficient statistics,

$$GT = 1 - \lambda_D^{1/\varepsilon},$$

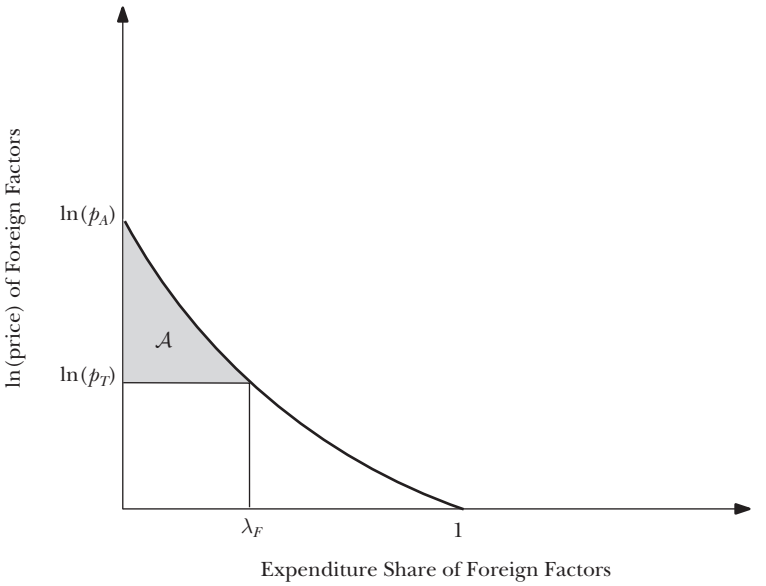
where $\lambda_D = 1 - \lambda_F$ denotes the share of expenditure on domestic goods in the trade equilibrium and $\varepsilon \geq 0$ denotes the trade elasticity.

The economics is straightforward. In spite of their different micro-theoretical foundations, the previous models all generate a similar demand for foreign factor services. Namely, in all these models, the share of expenditure on foreign factors, both compensated and uncompensated, takes a constant elasticity of substitution (CES) form. Depending on the specifics of the model, the trade elasticity may have a different structural interpretation. In Anderson (1979) and Krugman (1980), ε corresponds to the elasticity of substitution between domestic and foreign goods. In Eaton and Kortum (2002) and Melitz (2003), ε instead measures the dispersion of productivity across goods within a country.[4] In all models, however, domestic and

[3] For a small change in the log of the price of foreign factor services, $d \ln p$, Shepard's Lemma implies $d \ln e = \lambda_F d \ln p$. Using US factor services as our numeraire and integrating, we therefore obtain $\ln e(p_A, U_A) - \ln e(p_T, U_A) = \mathcal{A}$. Under this choice of numeraire, US factor income and US expenditure must also be unchanged between the autarky and trade equilibria: $e(p_A, U_A) = e(p_T, U_T)$. Starting from the equation in footnote 1, the two previous observations lead to equation in the text.

[4] Both Krugman (1980) and Melitz (2003) feature monopolistic rather than perfect competition. In general, this may lead to inefficiencies not captured in the analysis of Adao, Costinot, and Donaldson

Figure 3
Integrating Below the Factor Demand Curve



Source: Authors.
Note: To reduce the dimensionality of the problem of counterfactual analysis, we start from the equivalence between neoclassical economies and what are called "reduced exchange economies" in which countries simply trade factor services. This approach implies that by estimating the total demand for foreign factor services and examining the area below that demand curve, we can measure the overall gains from trade in the same way as one would compute the welfare gains from the introduction of new products. Under autarky, the expenditure share on foreign factors would be zero and the relative price of foreign factors would be equal to their reservation value, p_A. With trade, the expenditure share on foreign factors, λ_F, would be strictly positive and the relative price of foreign factors, p_T, would be strictly below its reservation value. (See text for further details.) The gains from trade are given by $GT = 1 - \exp(-\mathcal{A})$, where $\mathcal{A}$ is the gray area in the figure.

foreign factor services remain imperfect substitutes, with ε being the elasticity of substitution between them. Together with how much we trade, this is all we need to know to compute the gains from trade.[5]

To be clear, this approach is related to, but distinct from, the work of Feenstra (1994) and Broda and Weinstein (2006) on the gains from trade as a result of new

(2017). The assumption of constant elasticity of substitution preferences over goods in Krugman (1980) and Melitz (2003) rules that out.

[5] In the constant elasticity of substitution case, integrating below the demand for foreign factor services between ln(p_T) and infinity, which is the price of foreign factors under autarky, one can check that $\mathcal{A} = \ln(1 + p_T^{-\varepsilon})/\varepsilon$. To go from this expression to the ACR formula, one can then use the fact that the CES demand system is invertible. Thus, one can infer the relative price of foreign factors in the trade equilibrium, p_T, from the share of expenditure on those, λ_F. Specifically, given $\lambda_F = \frac{p_T^{-\varepsilon}}{1+p_T^{-\varepsilon}}$, we get $1 + p_T^{-\varepsilon} = 1/(1 - \lambda_F) = 1/\lambda_D$ and, in turn, $\mathcal{A} = -(\ln \lambda_D)/\varepsilon$. Combining the previous expression with the earlier equation showing gains from trade in Figure 3, we get the ACR formula in the text.

varieties becoming available, under the assumption of CES preferences for varieties from different countries. Here, we emphasize that by estimating the total demand for foreign factor services and examining the area below that demand curve, we can measure the overall gains from trade.[6]

Can It Be That Simple?

Although the formula just derived for welfare gains from trade is fairly general, the overwhelming number of applications of these concepts by trade economists involve a constant elasticity of substitution factor demand system, as illustrated by the ACR formula. Clearly, constant elasticity of substitution is a very strong functional-form restriction. There is no a priori reason to believe that the same assumption that has become popular in the trade literature to date—where popularity is often determined by tractability—should be the best guide to estimate the gains from trade in the United States in practice. In general, we do not expect "the" trade elasticity to be unique: for example, import demand may be much less elastic around autarky than around free trade, or much different in Costa Rica and France than in the United States, a scenario that the constant elasticity of substitution assumption necessarily rules out.

Rather than focus on a specific formula for the welfare gains from trade, the basic insight from Arkolakis, Costinot, and Rodríguez-Clare (2012) is that measuring the gains from trade requires addressing two questions: 1) How large are imports of factor services in the current trade equilibrium? And 2) how elastic is the demand for these imported services along the path from trade to autarky? The two sufficient statistics in the ACR formula are just specific answers to these two questions. If we do not trade much or if the factor services that we import are close substitutes to those that we would have access to under autarky, then the grey area in Figure 3 must be small, and so must be the gains from trade. In the next section, we will organize our discussion of the US gains from trade around these two questions.

Before delving into numbers, however, we want to highlight that aggregation issues will make addressing these questions more complex than the discussion so far might have suggested. By going from trade in goods to trade in factor services, we have argued that we can reduce the dimensionality of what needs to be estimated. The idea is that instead of estimating both production and demand functions for goods, one could just estimate the demand for factors from different countries. It does not follow, however, that factors from different countries can be aggregated into a single domestic factor and a single foreign factor. This aggregation holds true in the constant elasticity of substitution case, but not necessarily otherwise.

[6]It should also be clear that our approach simultaneously captures the benefits of importing and exporting. In general equilibrium, what matters is the relative price of US imports in terms of US exports, which is here summarized by the price of foreign factor services relative to US factor services. The reason why producer surplus does not appear in our computations is because firms' revenues are fully rebated to factors of production.

In general, the factor demand system that one needs to estimate may remain high-dimensional. Think of each country implicitly buying skilled workers, unskilled workers, and physical capital from around the world. For foreign factor services, there can be an empirical strategy for getting back to the one-factor case by treating all foreign factor services as a Hicks-composite good.[7] For domestic factor services, however, one cannot escape the fact that as the prices of all foreign factors are taken to their reservation values, one still needs to compute the relative prices of domestic factors under autarky. Again, this problem is similar to the one arising in the context of a single differentiated sector in the industrial organization literature. If Apple Cinnamon Cheerios (and a range of other brands) were no longer available, the prices of other ready-to-eat cereals might respond. Thus, when computing the equivalent variation associated with the removal of Apple Cinnamon Cheerios, such price responses should be included. Here, similar economic considerations are at play economy-wide. If foreign factor services complement certain domestic factor services (say, skilled labor), but substitute for others (say, unskilled labor), then moving to autarky will affect the relative price of these domestic factors, which will affect the utility that is possible under a situation of autarky as well as the income required to achieve that utility level.

Finally, it is worth emphasizing that the previous considerations arise under the maintained assumption that there exists a representative agent owning all domestic factors. We discuss the specific issues associated with the distributional consequences of trade later in this paper.

Gains from Trade in Practice

How Large Are the US Imports of Factor Services?

The simplest way to measure the US imports of foreign factor services is to look at the total value of all goods imported by the United States and assume that this is equal to the total payments to foreign factors used to produce these goods. Figure 4 shows the US import share, λ_F, computed as the share of total spending devoted to imports for each year from 1995 to 2014, using data from the World Input–Output Database (WIOD).[8] Despite a downwards spike during the Great Recession and a slight downtick in the latest years, the figure reveals a gradual increase in the importance of trade in the United States over the last two decades, with an increase in the US import share from about 6 to around 8 percent.

[7] For empirical purposes, this approach requires the existence of a price shifter that uniformly moves the price of all foreign factor services relative to domestic ones.

[8] As noted at the start of the paper, the ratio of imports is calculated here over total US spending, including spending on both final and intermediate goods, This number differs from the ratio of imports over GDP, which is 14 percent in 2014 according to the same database. Specifically, we use World Input–Output Database (WIOD) releases in 2013 and 2016; see Timmer, Dietzenbacher, Los, Stehrer, and de Vries (2015). The WIOD release 2013 has 35 sectors and 41 countries (including a synthetic Rest of the World); the WIOD release 2016 has 56 sectors and 44 countries (including a synthetic Rest of the World).

The low US import share partly reflects the fact that, as in most rich countries, US spending disproportionately falls on services, which are less likely to be traded. Although the US import share within manufacturing is much higher at 24.5 percent, manufacturing accounts for only 22 percent of total US spending. Another reason behind the low US import share is the size of the US economy, which mechanically raises the importance of intra-national relative to international trade. Thus, in spite of fairly low barriers to international trade, the US import share is one of the lowest in the world. In contrast, a small open economy like Belgium has an import share of expenditures above 30 percent.

The previous import shares implicitly abstract from global input–output linkages. In their absence, the US share of expenditure on imports, which we have plotted in Figure 4, must be equal to the share of spending by US consumers on foreign factor services, which is what our welfare formula emphasizes. In the presence of such linkages, it may not be.[9]

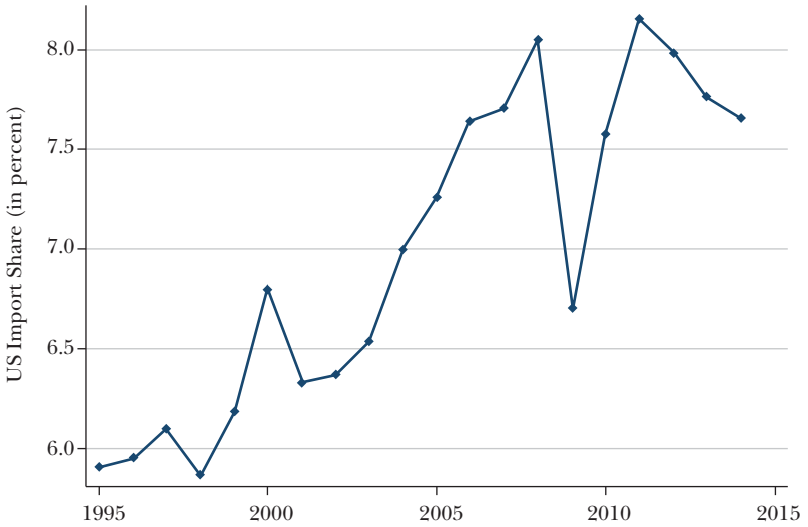
To assess the quantitative importance of these linkages, we again turn to the World Input–Output Database, which not only has the (gross) trade flows that we have used above, but also country-level input–output flows as well as final consumption and the value-added share in each sector-country. Under a proportionality assumption, this database provides a world input–output matrix that gives the share of output in each sector-country that comes from value added and from intermediate goods from each other sector-country.[10] As in Johnson and Noguera (2012), this matrix can be used to compute the factor services that every country i exports to every other country j both directly—that is, as value added embodied in exports for final consumption—and indirectly as part of worldwide input–output trade flows. This exercise reveals that import shares λ_F computed from gross trade flows (as in Figure 4) systematically understate the extent to which countries are open to international trade. When computed with value-added trade flows, the US import share becomes 11.4 percent rather than 8 percent. The same upward adjustment extends to all countries in the World Input–Output Database. The GDP-weighted average import share is 28 percent, whereas measured in gross flows it is 20 percent. In value-added terms, the United States and the world appear more open.

The key channel behind this adjustment is the fact that US domestic production also uses foreign factor services through imports of intermediate goods. This raises the share of spending by US consumers on foreign value added above the

[9] Because of intermediate goods, both measures also tend to be lower than the ratio of imports over GDP. Note, however, that even in the absence of intermediate goods, imports over GDP may differ from the expenditure share on imports because of trade imbalances. In the case of the United States, a country with a large trade deficit, this consideration further lowers the share of expenditure on imports relative to imports over GDP. Trade imbalances, of course, are related to the gains from intertemporal trade, which static models in the trade literature abstract from. We refer the reader interested in this issue to Costinot and Rodríguez-Clare (2014) and Heathcote and Perri (2014) for further discussions.

[10] Antras and de Gortari (2017) propose a model of global value chains potentially leading to deviations from such proportionality assumptions. In addition, de Gortari (2017) explores the issue empirically using Mexican firm-level data.

Figure 4
US Import Share



Source: Data for the years from 1995 to 2000 come from the 2013 release of the World Input–Output Database (WIOD) tables, while data for the years 2000 to 2014 come from the 2016 release of the WIOD tables.
Note: The figure contains the import share λ_F for the United States between the years 1995 and 2014. The time series coming from the two different releases of the WIOD were spliced so that they take the same value in the year 2000 (this adjustment is very small since the two series match very closely). The import share is calculated as gross total imports (adding across imports for intermediate input use and imports for final use) over total expenditure.

share of spending by US consumers on foreign goods. This is what happens, for instance, when some cars made and sold in the US include imports of German factor services used to produce the transmissions in those cars.

To add one more level of complexity, the import share computed with Johnson and Noguera's (2012) value-added flows excludes domestic factor services used to produce intermediate goods that were first exported and then imported back in the form of final goods—such services are counted as domestic, though they are traded in practice. Think of US imports of cars assembled in Mexico with US engines. Under autarky, however, all factor services that are currently traded would no longer be available. To compute the gains from trade, one should therefore expand the measure of spending on imports to include all traded factor services, not just the foreign ones.[11]

[11] In models with input–output loops, such as Alvarez and Lucas (2007), this means that the relevant price of foreign factor services is not the price of foreign value added, but the price of the bundle of inputs required for production abroad, a bundle that combines foreign value added with domestic value added, through imports of intermediate goods abroad.

Once this adjustment is made, again using the World Input–Output Database, the difference turns out to be small: for example, in the year 2014, the import ratio for the US increases from 11.4 to 12.1 percent, and similar results hold for the rest of the countries. Intuitively, the presence of trade costs tends to make this second adjustment—involving US factor services exported back and forth—small relative to the first adjustment—involving foreign factor services exported once.

How Elastic Is the Demand for Imported Factor Services?

The standard approach in the literature is to assume that the demand for factor services exhibits a constant elasticity of substitution, as in the models covered by the ACR formula. Letting λ_{ij} denote the share of expenditure by destination country j on factor services from origin country i and letting t_{ij} be an observable measure of the costs of trading factor services, then we can estimate this elasticity ε from the regression equation,

$$\ln \lambda_{ij} = \delta_i^o + \delta_j^d - \varepsilon \ln t_{ij} + \nu_{ij}.$$

The term δ_i^o is an origin-specific fixed effect that captures the role of the price of factor services in country i as well as anything that makes it costly for that country to export those services anywhere else. The term δ_j^d is a destination-specific fixed effect that captures how costly it is on average for country j to buy factor services. Finally, ν_{ij} is an error term that captures trade costs not included in t_{ij}.[12]

In practice, the observable costs of trading the factor services, t_{ij} may be tariffs, which vary across country-pairs thanks to preferential trade agreements (for example, Caliendo and Parro 2015) or freight costs (for example, Shapiro 2016; Adao, Costinot, and Donaldson 2017). The identifying assumption is that there is no correlation between the observable costs t_{ij} and the unobservable term ν_{ij}, so that one can estimate ε using a simple ordinary least squares regression. In their review of the literature using this general approach, Head and Mayer (2013) report a median estimate of $\varepsilon = 5$.

The standard approach raises some obvious concerns. One unfortunate feature of the empirical estimates of ε is that they come mostly from variation in trade costs and trade flows across foreign countries. Indeed, the previous regressions are often

[12] For historical reasons due to the pioneering work of Tinbergen (1962), such an equation is commonly referred to as a "gravity" equation; see Anderson (2011). Given the assumption that the demand for factor services exhibits a constant elasticity of substitution, the share of expenditure on factors from any origin country i in any destination j can be expressed as

$$\lambda_{ij} = \frac{(\tau_{ij} p_i)^{-\varepsilon}}{\sum_l (\tau_{lj} p_l)^{-\varepsilon}}$$

where p_i denotes the price of factor services from country i; τ_{ij} is a a summary measure of all trading frictions between i and j; and ε is the trade elasticity that we want to estimate. Assuming that $\tau_{ij} = \tau_i^o \tau_j^d t_{ij} \exp \nu_{ij}$, where t_{ij} is some observable trade cost shifter, and then taking the log of this equation leads to the regression equation in the main text, where the first two terms of that equation in the text become $\delta_i^o \equiv -\varepsilon \ln(\tau_i^o p_i)$ and $\delta_j^d \equiv -\ln\left(\sum_l (\tau_{lj} p_l)^{-\varepsilon}\right)$, respectively.

run excluding domestic flows. Yet, the relevant elasticity from a welfare standpoint is the elasticity of substitution between domestic and all foreign factors combined.[13]

Another concern is the assumption of a constant elasticity of demand for factor services. Adao, Costinot, and Donaldson (2017) have proposed a strict generalization of the constant elasticity of substitution assumption inspired by the work of Berry, Levinsohn, and Pakes (1995) in the field of industrial organization. Their "mixed CES" demand system allows the degree of substitutability across factor services from different countries to vary systematically with observable characteristics of those countries. While the average trade elasticity estimated by Adao, Costinot, and Donaldson (2017) is close to the estimates reviewed by Head and Mayer (2013), they find that poor countries, like China, are closer substitutes to other poor countries, like Bangladesh and Vietnam, than rich countries, like France and Germany. For the purposes of measuring the gains from trade, this suggests combining their estimates of factor demand with information about the entire vector of expenditure shares—that is, looking at factor demand for foreign shares on a country-by-country basis, not just overall.[14]

Although the demand system estimated in Adao, Costinot, and Donaldson (2017) relaxes the constant elasticity of substitution assumption, it is still very far from a nonparametric procedure that would flexibly trace out the response of the demand for foreign factors as we raise their prices to their reservation values. Given the limits on the number of observations and the amount of exogenous variation typically available in real world datasets, it is not surprising that few papers actually try to estimate the demand for foreign factors directly.

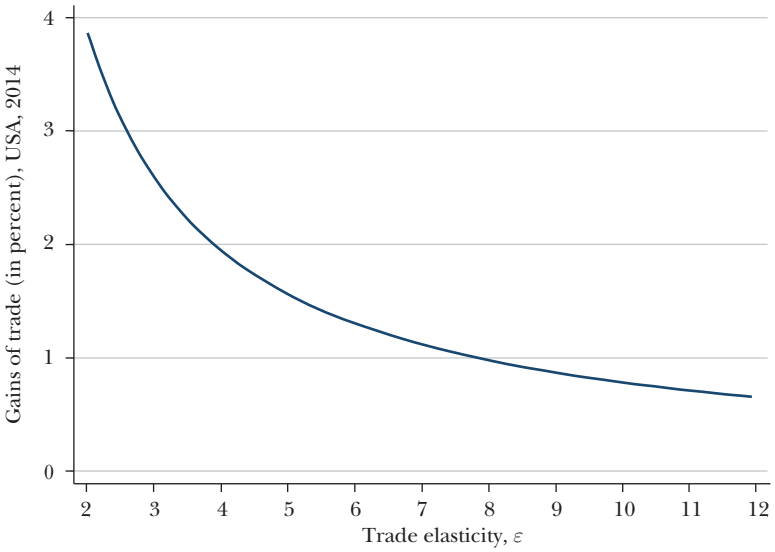
By far the most common alternative approach is to write down a computable general equilibrium model in which the values of all structural parameters indirectly pin down the demand for foreign factors. Such models typically involve a variety of nested relationships that each use their own constant elasticity of substitution assumption, both on the supply and demand side, though there is a tremendous amount of variation across these models in terms of the number of structural parameters that need to be estimated, from a single one in Eaton and Kortum (2002) to more than 13,000 in the latest model of the Global Trade Analysis Project (Hertel, McDougall, Narayanan, and Augiar 2012).

This leads to a trade-off. Parsimonious computable general equilibrium models can be estimated in a fairly transparent manner, which has contributed to their recent popularity in the field, but their predictions require stronger functional form restrictions, with many elasticities implicitly assumed to be identical or equal to one. This is what Dawkins, Srinivasan, and Whalley (2001) refer to as the "idiot's law of elasticities": all elasticities are equal to one until shown otherwise. Of course, both sets of parameters—those that are estimated in a transparent

[13]The work of Feenstra, Luck, Obstfeld, and Russ (forthcoming) is a notable exception. Their estimate points toward less substitutability between domestic and foreign sources than between different foreign sources.

[14]The issue of global-input linkages, of course, carries over to the estimation of factor demand. The empirical analysis of Adao, Costinot, and Donaldson (2017) is conducted under the assumption that the value of total factor services from a given country i sold in country j is equal to the total value of the goods sold by i in j.

Figure 5
The US Gains from Trade for Different Elasticities, ε



Source: Authors using data from the World Input–Output Database (WIOD) release 2016.
Note: Figure 5 shows the gains from trade for the United States in 2014, with $\lambda_D = 1 - \lambda_F$ equal to 92 percent and ε varying from 2 to 12, which is the range of values for ε estimated in Eaton and Kortum (2002). Gains from trade (GT) are computed according to $GT = 1 - \lambda_D^{1/\varepsilon}$.

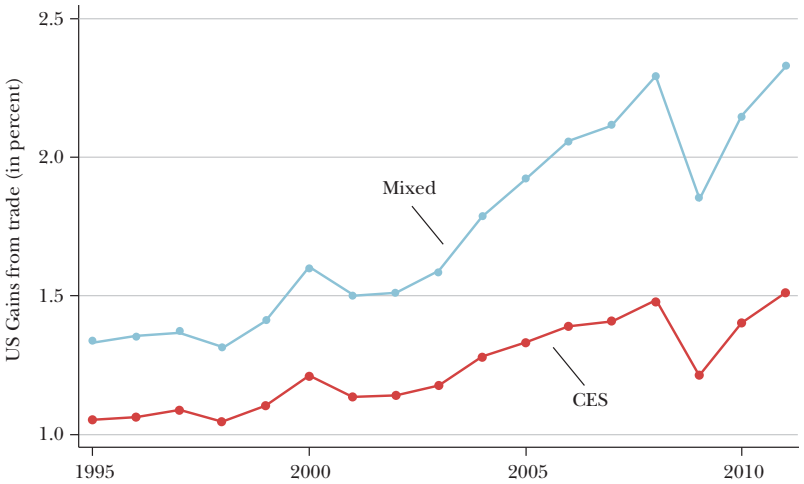
manner, like the trade elasticity within an industry, and those that are not, like the elasticity of substitution between goods from different industries or the elasticity of substitution between intermediate goods and primary factors—matter for the elasticity of the demand for foreign factor services, an important limitation that should be kept in mind.[15]

So How Large Are the US Gains from Trade?

We now put together the results of the previous two subsections and discuss their implications for the gains from trade. We first present results using the ACR formula for the size of gains from trade presented earlier. Figure 5 shows the gains from trade for the United States in 2014, with $\lambda_D = 1 - \lambda_F$ equal to 92 percent and ε varying from 2 to 12, which is the range of values for ε estimated in Eaton and Kortum (2002). Not surprisingly, the figure reveals that the gains from trade are highly sensitive to the value of epsilon: they are equal to 0.7 percent for $\varepsilon = 12$ and 4.1 percent for $\varepsilon = 2$. Evaluated at $\varepsilon = 5$, the median estimate in Head and Mayer (2013), the US gains from trade are equal to 1.7 percent. For comparison, a country like Belgium, with a much larger import share, has gains equal to 7.7 percent.

[15] In general, the elasticity of demand for foreign factors will be some average of upper-level and lower-level elasticities of substitution, with weights depending on the pattern of inter- and intra- industry trade (Ossa

Figure 6
The US Gains from Trade: CES (Constant Elasticity of Substitution) versus Mixed CES



Source: Data is from the World Input-Output Database (WIOD) release 2013.
Note: To explore the quantitative importance of the constant elasticity of substitution assumption, we apply the methodology described earlier using the more general factor demand system estimated by Adao, Costinot, and Donaldson (2017), a mixed constant elasticity of substitution framework (the line labeled "Mixed"). The line labeled "CES" shows results using the same data but restricting demand to be constant elasticity of substitution.

The previous estimates all implicitly rely on the assumption, embodied in a factor demand system using constant elasticity of substitution, that factor services from different countries are equally substitutable. To explore the quantitative importance of this restriction, we apply the methodology described earlier using the more general factor demand system estimated by Adao, Costinot, and Donaldson (2017). Results are reported in Figure 6.[16] From 1995 to 2011, the US gains from trade estimated in a mixed constant elasticity of substitution framework grow from about 1.3 to 2.3 percent. Using the same data, but restricting demand to be constant elasticity of substitution, one would have wrongly concluded that the US gains from trade over that same time period only grew from 1.1 to 1.5 percent.[17] The differential growth rates reflect the fact that over that time period, the United States not only imported more, but also imported relatively more from countries with lower GDP per capita,

2015; Costinot and Rodríguez-Clare 2014). Instead of assuming that the elasticity of substitution between industries is equal to one, one could therefore instead estimate that upper-level elasticity by targeting a trade elasticity for aggregate trade flows equal to that estimated in the literature. Such a procedure would guarantee that multi-industry computable general equilibrium models do not imply a demand for foreign factor services whose elasticity is counterfactually too low.

[16] We thank Rodrigo Adao for help with these counterfactual exercises.

[17] When restricting demand to be constant elasticity of substitution, Adao, Constinot, and Donaldson (2017) estimate a trade elastisticity ε around 6, hence the somewhat smaller gains from trade than before.

namely China, whose factor services have been estimated to be less substitutable to domestic factors in the United States.

As mentioned previously, a more common way in which the literature has departed from a factor demand system assuming constant elasticity of demand is indirectly via computable general equilibrium models. In Costinot and Rodríguez-Clare (2014), we report results for the gains from trade for the countries in the World Input–Output Database for the year 2008 under various structural assumptions. Focusing on the United States, the gains from trade increase from 1.8 percent in the model with a single sector to 4.4 percent of GDP in the multisector model with unitary elasticity of substitution across different industries. If one further allows for intermediate goods and assumes as in Caliendo and Parro (2015) that the elasticity of substitution between primary factors and inputs from different sectors is again one, the implied gains from trade increase to 8.3 percent of GDP.[18] Beside assumptions on the values of the previous elasticities, this much higher number reflects the fact that trade in intermediate goods raises the share of consumers' expenditure on factor services that are traded internationally.

One notable omission of the previous computable general equilibrium models is an explicit treatment of natural resources. If a country has no oil or gas resources and needs to import all that it uses, one would expect substantial losses from moving to autarky. If these resources are critical in the production of certain goods, and if those goods are not easily substitutable for other goods, this could lead to a demand for foreign factor services that becomes very inelastic as a country moves closer to autarky or that is much less elastic than the (relative) demand for factor services from different countries (as discussed earlier in footnote 13). Neither of these channels is present in the models reviewed in Costinot and Rodríguez-Clare (2014). In recent work, Fally and Sayre (2017) show that these considerations can lead to much larger gains from trade, especially in countries that lack diversified endowments of primary resources. For the United States, however, with its reasonably extensive endowments of natural resources, the effects that they uncover are small.

Growth, Distortions, and Redistribution

The Dynamic Gains from Trade

Many economists have a gut instinct that the gains from trade are much larger than those presented in the previous section. Could it be that our static approach misses important sources of "dynamic gains"? We briefly discuss a number of channels through which dynamic considerations may affect the magnitude of the gains from trade.

[18] The result for the gains from trade in the single-sector model is again computed using $\varepsilon = 5$, while for the multisector model with and without intermediate goods, the sector-level trade elasticities are those estimated in Caliendo and Parro (2015).

Perhaps more than any other dynamic consideration, the existence of innovation is most responsible for the view that the dynamic gains from trade may be large. A useful way to start thinking about this issue is to interpret the creation of new good varieties in the models by Krugman (1980) and Melitz (2003) as product innovation. The fact that these models are covered by the ACR formula suggests that the static model considered above includes an element of dynamic change. A similar insight holds for the growth model of Eaton and Kortum (2001). For dynamic gains from trade to arise, trade must lower the cost of innovation relative to the cost of production, as in the lab-equipment model of Rivera-Batiz and Romer (1991). Evidence on this connection is meager and contradictory. Bloom, Draca, and Reenen (2016) report evidence of a positive response of innovation to Chinese import competition among European firms. In the United States, however, the results of Autor, Dorn, Hanson, Pisano, and Shu (2016) point towards a negative effect of the China trade shock on innovation.

Another dynamic issue is related to frictions in the reallocation of factors of production, as emphasized, for instance, by Artuç, Chaudhuri, and McLaren (2010) and Caliendo, Dvorkin, and Parro (2015) in a labor market context. Intuitively, such frictions lead the economy to adjust slowly to changes in trade costs, implying a lower trade elasticity in the short run than in the long run. A sudden move to autarky would then imply losses that are high in the short run, but decline over time as resources get reallocated. In this case, the discounted stream of losses would be higher than the steady-state losses computed disregarding frictions in the static analysis of gains from trade.[19]

Our static analysis also abstracts from capital accumulation. Because some investment goods are tradable (like equipment goods), we expect allowing for capital accumulation leads to higher measured gains from trade, just like when allowing for trade in intermediate goods. Consistent with this view, Ravikumar, Santacreu, and Sposi (2017) find that the welfare effects of trade liberalization—specifically a decline of 20 percent in all international trade costs—are 23 percent higher when taking into account the impact on capital accumulation, relative to a static model.

More generally, very different welfare effects of trade may arise in the presence of the markup distortions typical in growth models, or because of knowledge spillovers within or across countries, as in some of the models covered in Grossman and Helpman (1991) as well as in more recent work by Sampson (2016), Perla, Tonetti, and Waugh (2015), and Buera and Oberfield (2016). As emphasized by Young (1991), trade could either increase or decrease growth, with ambiguous welfare consequences. The key force in these models, however, is not dynamics per se, but the presence of distortions, to which we now turn.

[19] Burstein and Melitz (2013) and Alessandria, Choi, and Ruhl (2014) make a related point in the context of models with firm-level heterogeneity, sunk costs, and transitional dynamics.

Distortions and the Gains from Trade

In the earlier textbook analysis of gains from trade, the private and social marginal benefits of importing foreign factor services are equalized. In the presence of distortions, private and social marginal benefits may not be aligned, further complicating the measurement of the welfare gains from trade.

As a matter of theory, the gains from trade could be larger, if trade alleviates distortions, or smaller, if it worsens them. This insight is just a corollary of the theory of the second-best: in the presence of domestic distortions, imposing another distortion (in this case, imposing autarky through an infinite tariff) could in theory make the economy better off.

One recent area of research has focused on the pro-competitive effects of trade and the extent to which opening up to trade may reduce markup distortions (for example, Epifani and Gancia 2011; Edmond, Midrigan, and Xu 2015; Arkolakis, Costinot, Donaldson, and Rodríguez-Clare 2012). After calibrating a model to the US economy in Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012), we find only small differences between the overall welfare gains predicted by models with and without variable markups, though our analysis relies on strong functional-form assumptions.

Another source of distortions that could affect the gains from trade is the presence of differential wage premia across firms or sectors (for example, Helpman, Itskhoki, and Redding 2010; Davis and Harrigan 2011; Święcki 2017). For instance, because wages for observably identical workers tend to be lower in agriculture than in manufacturing (and assuming that this pattern reflects differences in wage premia rather than sorting on unobservable characteristics), the gains from trade for countries that specialize in agriculture would be lower than those predicted by the earlier ACR formula. Following this logic, Święcki estimates that the gains from trade in the United States are slightly lower than those reported in the estimates in the previous section.

Redistribution

Our discussion has focused on the aggregate gains from trade: that is, the gains that would accrue to a representative consumer. In practice, of course, trade tends to create both winners and losers. How should one think about the overall gains from trade in such a situation?

This issue can of course be analyzed under the assumptions of a benevolent social planner and lump-sum transfers, but in practice, tax instruments are limited and shaped by political-economy considerations. To shed light on the magnitude of the gains from trade in such environments, Antràs, de Gortari, and Itskhoki (2017) and Galle, Rodríguez-Clare, and Yi (2017) propose to focus on an environment with exogenous income taxes and posit a social welfare function that displays constant inequality aversion. In the context of the United States, Galle, Rodríguez-Clare, and Yi focus on 1,444 worker groups defined by commuting zones and level of education. They find that 26 groups lose from trade, with one of those groups experiencing losses equal to 3.2 percent, while the overall gains are 1.5 percent. Poorer groups

experience stronger degrees of import competition and hence have lower gains from trade, so trade worsens between-group inequality. In their preferred specification, when the gains from trade are adjusted for an inequality-averse social welfare function, the gains for the United States remain positive and only slightly lower than the standard measure which ignores inequality (1.4 versus 1.5 percent), though this conclusion is clearly sensitive to the assumed degree of inequality aversion and the level of aggregation at which gains and losses are considered.

Concluding Remarks

The share of US expenditure on imports is smaller than in most other countries. To a large extent, this reflects the fact that for a large country like the United States, a significant fraction of trade occurs intra- rather than internationally. This basic observation implies that that the welfare gains from international trade in the United States are smaller than in most other countries. Although magnitudes vary greatly depending on how one infers the shape of the US demand for foreign factor services, the estimates of gains from trade for the US economy that we review range from 2 to 8 percent of GDP.

Although such gains are nothing to spit at, they may appear surprisingly small to some. It may be the case that if the United States were to approach autarky, the US demand for foreign factor services would become much less elastic, revealing the importance of some critical foreign inputs and, in turn, much larger gains from trade. Both extrapolations based on the direct estimation of demand and the predictions of computable general equilibrium models may miss that. As mentioned in our introduction, the only direct evidence of the cost of autarky for the United States that we have comes from the Jeffersonian trade embargo between December 1807 and March 1809. Irwin (2005) estimates its welfare cost to be around 5 percent of US GNP in 1807. Interestingly, this is similar in magnitude to the welfare gains from trade estimated by Bernhofen and Brown (2005) in the context of Japan's emergence from autarky after 1858.

Of course, the estimates reviewed in this article are not meant as a guide for future economic policy. Current trade volumes are crucial for measuring the welfare cost of autarky, not the potential gains from further trade expansion. If international trade flows were to grow, so would the gains from international trade.

■ *The authors thank Rodrigo Adao, Dave Donaldson, Thibault Fally, Mark Gertler, Gordon Hanson, Enrico Moretti, and Timothy Taylor for very helpful comments and Mauricio Ulate for excellent research assistance.*

References

Adao, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2017. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade." *American Economic Review* 107(3): 633–89.

Alessandria, George, Horag Choi, and Kim Ruhl. 2014. "Trade Adjustment Dynamics and the Welfare Gains from Trade." NBER Working Paper 20663.

Alvarez, Fernando, and Robert E. Lucas. 2007. "General Equilibrium Analysis of the Eaton–Kortum Model of International Trade." *Journal of Monetary Economics* 54(6): 1726–68.

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review* 69(1): 106–16.

Anderson, James E. 2011. "The Gravity Model." *Annual Review of Economics* 3: 133–160.

Antràs, Pol, and Alonso de Gortari. 2017. "On the Geography of Global Value Chains." NBER Working Paper 23456.

Antràs, Pol, Alonso de Gortari, and Oleg Itskhoki. 2017. "Globalization, Inequality and Welfare." *Journal of International Economics* 108: 387–412.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2012. "The Elusive Pro-Competitive Effects of Trade." NBER Working Paper 21370.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102(1): 94–130.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review* 100(3): 1008–45.

Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. 2016. "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents." NBER Working Papers 22879.

Baldwin, Richard, and James Harrigan. 2011. "Zeros, Quality and Space: Trade Theory and Trade Evidence." *American Economic Journal: Microeconomics* 3(2): 60–88.

Bernhofen, Daniel M., and John C. Brown. 2005. "An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan." *American Economic Review* 95(1): 208–25.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4): 841–90.

Bhagwati, Jagdish N., Arvind Panagariya, and T. N. Srinivasan. 1998. *Lectures on International Trade.* Boston, Massachussetts: MIT Press.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2016. "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." *Review of Economic Studies* 83(1): 87–117.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121(2): 541–85.

Buera, Francisco J., and Ezra Oberfield. 2016. "The Global Diffusion of Ideas." NBER Working Paper 21844.

Burstein, Ariel, and Marc J. Melitz. 2013. "Trade Liberalization and Firm Dynamics." Chap. 7 in *Advances in Economics and Econometrics: Tenth World Congress,* Vol. 2: *Applied Economics,* Cambridge University Press.

Caliendo, Lorenzo, Maximiliano A. Dvorkin, and Fernando Parro. 2015. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." Federal Reserve Bank of St. Louis Working Paper 2015-9. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2646148##.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82(1): 1–44.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." Chap. 4 in *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. North Holland.

Davis, Donald R., and James Harrigan. 2011. "Good Jobs, Bad Jobs, and Trade Liberalization." *Journal of International Economics* 84(1): 26–36.

Dawkins, Christina, T. N. Srinivasan, and John Whalley. 2001. "Calibration." Chap. 58 in *Handbook of Econometrics,* vol. 5. Elsevier.

de Gortari, Alonso. 2017. "Disentangling Global Value Chains." https://scholar.harvard.edu/alonsodegortari/publications/disentangling-global-value-chains.

Eaton, Jonathan, and Samuel Kortum. 2001. "Trade in Capital Goods." *European Economic Review* 45(7): 1195–1235.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography and Trade." *Econometrica* 70(5): 1741–79.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, Markups and the Gains from Trade." *American Economic Review* 105(10): 3183–3221.

Epifani, Paolo, and Gino Gancia. 2011. "Trade, Markup Heterogeneity and Misallocations."

Journal of International Economics 83(1): 1–13.

Fally, Thibault, and James Sayre. 2017. "Commodity Trade Matters." Unpublished paper, University of California, Berkeley.

Feenstra, Robert C. 1994. "New Product Varieties and the Measurement of International Prices." *American Economic Review* 84(1): 157–77.

Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ. Forthcoming. "In Search of the Armington Elasticity." *Review of Economics and Statistics*.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi. 2017. "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade." NBER Working Paper 23737.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

Hausman, Jerry A. 1997. "Valuation of New Goods under Perfect and Imperfect Competition." Chap. 5 in *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon. University of Chicago Press.

Head, Keith, and Thierry Mayer. 2013. "Gravity Equations: Workhorse, Toolkit, Cookbook, Workhorse." Chap. 3 in *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. North Holland, Elsevier.

Heathcote, Jonathan, and Fabrizio Perri. 2014. "Assessing International Efficiency." Chap. 9 in *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. North Holland, Elsevier.

Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica* 78(4): 1239–83.

Hertel, Thomas W., Robert A. McDougall, Badri G. Narayanan, and Angel H. Aguiar. 2012. "Behavioral Parameters." Chap. 14 in *GTAP Data Bases: The GTAP 8 Data Base Documentation*. Center for Global Trade Analysis, Purdue University. https://www.gtap.agecon.purdue.edu/databases/v8/v8_doco.asp.

Irwin, Douglas A. 2005. "The Welfare Costs of Autarky: Evidence from the Jeffersonian Embargo, 1807–1809." *Review of International Economics* 13(4): 631–45.

Johnson, Robert C., and Guillermo Noguera. 2012. "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of International Economics* 86(2): 224–36.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70(5): 950–59.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–1725.

Nevo, Aviv. 2003. "New Products, Quality Changes, and Welfare Measures Computed from Estimated Demand Systems." *Review of Economics and Statistics* 85(2): 266–75.

Ossa, Ralph. 2015. "Why Trade Matters After All." *Journal of International Economics* 97(2): 266–77.

Perla, Jesse, Christopher Tonetti, and Michael E. Waugh. 2015. "Equilibrium Technology Diffusion, Trade and Growth." NBER Working Paper 20881.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110(4): 705–29.

Ravikumar, B., Ana Maria Santacreu, and Michael Sposi. 2017. "Capital Accumulation and Dynamic Gains from Trade." *Federal Reserve Bank of St. Louis Working Paper* 2007-005D.

Rivera-Batiz, Luis A., and Paul M. Romer. 1991. "Economic Integration and Endogenous Growth." *Quarterly Journal of Economics* 106(2): 531–55.

Sampson, Thomas. 2016. "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth." *Quarterly Journal of Economics* 131(1): 315–80.

Shapiro, Joseph S. 2016. "Trade Costs, CO_2, and the Environment." *American Economic Journal: Economic Policy* 8(4): 220–54.

Święcki, Tomasz. 2017. "Intersectoral Distortions and Welfare Gains from Trade." *Journal of International Economics* 104: 138–56.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. 2015. "An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production." *Review of International Economics* 23(3): 575–605.

Tinbergen, Jan. 1962. *Shaping the World Economy*. New York: The Twentieth Century Fund.

Young, Alwyn. 1991. "Learning by Doing and the Dynamic Effects of International Trade." *Quarterly Journal of Economics* 106(2): 396–406.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 25–46

Alternative Sources of the Gains from International Trade: Variety, Creative Destruction, and Markups

Robert C. Feenstra

The modern theory of international trade allows for several sources of the gains from trade in addition to traditional comparative advantage. We discuss these sources and provide estimates of the gains for the United States and other countries. It turns out that the formula used to measure these new gains can be used to measure the gains from traditional comparative advantage, too, as we shall explain.

The first alternative source of gains from trade is not that new, and refers to the gains from increased variety of products. These gains were recognized by David Ricardo (1817, chap. 7) when he wrote: "Foreign trade, then, … [is] highly beneficial to a country, as it increases the amount and variety of the objects on which revenue may be expended." In Ricardo's time, these new varieties included cacao from Africa, spices from Southeast Asia, sugar from the Caribbean, and tea from India, all shipped to Europe and America. The production of these goods relies heavily on climate and soil, so they might be thought of as reflecting fundamental comparative advantage. The modern theory of international trade, however, allows countries to trade product varieties that do not have such fundamental differences, as with different types of cheese from France and Holland or different types of cars from Germany and Japan. Indeed, the modern theory often assumes that *all* countries and industries are producing differentiated varieties, which means that firms

■ *Robert C. Feenstra is the C. Bryan Cameron Distinguished Chair in International Economics, University of California-Davis, Davis, California, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is rcfeenstra@ucdavis.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.25

doi=10.1257/jep.32.2.25

are operating under monopolistic competition rather than perfect competition. That is, firms retain some limited monopoly power in their unique product varieties so that prices are above marginal cost, but there is free entry of firms in the long run, so that markups just cover the fixed costs of entry and industry profits are zero.[1]

A second additional source of gains from trade also comes from the monopolistic competition model, but in contrast with the first-generation models that had homogeneous firms (Krugman 1979, 1980, 1981; Helpman and Krugman 1985), the second-generation model of monopolistic competition allows firms to be heterogeneous in their productivity levels (Melitz 2003; Chaney 2008). In this setting, international trade allows the most-productive firms to expand their sales through exports, while the least-productive firms are forced to exit because of competition with imports. This is the process that Joseph Schumpeter (1942, part 2, chap. 7) had in mind when he wrote: "The opening up of new markets, foreign or domestic, … revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism." Through this process, average productivity in the industry rises due to increased sales of the most-productive firms. These gains are analogous to those from traditional comparative advantage in a Ricardian model, when the relatively most-productive industries expand their share of output through exporting. Indeed, as we shall see, the Ricardian model with stochastic technologies due to Eaton and Kortum (2002) has a very similar formula for the gains from trade as in the heterogeneous firm model.

A third alternative source of the gains from trade occurs when competition between firms in different countries leads them to reduce the markups that they charge. If there is only a single monopolistic domestic firm, then the reduction in markups leads to consumer gains and a reduction in the monopoly distortion, but the consumer gains would be substantially offset by the firm losses. In a situation of monopolistic competition, however, the *entire* reduction in consumer prices potentially becomes a social gain. That is because the free entry of firms under monopolistic competition drives industry profits to zero, and so the consumer gain from a reduction in markups is not offset by any fall in profits (which remain at zero). There is still a potential offset to the consumer gain, however, because with some firms exiting there will be reduced product variety, and the net effect on welfare must take all these effects into account.

To illustrate the potentially ambiguous effect of competition between markets on social welfare, consider the comment from the popular food author Michael Pollan, who has said: "America ships tons of sugar cookies to Denmark and Denmark ships tons of sugar cookies to America. Wouldn't it be more efficient just to swap recipes?"[2]

[1] Research in international trade is nearly always done in a long-run general equilibrium setting because it is felt that short-run or partial equilibrium settings do not lend themselves to theories of the pattern of trade. Likewise, the gains from trade are usually evaluated in a long-run general equilibrium model, where allowing free entry in all industries eliminates the need to keep track of the distribution of industry profits throughout the economy.

[2] Remarks by Michael Pollan at the University of California, Davis, Mondavi Center, February 12, 2007.

Pollan is referring to the social cost of shipping goods between markets, which he views as a waste, but he is not taking into account the beneficial effect of competition between American and Danish firms. If having the cookies sold between the two countries leads to lower prices due to reduced markups, then there can be social gains despite the shipping costs. Whether there *are* social gains or not will depend on whether there is free entry into the sugar cookie industry, as we shall discuss.

The gains from trade are often computed as compared to autarky (that is no trade at all), and we will report some of these estimates. But it is important to consider also the gains from free trade as compared to actual restrictions on trade such as tariffs and quotas. In other words, we are interested in the cost of such trade restrictions as they are used in practice. These restrictions are often more complex than the simplified version used in modern models, and therefore have some unexpected costs that we shall describe.

Product Variety in Trade

Measuring gains from new varieties is difficult because there is no observed price for a product before it is available. The solution given many years ago by Hicks (1940) is that the relevant price of a product before it is available is the *reservation price* for consumers, which is the price that is so high that demand is zero, and the drop from the reservation price to the observed price can be used to measure the consumer gains from the appearance of that new good. This idea of Hicks has been applied to new products by Hausman (1997, 1999). But when we try to apply this idea to the appearance of new product varieties from many countries due to international trade, we run into a problem: if each exporting country is providing a different variety, then we potentially have hundreds or thousands of new product varieties through trade, and it is impractical to estimate the reservation price for each.

A common way to address this problem in this literature is to adopt a constant-elasticity-of-substitution utility function. Provided that there are many goods so that the share of income spent on any good is small, then the elasticity of demand is approximately equal to the elasticity of substitution, which we denote by $\sigma > 1$. The constant-elasticity demand curve approaches the vertical axis asymptotically as the price goes to infinity, but it does not touch this axis. So the reservation price is infinity, but the area under the demand curve measuring the consumer surplus gain from a new product is bounded and well-behaved: the consumer gains from the new variety shrink as the elasticity of substitution is higher, indicating that the new good is a closer substitute for an existing good.

With the assumption of a constant-elasticity-of-substitution (CES) utility function in Feenstra (1994), I show how new varieties affect an exact cost-of-living index (which focuses on the cost of purchasing goods that provide a constant level of utility). After making some assumptions—for example, that profits are zero due to free entry so that consumer gains from trade are equal to the social gains—one can invert the change in the cost-of-living index to obtain an expression for the welfare

gain from the new products.[3] Suppose that we compare two different equilibria of the home economy with different amounts of imports. Let the variable λ denote the *share of home spending on its domestic products* in the first equilibrium, and let λ' denote this share in the second equilibrium. If we start in autarky where imports are zero and then move to free trade, it follows that $\lambda = 1$ initially and $\lambda' < 1$ under free trade. Then the social gains from trade can be expressed as:

$$\text{Gains from trade} = \left(\frac{\lambda'}{\lambda}\right)^{-1/(\sigma-1)}.$$

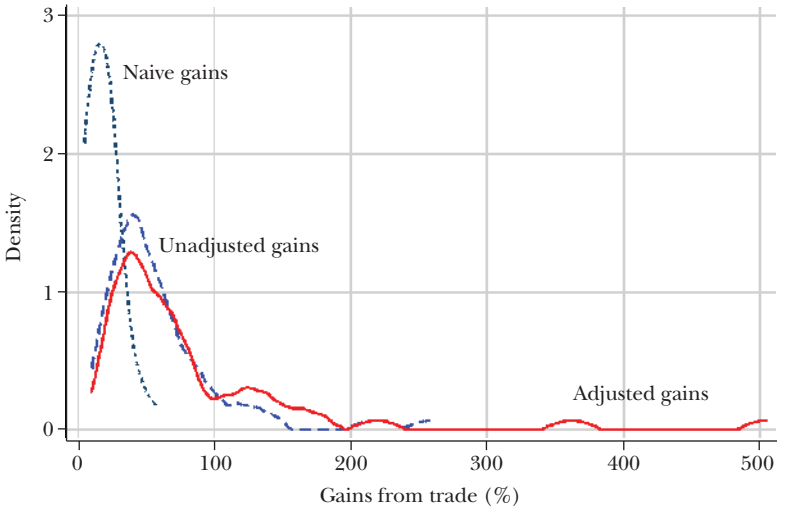
For example, suppose that $\sigma = 3$ so that $-1/(\sigma - 1) = -1/2$, and that 90 percent of expenditure with trade goes to domestic goods, with 10 percent going to imports. Then, the formula equals $(0.9)^{-1/2} \approx 1.05$, which suggests that the availability of the new imported varieties, with the specified elasticity and spending level, raises welfare by about 5 percent.

Broda and Weinstein (2006) use this approach, extended to many industries, to estimate the gains from trade for the United States due to the expansion of import varieties. As in Armington (1969), varieties of a good from different countries (like cars from the United States, Germany, and Japan) are treated as imperfect substitutes, and the elasticity of substitution between countries is estimated using the techniques in Feenstra (1994). So importing a "new variety" really means that the United States starts importing a good from a country that did not export that good to the United States before. Broda and Weinstein find gains due to import variety that grow by 1.2 percentage points per year over 1972–2001 to a total of 28 percent of import expenditure in 2001, or 2.6 percent of GDP. While this number may seem modest, remember that it is an estimate of the *incremental* gains from growing import variety from 1972 to 2001, not the total gains from trade as compared to autarky.

Gains from trade as compared to autarky for many countries are computed by Ossa (2015), using the above formula with $\lambda = 1$ in each industry. For the United States, he finds gains from trade equal to 13.5 percent of GDP, which is at the low end as compared to other countries. The median gain from trade in varieties across all countries is 55.9 percent of GDP. A number of other small European countries like the Netherlands, Norway, and Sweden are close to those median gains. The country with the highest gains is Belgium at 505.2 percent of GDP. Larger economies tend to have lower gains, because their share of spending on domestic firms is higher. Thus, gains from trade in varieties as compared to autarky are 30.8 percent of GDP for China, 35.3 percent for France, and 21.4 percent for Japan. Ossa also takes into account nontraded goods (which lowers the measured gains from trade) and interindustry flows of intermediate inputs (which raises the measured gains).

[3] Some other relevant assumptions are that labor is the only factor of production and that marginal labor costs are fixed. Normalizing the wage at unity and with constant markups due to the constant elasticity of demand, it follows that prices are also fixed. In Feenstra (1994), I show that the cost of living index for the home consumer between these two equilibria is $(\lambda'/\lambda)^{1/(\sigma-1)}$. The formula in the text for the social gains is the inverse of this expression.

Figure 1
Gains from Trade
(percent of GDP)



Source: This figure graphs results from Ossa (2015).
Note: The horizontal axis measures the gains from trade (in percent), and the vertical axis is the frequency (or density) of countries with those gains. Three cases are graphed: "naive" gains, which, rather than allowing elasticity of substitution between countries to vary across industries, uses an *average* elasticity for all industries; "unadjusted" gains, which do not take into account that some goods are nontraded or that intermediate inputs are used in production; and "adjusted" gains, which take all these factors into account. Gains from trade are graphed with a bandwidth of 0.1.

A histogram of Ossa's (2015) estimates of the gains across all countries is provided in Figure 1, where the horizontal axis measures the gains from trade (in percent), and the vertical axis is the frequency (or density) of countries with those gains. We distinguish several different cases: "naive" gains, which, rather than allowing elasticity of substitution between countries to vary across industries, uses an *average* elasticity for all industries; "unadjusted" gains, which do not take into account that some goods are nontraded or that intermediate inputs are used in production; and "adjusted" gains, which take all these factors into account. The naive gains are the smallest because they do not allow for low estimates of the elasticity in some industries, which would increase the gains from trade. The unadjusted gains are less than the adjusted gains for most countries, because, in making the adjustment, the extra gains from intermediate input linkages usually exceed the reduced gains from having nontraded goods. The importance of input–output linkages to the calculation of gains from trade is stressed by Chaney and Ossa (2013) and Melitz and Redding (2014).

The median gains across all 50 countries considered by Ossa (2015) are 16.5 percent of GDP for the naive gains, 48.6 percent for the unadjusted gains, and 55.9 percent for the adjusted gains (as noted above). These estimates may seem high, for two reasons. First, since individuals often consume just one variety of a differentiated product, we can ask whether the increase in variety due to international trade really translates into the gains computed from a constant elasticity of substitution utility function. This question is answered in the affirmative by Anderson, de Palma, and Thisse (1992). They show that even when each individual consumes their preferred variety, social welfare can quite possibly be measured by a CES utility function.[4]

A different concern with the gains from imported variety are that they might be offset by reduced *domestic* variety as home firms leave the market due to import competition. Such exit of home firms does not occur in the simplest monopolistic competition model due to Krugman (1980), which assumes constant-elasticity-of-substitution preferences and homogeneous firms, that is, all firms in each country have the same marginal and fixed costs. If transportation costs are zero, then it turns out that home firms faced with import competition can still sell enough abroad that they do not exit as trade is opened. But that result is highly stylized, and in reality we expect to observe the exit of some home firms, as we will find in a model with heterogeneous firms discussed in the next section. The heterogeneous firm model will allow for additional gains from trade due to creative destruction, as the least-efficient firms exit. Even though that exit results in less product variety, the gains from trade may very well be larger than Ossa (2015) measured using the formula for product variety described in this section.

Productivity Differences and Gains from Trade

International trade can also lead to gains through the creative destruction process in which firms of higher productivity expand their output through exporting, while firms of lower productivity exit through bankruptcy. This source of gains from trade is emphasized in the model of monopolistic competition with heterogeneous firms due to Melitz (2003; also exposited in this journal in Melitz and Trefler 2012).

A good example of an industry where firms differ in their productivities and where the most-productive firms benefit from exports is the production of solar (photovoltaic) cells. There are many such firms in the United States and worldwide, with rapidly growing capacity in China. Indeed, import competition from

[4]The assumptions needed to obtain a CES utility function are that individuals have utility that includes a "random" component reflecting their own tastes for each variety, and that this random component is distributed across varieties in the same way as for the logit demand system. In order to obtain the CES rather than the logit demand system, individuals have to be able to consume multiple units of their preferred variety rather than just a single unit.

China led two US-based manufacturers, Suniva and SolarWorld, to file a petition with the US International Trade Commission in 2017 asking for the application of import tariffs. A tariff of 30 percent on solar cells and modules was approved by President Trump in January 2018 (which will fall by 5 percent per year over the next four years). The tariff was opposed by US industry groups such as the Solar Energy Industries Association, which represents companies that install solar panels. Interestingly, the import tariff will not protect another US firm, First Solar, which both produces cells with a more advanced technology (to which the tariffs do not apply) and is also a leading US installer of solar systems worldwide. First Solar's stock price fell on the announcement of the tariff. Thus, it appears that the tariffs applied by the United States will enable less-efficient firms like Suniva and SolarWorld to survive, while not benefiting the most-efficient, exporting firms.

There are many other industries where plants differ in their productivities, as has been well documented in the industrial organization literature. Syverson (2004a, b; 2011) finds that the productivity distribution across plants is related to the extent of product substitutability: when substitution between products is greater, then dispersion is smaller, because the lower-productivity plants find it harder to survive. In the macro literature, too, imperfect substitution across products is key when examining the potential productivity gains due to creative destruction. For example, in Garcia-Macia, Hsieh, and Klenow (2016) and Aghion, Bergeaund, Boppart, Klenow, and Li (2017), an approach similar to the one described here for looking at gains from variety is used to measure productivity gains when one plant is replaced by another.

With international trade, the recent example of the US tariff on solar cells shows us that *barriers to trade* between countries allow for a dispersion of productivities that is higher than would otherwise occur. Such trade barriers can take many forms: tariffs, shipping costs, difficulties of communication or making contracts between countries, and others. In the initial formulation of the Melitz model, the trade costs were modeled as "iceberg" costs of trade—that is, some fraction of the good "melts" in route to the destination country. While Samuelson (1952) proposed this way of modeling transport costs as a theoretical simplification, it has been used so often since then, including in the Melitz model, that it is sometimes not realized how sensitive certain results are to that assumption. We will continue with the simplification of "iceberg" costs of trade throughout this section, but then broaden the discussion to incorporate actual tariffs (as in the solar cell example) later in the paper.

When the "iceberg" costs of trade are reduced, the most-productive firms expand their output by exporting more and the least-productive firms exit. As a result, average industry productivity grows and consumer prices fall. This productivity-enhancing effect of creative destruction has been demonstrated for Canada by Trefler (2004) under the Canada–US Free Trade Agreement, and for a broader sample of countries and free-trade agreements by Badinger (2007a, 2008; for evidence on Chile and Mexico, see also Tybout, de Melo, and Corbo 1991; Tybout

and Westbrook, 1995). As an example of this literature, Trefler (2004, p. 870) finds that:

> The Canada-U.S. Free Trade Agreement provides a unique window onto the effects of a reciprocal trade agreement on an industrialized economy (Canada). For industries that experienced the deepest Canadian tariff cuts, the contraction of low-productivity plants reduced employment by 12 percent while raising industry-level labor productivity by 15 percent. For industries that experienced the largest U.S. tariff cuts, plant-level labor productivity soared by 14 percent. These results highlight the conflict between those who bore the short-run adjustment costs (displaced workers and struggling plants) and those who are garnering the long-run gains (consumers and efficient plants).

This comment alerts us that the heterogeneous firm model has within it the potential for losses by less productive firms and their workers. In the Melitz (2003) model, such losses are assumed away with a single type of labor that is perfectly mobile between industries. An alternative model due to Yeaple (2005) has workers varying by ability, with heterogeneous technologies available in one sector. Higher-ability workers match with the more advanced technology, workers with mid-level abilities use the inferior technology, and the lowest-ability workers are in a traditional sector, which is not traded in equilibrium. In this setting, opening trade between countries generates gains for the highest-ability workers without any losses to the lowest-ability workers, but the mid-level workers using the inferior technology suffer a loss in wages. As noted by Helpman (forthcoming), this pattern of wage changes accords with the "hollowing out" of the US labor market, whereby workers at the upper and lower ends of the skills distribution have seen real gains, but not workers in the middle of the distribution (Autor 2014). Helpman (forthcoming) comprehensively discusses the potential losses to workers in recent models of international trade with matching between heterogeneous workers and firms.

Our goal here is to develop a general formula for the long-run gains due to creative destruction, which in turn requires thinking about the distribution of productivity across firms and how it would be altered by trade. To that end, it is convenient to simplify the Melitz model by adopting a Pareto distribution of productivities across firms, as proposed by Chaney (2008), in what is called the Melitz–Chaney model. In this setting, one parameter governs the variance of productivities across firms, which we will denote by θ. If $\theta = 1.16$, for example, then 20 percent of the firms account for 80 percent of the output. The higher is θ, the less spread-out are the firms in their productivities, and as θ approaches infinity then we are back in the case of homogeneous firms, all with the same productivity.

Once we know the parameter θ, then the only other information needed to compute the gains from trade in the one-sector Melitz–Chaney model is the *share of home spending on domestic products,* which we again denote by the variable λ in an initial equilibrium. Let λ' denote this share in a second equilibrium. As discussed

earlier, if we start in autarky and then move to trade, then $\lambda = 1$ initially and $\lambda' < 1$. With these assumptions, Arkolakis, Costinot, and Rodríguez-Clare (2012) show that the home gains from trade between these two equilibria are:

$$\text{Gains from trade} = \left(\frac{\lambda'}{\lambda}\right)^{-1/\theta}$$

For example, compare the United States in autarky (so $\lambda = 1$) with the imports of 15 percent of GDP, so that the share of home spending on its domestic products equals 0.85. Suppose we choose the Pareto parameter $\theta = 4$. Then the US gains as compared to autarky equal $(0.85)^{1/4} \approx 1.04$, which represents gains of 4 percent of GDP. This example gives gains that are unrealistically small, however, because the above formula applies to a one-sector economy, with trade in final goods only. When we take into account many industries, some of which have lower values of θ (so the firms with higher productivities are more spread out), and also include realistic input-output flows between industries, then the gains from trade become much larger, as we illustrated earlier in Figure 1.

Notice that the formula above is much the same as the formula in the previous section, except that they differ by the exponent used in each case. It is fair to ask what has happened to the elasticity of substitution σ, which appeared in the formula of the previous section, and measures the preference for different product varieties. I answer this question in Feenstra (2010; 2015, chap. 6), arguing that the gains from new import varieties in the Melitz–Chaney model *cancel out* with the losses due to reduced domestic varieties, as some domestic firms faced with import competition will exit the market. In other words, there are zero net gains from product variety in the one-sector Melitz–Chaney model; all of the gains come from *rising average productivity* in the industry.

In place of gains from trade due to product variety, the gains from trade in the one-sector Melitz–Chaney model are *entirely* due to creative destruction—that is, to low-productivity firms exiting and high-productivity firms expanding their output through exporting. Despite having no gains from product variety, it is still quite possible that the Melitz–Chaney model (with heterogeneous firms) has gains that are *greater* than would occur due to importing new varieties (in the model with homogeneous firms). This result holds for two reasons. First, Melitz and Redding (2015) show that the heterogeneous firm model allows for the greatest expansion of output by the high-productivity firms when trade is opened, leading to more trade and therefore a *lower value* of λ' (the share of home spending on domestic products in the second equilibrium) than in the homogeneous firm model. With a lower value of λ' in the formula from gains to trade just above, and a negative exponent, the gains from trade are higher.

Second, even if we simply take the domestic share λ', used in the formula just above and in the previous section, as given from the data, we can *still find* that the gains are higher in the heterogeneous firm model due to the differing exponents used in each case. Simonovska and Waugh (2014) present a method to estimate the parameters of the gains from trade formula: either the elasticity of substitution σ

from the homogeneous firm model of the previous section, or the Pareto parameter θ from a heterogeneous firm model. They show that when estimating each of these models on the same data, there is a systematic relationship between the estimated parameters: namely, that $\hat{\theta} < (\hat{\sigma} - 1)$.[5] So even when using the same value of λ', we will still get greater gains from the process of creative destruction.

Several studies have compared gains from trade with homogenous firms reflecting product variety versus with heterogeneous firms reflecting creative destruction. For example, Balistreri, Hillberry, and Rutherford (2011) find gains that are four times higher with heterogeneous firms than the gains from product variety.[6] Smaller differences are found by Costinot and Rodríguez-Clare (2014). Moving from free trade to a hypothetical 40 percent worldwide tariff, they find average country losses of 7.0 percent with heterogeneous firms as compared to 5.3 percent with homogeneous firms; but in a model without intermediate inputs, the loss due to the tariff with heterogeneous firms is slightly smaller than with homogeneous firms.[7]

We conclude that while there is a *presumption* that the gains from creative destruction exceed the gains from new varieties, this result does not always hold in calibrated models. The same is true when we consider results from actual changes in trade barriers. For the Canada–US Free Trade Agreement, the above quotation from Trefler (2004) shows that productivity in Canadian industries was indeed boosted, and by a considerable amount. But Hsieh, Li, Ossa, and Yang (2018) have found that exit of Canadian firms led to such a fall in product variety that it overpowered the entry of new US exporters, such that overall consumer variety in Canada fell in some industries. That empirical result could not occur in the Melitz–Chaney model (where overall variety is constant due to the Pareto assumption). Canadian welfare still rose from the free trade agreement, but not by as much as in the formula we have described in this section.

Finally, it is important to point out that the formula for the gains from trade in this section also applies as a measure of the gains due to traditional comparative advantage in the perfectly competitive model of Eaton and Kortum (2002). These authors consider a continuum of products that have constant-elasticity-of-substitution preferences between them. The technologies for producing each product are independently drawn in each country and follow a Fréchet distribution with the parameter θ. This distribution has the convenient property that the minimum value in a sample

[5] As already explained from the work of Syverson (2004a, 2004b, 2011), the dispersion of productivity across plants is limited by the elasticity of substitution. That restriction takes the form of the inequality $\theta > (\sigma - 1)$, which is required for firms of differing productivities to occur in equilibrium. The estimates from Simonovska and Waugh (2014) do not contradict that theoretical condition because the estimates are obtained from *different models* applied to the same data.

[6] Balistreri, Hillberry, and Rutherford (2011) consider a simplified version of the gains from product variety, where the differences in product varieties across countries are exogenously given in what is called the "Armington model." This model does not include any increasing returns to scale or monopolistic competition, so as explained by Simonovska and Waugh (2014), it gives low gains from trade.

[7] As noted by Costinot and Rodríguez-Clare (2014, p. 233), tariffs can lead to changes in product variety in a multisector Melitz–Chaney model. For this reason the comparison of the heterogeneous and homogeneous firm models is not a pure comparison of creative destruction versus product variety gains.

of N independent draws from a Fréchet distribution is also distributed as Fréchet, but with a reduced variance. There is a natural application of this property to international trade, since for each product, the buyer will be choosing the minimum-cost supplier over all N potential supplying countries.

The Eaton–Kortum (2002) model has many of the same reduced-form properties as the Melitz–Chaney model, but it assumes perfect competition rather than monopolistic competition. One of these properties is the formula for the gains from trade, which in the Eaton–Kortum model is identical to the equation presented earlier in this section, but with θ now reflecting the parameter of the Fréchet distribution. In the Eaton–Kortum model, these gains are due to comparative advantage—that is, from the ability to import from countries with the lowest relative costs. The close similarity of the formula used to measure the gains from trade in different models—as pointed out by Arkolakis, Costinot, and Rodríguez-Clare (2012)—is what makes this formula so important for international trade research.

Pro-competitive Effects of Trade

The insight that import competition might create gains from trade by reducing the markups charged by firms was emphasized in the early monopolistic competition literature by Krugman (1979), but the broad idea that trade can be a source of pro-competitive gains predates that literature. In Bhagwati's (1965) well-known analysis of import tariffs versus quotas, he emphasized that import quotas would inhibit the competitive pressure from imports, because the quota fixes the import quantity regardless of the price charged by domestic firms. As a result, a home monopoly would raise its price more in the presence of a quota than for an "equivalent" tariff (that is, a tariff that results in the same quantity of imports). Harris (1985) and Krishna (1989) show a similar result, but in an oligopoly rather than a monopoly setting. The idea that trade barriers—especially quotas—can lead domestic firms to exercise their market power has been shown empirically for many countries: see Levinsohn (1993) for Turkey; Harrison (1994) for the Ivory coast; Krishna and Mitra (1998) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for India; Kim (2000) for Korea; Bottasso and Sembenelli (2001) for Italy; Konings, Cayseele, and Warzynski (2005) for Bulgaria and Romania; and Badinger (2007b) for other European countries.

To think about how trade can affect markups, we return to the discussion of trade in sugar cookies between the United States and Denmark (in the spirit of Michael Pollan's comments earlier). Suppose we take the extreme assumption that there is no difference between American and Danish sugar cookies, so that there are no gains whatsoever from product variety. We also ignore differences in productivity between firms in the two countries. In this setting, the only potential source of gains from trade is the pro-competitive reduction in prices as the firms from each country enter the other market. Against these gains, we must count the resources spent in transporting the cookies between the countries.

Brander and Krugman (1983) call this a situation of intraindustry trade in homogeneous products. They argue that with a *fixed* number of firms, moving from autarky to free trade in this context has an ambiguous effect on global welfare. To grasp the intuition behind this result, suppose that there are two countries with a single firm in each that have identical production costs. Under autarky, the firms sell at the (same) monopoly price in their respective markets. Provided that shipping costs are less than the autarky markup, then with trade the firms will "cross-haul" into each other's market, because for the first unit sold abroad the marginal revenue equals the price, which exceeds the marginal cost inclusive of shipping costs. That trade will reduce the price charged in each market, with a resulting consumer gain. But from a social welfare perspective, against that gain we must count the reduction in the profits for each firm due to competition from abroad and also due to the shipping costs. That loss in profits can be greater or less than the consumer gain, so it follows that the *social* gain can be positive or negative.

With free entry of identical firms and a homogenous product in both countries, however, Brander and Krugman (1983) argue that global welfare necessarily improves due to free trade. In this case, industry profits are zero both before and after trade, so any reduction in consumer prices becomes a social gain: the source of that gain is the exit of firms and the resulting savings of their fixed costs, which balances against the loss of resources in shipping. Welfare rises going from autarky to free trade if and only if the market price falls, which occurs whenever the firms cross-haul into the foreign market. This is an example of pro-competitive gain that can occur despite the fact that trade is intrinsically wasteful, due to trading a homogeneous product in the presence of shipping costs.

This social gain from trade depends, however, on the assumption that the transportation charges are priced at their true, social value. If gasoline and other fuels are priced too low, without fully reflecting their environmental costs, then the resources devoted to transportation could lead to a social loss. The social costs of international and intranational transportation have been analyzed empirically by Shapiro (2016), who compares the social cost of CO_2 emissions from international trade with the accompanying gains from trade. Shapiro measures the gains from product variety, so these are comparable to the gains we reported in Figure 1. For the world, he computes gains from trade of $5.5 trillion, or 10 percent of world GDP. By his computation, the added CO_2 emissions due to international air and sea shipments represent 5 percent of global CO_2 emissions. He adopts a social external cost of $29 per ton to these CO_2 emissions, in which case the global cost of the international transportation equals $34 billion, or 0.06 percent of world GDP. In other words, the gains from trade are more than 100 times greater than the social cost of these emissions, and the aggregate social gains from international trade vastly exceed the environmental externality from the international shipping.

Shapiro (2016) also analyzes the impact of a global tax on emissions of carbon dioxide from air and sea shipping. He finds that implementing such a tax would raise global welfare, but in fact, would raise country welfare only for wealthy countries and decrease it for poor countries. That result occurs because poor countries

depend more on exporting products with low value/weight ratios, such as unprocessed resources. Thus, despite the fact that the poorer countries are harmed disproportionately from global climate change—because they depend on agriculture and tend to be located in the equatorial region—they would also be harmed by policies to tax the carbon used in international transportation.

Let us return to the theoretical discussion of international trade and markups. The model of Brander and Krugman (1983) assumes Cournot competition between firms (which means that each firm treats the other firms' quantities as given). This paper is an example of the short-lived literature on "strategic trade policy" (for example, Brander and Spencer 1984, 1985). Trade researchers quickly moved away from oligopoly, however, because it was felt that the results of that literature were too sensitive to the form of conduct between firms (like whether Bertrand or Cournot competition was being assumed) to be reliable for economic policy (Eaton and Grossman 1986).

Thus, trade economists have returned to the monopolistic competition framework but without assuming a constant elasticity in order to analyze markups that can change. One approach is to use the demand curve that arises from a translog expenditure function for the consumer. To build some intuition about this demand curve, consider a linear demand curve and a curved constant-elasticity demand curve, which are tangent to each other at the point of consumption. Now consider a family of demand curves all of which are convex (they lie above the linear demand curve) but have finite reservation prices (they lie below the constant-elasticity demand curve), with all of the demand curves tangent at this same consumption point. One of these intermediate demand curves will be based on the translog expenditure function. More specifically, with a single good it can be shown that the area under the translog demand curve is *one half as large* as under the constant-elasticity demand curve (Feenstra and Shiells 1997, fn. 27).

Translog preferences allow markups to vary systematically with the elasticity of demand. Specifically, markups fall when the market shares of domestic firms are reduced due to import competition. Industry profits remain at zero, however, due to the exit of some firms, which leads to savings in fixed costs. It follows that the entire reduction in consumer prices is a social gain for the economy. In Feenstra and Weinstein (2017), my coauthor and I adopt a translog expenditure function to measure the gains from new import varieties for the United States over 1992–2005, while also incorporating the losses from reduced domestic varieties along with changes in mark-ups. We find average growth of gains from trade of 0.85 percent of GDP during 1992–2005, with about half of that amount (0.44 percentage points) due to product variety and the rest being the pro-competitive gains due to reduced markups. *Adding together* the gains from variety and the pro-competitive gains from trade for the United States, we find about the same *total* gains in the translog case as the pure variety gains under constant-elasticity-of-substitution preferences as measured by Broda and Weinstein (2006).[8]

[8] To be specific, Broda and Weinstein find US gains from increased variety of 2.6 percent of GDP, which are the cumulative gains from new import varieties in the final year of their 1972–2001 period. If we instead

In Feenstra and Weinstein (2017), we do not attempt to measure the productivity gains coming from creative destruction. However, in Feenstra (2018), that task is undertaken, using translog preferences and a *truncated* Pareto distribution for firm productivities, where the truncation means that the most-productive firm has a productivity that is bounded above. These assumptions allow all three sources of gains from trade to operate: variety gains, pro-competitive gains, and productivity gains. In this case, the *total* gains from trade are larger than when only variety and pro-competitive gains are included: the total annual average gains for the United States are about 1.1 percent of GDP, with roughly one-quarter from productivity gains (as an upper-bound), with the remaining three-quarters evenly divided between variety and pro-competitive gains.

Those total annual average gains of 1.1 percent of GDP for the United States are *less than* the gains that would be expected in theory from a model that assumed an *untruncated* Pareto distribution for firms' productivities. But in that case, with no upper bound to the highest-productivity firms, these *extremely* productive firms contribute the most to the gains from trade. Indeed, without any upper bound to firms' productivities, it turns out that the pro-competitive (and product variety) gains no longer operate. This surprising result comes about because even with the markups of domestic firms falling as trade costs fall, the markups of foreign exporting firms increase, so the overall distribution of markups (and product variety) is not affected. So with the untruncated Pareto distribution, the *entire* gains from trade come from creative destruction. Essentially, the earlier formula showing social gains from the creative destruction process becomes an upper bound to the total gains from trade, which applies even when all three sources of gains operate (as with truncated Pareto and translog preferences).[9]

Beyond Iceberg Trade Costs

The discussion in the previous sections has mainly considered "iceberg" costs of trade, but in reality, trade costs take many forms: tariffs, quotas, tariff-rate quotas that apply the tariff only to imports exceeding a certain quota level, the threat of tariffs as in dumping cases, export taxes and subsidies, and other policy instruments. These policies differ most obviously from iceberg trade costs in that iceberg costs are paid in terms of a firm's *own output*, which hypothetically "melts" along the way to the destination country. An increase in iceberg costs leads to an increase in the cost inclusive of shipping and therefore in the price charged abroad. An increase in this price reduces the quantity demanded and lowers the

consider *average annual gains* relative to GDP, they are 0.8 percent during 1990–2001, which is close to the 0.85 percent average annual gains over 1992–2005 obtained in Feenstra and Weinstein (2017).

[9] In Feenstra (2018), I obtain this upper bound theoretically for a wide class of homothetic preferences that includes the translog as a special case. A similar upper bound also occurs in Bertoletti, Etro, and Simonovska (2016), which examines certain non-homothetic preferences, and in Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (forthcoming), for estimated non-homothetic preferences.

profits earned by the exporting firm. But unlike a tariff, iceberg costs also *generate demand* for the firm's own product, since more of the product is shipped than arrives in the destination market (with the remainder "melting" along the way). Because these costs are literally used for shipping, they do not lower the firm's profits as much as would an import tariff of the same magnitude. That distinction is enough to make a difference between iceberg trade costs and tariffs.[10] In addition, tariff revenue might be redistributed back to consumers, rather than melting away in transit, which is another potential difference between iceberg costs and tariffs.

Thus, we conclude that it is important to keep track of the tariff revenue (or its equivalent) in the economy: who is paying it, who receives it, and so on. This idea is not new. It has received substantial prior attention in the analysis of import quotas, where instead of generating tariff revenue, an import quota will generate *rents* as the difference between the selling price in the quota-controlled market and under free trade. In the "voluntary export restraints" used to restrict US imports of Japanese automobiles in the 1980s, for example, the rents exceeded $1,000 per imported car or $2 billion annually in 1983–84 (Feenstra 1988). These rents were earned by the Japanese exporting firms, leading to a substantial increase in the stock market value of these firms when the quota was announced (Ries 1993). In addition, European automakers reacted to the voluntary export restraints by increasing their prices in the US market, as we would expect under oligopoly pricing, which cost US consumers an additional $3.4 billion (Dinopoulos and Kreinin 1988). These welfare costs to the United States are far in excess of the deadweight losses associated with import quotas in automobiles, and are outside the scope of the formula for social welfare gains from trade presented earlier.

A similar lesson applies from more recent policy actions, such as the tariff applied to US imports of tires from China by President Obama during 2009–2012. Because this tariff applied only to imports from China, other Asian countries (and also Mexico) exporting to the United States were able to earn rents due to the higher price of tires in the United States. These rents captured by foreign producers amounted to $0.8 billion annually (Hufbauer and Lowry 2012), which far exceeds the wages earned in the additional jobs in the United States. This situation also applies to the tariff on US imports of solar cells, where some emerging economies that export are exempted from the tariff and can therefore collect rents from higher prices in the United States. Interestingly, one commissioner of the US International Trade Commission recommended instead that the United States adopt a quota on imports at the 2016 level (increasing each year thereafter), with quota licenses *auctioned* by the United States, which would have allowed the rents to be collected

[10]We are assuming that the tariff is charged on imports valued at their price inclusive of the markup over marginal costs. In the alternative case where the tariff is charged only on the variable cost of the import, it has much the same effect as iceberg costs; in that case, the two instruments differ only by the redistribution of tariff revenue.

as auction revenue.[11] Despite the benefit of collecting the revenue, however, a quota of this type would have allowed US firms to exercise their market power and raise their prices, much as occurred with the voluntary export restraints on U.S. imports of cars from Japan in the 1980s.

A different example of how certain trade policies can generate costs above and beyond those in the earlier formula for social welfare gains comes from the application of antidumping policy. Every year, dozens of antidumping cases are filed by firms in the United States and in other countries, alleging that foreign firms are selling at "less than fair market value," which can lead to the imposition of antidumping tariffs. To avoid such tariffs, foreign firms may raise their prices so that they are less likely to be found guilty of selling at "less than fair market value." That price increase leads to a welfare cost in the importing country *even when the antidumping tariff is not actually applied*. Because foreign firms earn these higher prices, and there is no offsetting tariff revenue, these antidumping actions have a substantial welfare cost.

Staiger and Wolak (1994) find evidence of such price increases in the absence of antidumping duties being applied. Ruhl (2014) quantifies the magnitude of the resulting welfare loss for the United States in 1992. In a calibrated model, he finds that antidumping policy reduces US consumption (and therefore welfare) by 3.2 percent. If we apply Ruhl's estimate of a 3.2 percent loss in consumption to US imports of $533 billion in 1992, then we obtain a loss of $17 billion. This total is four times greater than the welfare cost of US antidumping and countervailing duties from Gallaway, Blonigen, and Flynn (1999), estimated at $4 billion in 1993—but this estimate considers only the *actual* antidumping and countervailing duties in effect in 1993, whereas Ruhl allows firms to increase their price to lower the *probability* of tariffs being applied. Ruhl also allows for firm heterogeneity, which is important because antidumping duties are higher for the most-productive (and hence lowest-price) firms.

Turning from these specific examples of trade policies back to the theory, what is the impact of import tariffs in the Melitz–Chaney model? To add a dose of realism to that framework, suppose that in addition to the differentiated-goods sector, we add a perfectly competitive sector that is nontraded. Assume that trade is balanced. Then applying an import tariff to the differentiated goods sector can be expected to have the same impact as applying an export tax to that sector: both will drive down levels of trade, and with the assumption of balanced trade, both will result in reduced imports and exports. This equivalence of import tariffs and export taxes in general equilibrium is called "Lerner symmetry."

In this general framework, in Caliendo, Feenstra, Romalis, and Taylor (2017), we find that starting from a zero-tariff equilibrium, an import tariff leads to reduced entry into the differentiated goods sector *at home*. That is, an import tariff or an

[11] See the statements of the commissioners in this United States International Trade Commission document dated October 31, 2017: https://www.usitc.gov/press_room/documents/solar201_remedy_commissionerstatements.pdf.

export tax equivalently leads to the exit of home firms, which creates a social cost that is above and beyond the cost of reducing trade based on the earlier formula on gains from trade. Because of this distortionary effect that tariffs have on entry, it even turns out that the optimal tariff for countries can be *negative* instead of positive.[12] In the quantitative model, it turns out the fully *one-quarter* of the countries in the world have negative optimal tariffs, with most of these being very remote economies with little trade as a result (for example, small island economies), but with some examples of strong international linkages (for example, France).

The results in Caliendo, Feenstra, Romalis, and Taylor (2017) imply that ongoing efforts to liberalize trade under the World Trade Organization are important because it is quite possible that countries have first-order gains from reducing tariffs even when tariffs are small (indeed, that is the case when the optimal tariff is negative). This is in contrast to a competitive model, where the worldwide gains from reducing tariffs that are small will also be very small. The idea that the gains from tariff reductions are not necessarily small, especially for highly linked and for very remote countries, argues for renewed attention to multiregional and multilateral negotiations to liberalize trade.

Conclusions

Recent theories of international trade have allowed for sources of the gains from trade that go well beyond conventional comparative advantage, including gains from increased variety; a shift toward firms with higher productivity; and lower mark-ups. In the context of the prominent models of these effects, the social gains can be presented in a straightforward formula. For a given sector, the implied gains depend on the *share of home spending on its domestic goods,* which together with a parameter reflecting the taste for variety or the Pareto productivity distribution of firms, are "sufficient statistics" for the gains from trade, to use the term popularized by Arkolakis, Costinot, and Rodríguez-Clare (2012). This formula tells us that we can expect welfare gains in proportion to the rise in trade shares (or more precisely, to the fall in home shares), regardless of the underlying model of trade.

The gains from trade that we have discussed are "overall" gains in the sense that they reflect increased purchasing power for the economy overall without consideration of how those gains are distributed across individuals, some of whom can be expected to lose. We conclude our paper with a further discussion of these distributional concerns (which are also discussed in the papers by Fort, Pierce, and Schott, and by Rodrik, in this volume).

Krugman (1981) was optimistic that trade in product varieties had the potential to *offset* certain losses from trade. He started with the logic of the Stolper–Samuelson theorem, where under perfect competition the factor of production used intensively

[12] However, if the entry distortion is *offset* using some domestic policies, then the optimal tariff in the Melitz–Chaney model would remain positive, as shown by Costinot, Rodríguez-Clare, and Werning (2016).

in export production will gain but the factor of production used intensively in import production must lose. Krugman argued that this result could be overturned under monopolistic competition, since in that case, the gains from product variety could potentially benefit all factors of production. For example, think of workers benefiting from the low prices and product variety offered at Walmart even if their nominal wages suffer due to import competition. Is there any evidence to support the idea that they might experience an overall gain in real terms?

Fajgelbaum and Khandelwal (2016) begin to address this question by examining how consumers *of differing incomes* benefit from import variety. They find that poor consumers tend to gain the most from imported goods, because they concentrate their spending in sectors that have more trade. That conclusion holds especially for higher-income countries that export high–income-elastic goods (luxuries) and import low–income-elastic goods (necessities). As a result, the poor in high-income countries have the greatest consumption gains from trade as a share of income. Fajgelbaum and Khandelwal do not incorporate the differing wages earned by various individuals, however. That step is taken in the recent work by Borusyak and Jaravel (2018) for the United States, who distinguish the expenditures and earnings of different educational groups. They find that a reduction of all import and export trade barriers generates a modest increase in inequality between education groups, which is primarily due to the earnings (that is, wages) channel. Those with and without a college degree both gain on average from such a reduction in trade barriers, as in Krugman's original claim, but the college-educated gain by more.

This optimistic outcome due to the product variety gains from trade may well apply to the gains from reduced markups, too, which can be expected to benefit many consumers. In contrast, creative destruction will most likely negatively affect some workers much more than others. We have already mentioned the sorting of workers across firms, whereby workers with mid-level skills are matched to firms with inferior technologies, and these workers and firms face the greatest import competition and experience losses. Indeed, we have suggested that the US tariff on solar cells is intended to protect firms and workers in that industry, but it does so at the cost of disrupting the industry engaged in the installation of solar panels. More work is needed to determine how the alternative sources of gains from trade discussed here interact with their distributional consequences, so that we know whether these gains are widely shared, and if there are policies that can help to achieve that goal.

■ *The author thanks Eric Bartelsman and Gordon Hanson for helpful comments.*

References

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. 2017. "Missing Growth from Creative Destruction." Federal Reserve Bank of San Francisco Working Paper 2017-04.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Arkolakis, Costas, Arnaud Costinot, David Donaldson, and Andrés Rodríguez-Clare. Forthcoming. "The Elusive Pro-Competitive Effects of Trade." *Review of Economic Studies*.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102(1): 94–130.

Armington, Paul S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16(1): 159–78.

Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality among the 'Other 99 Percent.'" *Science,* May, 23, 344(6186): 843–51.

Badinger, Harald. 2007a. "Market Size, Trade, Competition and Productivity: Evidence from OECD Manufacturing Industries." *Applied Economics* 39(17): 2143–57.

Badinger, Harald. 2007b. "Has the EU's Single Market Programme Fostered Competition? Testing for a Decrease in Markup Ratios in EU Industries." *Oxford Bulletin of Economics and Statistics* 69(4): 497–519.

Badinger, Harald. 2008. "Trade Policy and Productivity." *European Economic Review* 52(5): 867–91.

Balistreri, Edward J., Russell H. Hillberry, and Thomas F. Rutherford. 2011. "Structural Estimation and Solution of International Trade Models with Heterogeneous Firms." *Journal of International Economics* 83(2): 95–108.

Bertoletti, Paolo, Federico Etro, and Ina Simonovska. 2016. "International Trade with Indirect Additivity." NBER Working Paper 21984.

Bhagwati, Jagdish N. 1965. "On the Equivalence of Tariffs and Quotas." In *Trade Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler,* edited by Robert E. Baldwin, et al. Chicago: Rand-McNally, 53–67.

Borusyak, Kirill, and Xavier Jaravel. 2018. "The Distributional Effect of Trade: Theory and Evidence from the United States." Available at: https://scholar.harvard.edu/borusyak/publications/distributional-effects-trade-theory-and-evidence-united-states-job-market.

Bottasso, Anna, and Alessandro Sembenelli. 2001. "Market Power, Productivity and the EU Single Market Program: Evidence from a Panel of Italian Firms." *European Economic Review* 45(1): 167–86.

Brander, James A., and Paul Krugman. 1983. "A 'Reciprocal Dumping' Model of International Trade." *Journal of International Economics* 15(3–4): 313–21.

Brander, James A., and Barbara J. Spencer. 1984. "Trade Warfare: Tariffs and Cartels." *Journal of International Economics* 16(3–4): 227–42.

Brander, James A., and Barbara J. Spencer. 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 18(1–2): 83–100.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121(2): 541–85.

Caliendo, Lorenzo, Robert C. Feenstra, John Romalis, and Alan M. Taylor. 2017. "Tariff Reductions, Entry, and Welfare: Theory and Evidence for the Last Two Decades." NBER Working Paper 21768.

Chaney, Thomas. 2008. "Distorted Gravity: the Intensive and Extensive Margins of International Trade." *American Economic Review* 98(4): 1707–21.

Chaney, Thomas, and Ralph Ossa. 2013. "Market Size, Division of Labor, and Firm Productivity." *Journal of International Economics* 90(1): 177–80.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–262. New York: Elsevier.

Costinot, Arnaud, Andrés Rodríguez-Clare, and Iván Werning. 2016. "Micro to Macro: Optimal Trade Policy with Firm Heterogeneity." NBER Working Paper 21989.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, Nina Pavcnik. 2016. "Prices, Markups and Trade Reform." *Econometrica* 84(2): 445–510.

Dinopoulos, Elias, and Mordechai E. Kreinin. 1988. "Effects on the U.S.–Japan Auto VER on European Prices and on U.S. Welfare." *Review of Economics and Statistics* 70(3): 484–91.

Eaton, Jonathan, and Gene M. Grossman. 1986. "Optimal Trade and Industrial Policy under Oligopoly." *Quarterly Journal of Economics* 101(2): 383–406.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography and Trade." *Econometrica*

70(5): 1741–80.

Fajgelbaum, Pablo D., and Amit K. Khandelwal. 2016. "Measuring the Unequal Gains from Trade." *Quarterly Journal of Economics* 131(3): 1113–80.

Feenstra, Robert C. 1988. "Quality Change Under Trade Restraints in Japanese Autos." *Quarterly Journal of Economics* 103(1): 131–46.

Feenstra, Robert C. 1994. "New Product Varieties and the Measurement of International Prices." *American Economic Review* 84(1): 157–77.

Feenstra, Robert C. 2010. "Measuring the Gains from Trade under Monopolistic Competition." *Canadian Journal of Economics* 43(1): 1–28.

Feenstra, Robert C. 2015. *Advanced International Trade: Theory and Evidence*, 2nd edition. Princeton University Press.

Feenstra, Robert C. 2018. "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity." *Journal of International Economics* 110: 16–17.

Feenstra, Robert C., and Clinton R. Shiells. 1997. "Bias in U.S. Import Prices and Demand." Chap. 6 in, *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon. NBER and University of Chicago Press.

Feenstra, Robert C., and David E. Weinstein. 2017. "Globalization, Markups, and U.S. Welfare." *Journal of Political Economy* 125(4): 1040–74.

Gallaway, Michael P., Bruce A. Blonigen, and Joseph E. Flynn. 1999. "Welfare Costs of the US Antidumping and Countervailing Duty Laws." *Journal of International Economics* 49(2): 211–44.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow. 2016. "How Destructive is Innovation?" NBER Working Paper 22953.

Harris, Richard. 1985. "Why Voluntary Export Restraints Are 'Voluntary.'" *Canadian Journal of Economics* 18(4): 799–809.

Harrison, Ann E. 1994. "Productivity, Imperfect Competition and Trade Reform: Theory and Evidence." *Journal of International Economics* 36(1–2): 53–73.

Hausman, Jerry A. 1997. "Valuation of New Goods under Perfect and Imperfect Competition." Chap. 5 in *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon. NBER Studies in Income and Wealth. University of Chicago Press.

Hausman, Jerry A. 1999. "Cellular Telephone, New Products and the CPI." *Journal of Business and Economic Statistics* 17(2): 188–94.

Helpman, Elhanan. Forthcoming. *Globalization and Inequality*. Princeton University Press.

Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge: MIT Press.

Hicks, John R. 1940. "The Valuation of Social Income." *Economica* 7(26): 105–24.

Hsieh, Chang-Tai, Nicholas Li, Ralph Ossa, and Mu-Jeung Yang. 2018. "Accounting for the New Gains from Trade Liberalization." NBER Working Paper 22069, revised.

Hufbauer, Gary Clyde, and Sean Lowry. 2012. "US Tire Tariffs: Saving Few Jobs at High Cost." Peterson Institute for International Economics, Policy Brief no. PB12-9, April. https://piie.com/publications/pb/pb12-9.pdf.

Kim, Euysung. 2000. "Trade Liberalization and Productivity Growth in Korean Manufacturing Industries: Price Protection, Market Power, and Scale Efficiency." *Journal of Development Economics* 62(1): 55–83.

Konings, Jozef, Patrick Van Cayseele, and Frederic Warzynski. 2005. "The Effects of Privatization and Competitive Pressure on Firms' Price-Cost Margins: Micro Evidence from Emerging Economies." *Review of Economics and Statistics* 87(1): 124–34.

Krishna, Kala. 1989. "Trade Restrictions as Facilitating Practices." *Journal of International Economics* 26(3–4): 251–70.

Krishna, Pravin, and Devashish Mitra. 1998. "Trade Liberalization, Market Discipline and Productivity Growth: New Evidence from India." *Journal of International Economics* 56(2): 447–62.

Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9(4): 469–79.

Krugman, Paul R. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70(5): 950–59.

Krugman, Paul R. 1981. "Intraindustry Specialization and the Gains from Trade." *Journal of Political Economy* 89(5): 959–73.

Levinsohn, James. 1993. "Testing the Imports-as-Market-Discipline Hypothesis." *Journal of International Economics* 35(1–2): 1–22.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–1725.

Melitz, Marc J., and Stephen J. Redding. 2014. "Missing Gains from Trade?" *American Economic Review* 104(5): 317–21.

Melitz, Marc J., and Stephen J. Redding. 2015. "New Trade Models, New Welfare Implications." *American Economic Review* 105(3): 1105–46.

Melitz, Marc J., and Daniel Trefler. 2012. "Gains from Trade When Firms Matter." *Journal of Economic Perspectives* 26(2): 91–118.

Ossa, Ralph. 2015. "Why Trade Matters After

All." *Journal of International Economics* 97(2): 266–77.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*. London: John Murray. Available at: http://www.econlib.org/library/Ricardo/ricP.html.

Ries, John C. 1993. "Windfall Profits and Vertical Relationships: Who Gained in the Japanese Auto Industry from VERs?" *Journal of Industrial Economics* 41(3): 259–76.

Ruhl, Kim J. 2014. "The Aggregate Impact of Antidumping Policies." Unpublished paper, Pennsylvania State University.

Samuelson, Paul A. 1952. "The Transfer Problem and Transport Costs: The Terms of Trade When Impediments Are Absent." *Economic Journal* 62(246): 278–304.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism, and Democracy*. Harper & Brothers.

Shapiro, Joseph S. 2016. "Trade Costs, CO_2, and the Environment." *American Economic Journal: Economic Policy* 8(4): 220–54.

Simonovska, Ina, and Michael E. Waugh. 2014. "Trade Models, Trade Elasticities, and the Gains from Trade." NBER Working Paper 20495.

Staiger, Robert W., and Frank A. Wolak. 1994. "Measuring Industry-Specific Protection: Antidumping in the United States." *Brooking Papers on Economic Activity: Microeconomics,* no. 1, pp. 51–118.

Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 112(6): 1181–1222.

Syverson, Chad. 2004b. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86(2): 534–50.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–65.

Trefler, Daniel. 2004. "The Long and Short of the Canada–U.S. Free Trade Agreement." *American Economic Review* 94(4): 870–95.

Tybout, James, Jaime de Melo, and Vittorio Corbo. 1991. "The Effects of Trade Reform on Scale and Technical Efficiency: New Evidence from Chile." *Journal of International Economics* 31(3–4): 231–50.

Tybout, James R., and M. Daniel Westbrook. 1995. "Trade Liberalization and Dimensions of Efficiency Change in Mexican Manufacturing Industries." *Journal of International Economics* 39(1–2): 53–78.

Yeaple, Stephen Ross. 2005. "A Simple Model of Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics* 65(1): 1–20.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 47–72

New Perspectives on the Decline of US Manufacturing Employment

Teresa C. Fort, Justin R. Pierce, and Peter K. Schott

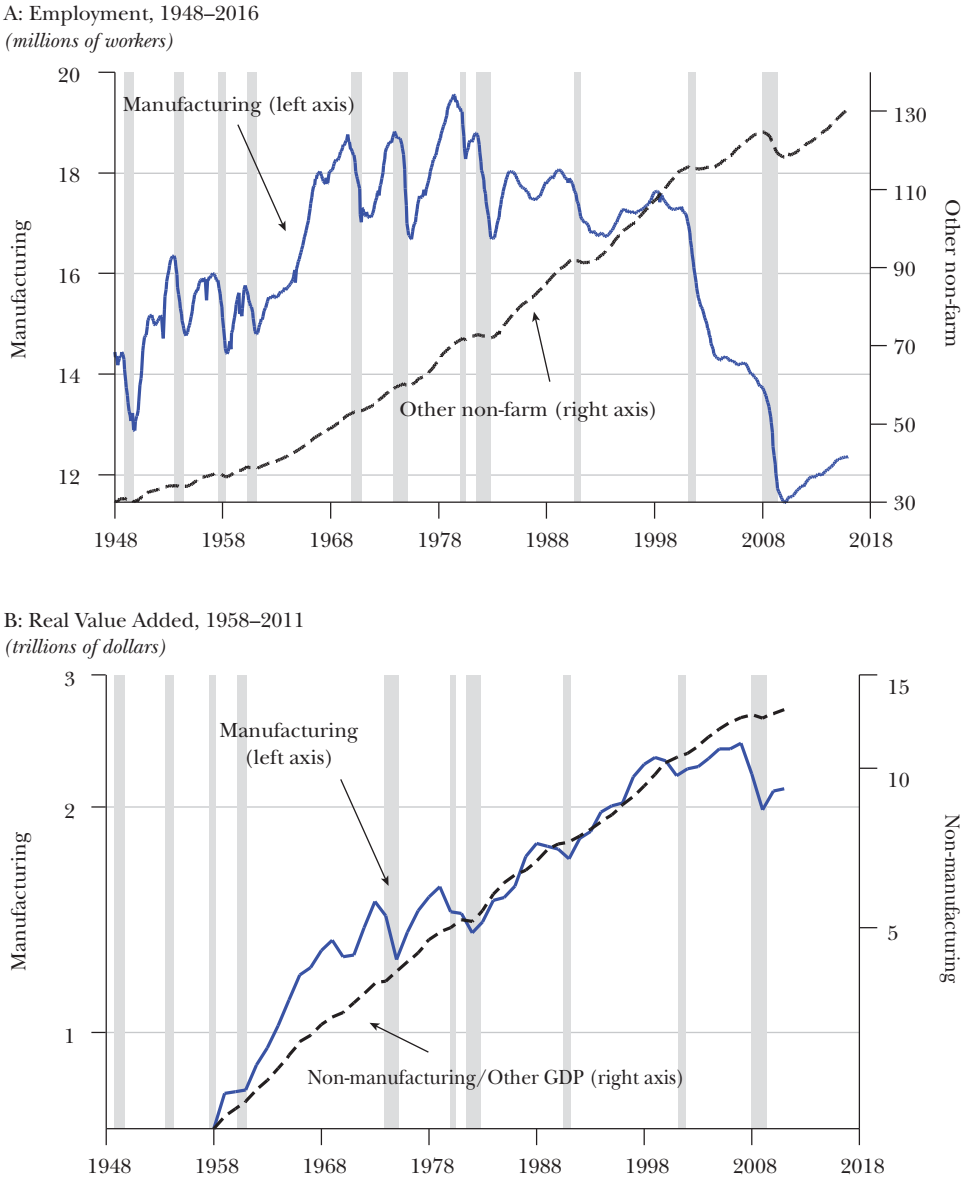
US manufacturing since World War II exhibits three notable trends, illustrated in the two panels of Figure 1. First, manufacturing employment has diverged from non-manufacturing employment, as shown on different axes in Figure 1A. While both series moved upward until the late 1970s, manufacturing employment then begins to decline, even as other non-farm employment continues a steady rise. As a result, there is a continual decline in manufacturing employment's share of total US non-farm employment, from 32 percent in 1948 to 8 percent in 2017. Second, while US manufacturing employment fell just 12 percent over the 21 years between the post-war peak in 1979 and 2000, it then dropped by more than twice as much—25 percent—from 2000 to 2012. Third, despite the relative flatness and subsequent sharp decline in US manufacturing employment, the bottom panel of Figure 1 shows a steady rise in manufacturing real value added at more or less the same rate as non-manufacturing GDP over the same period, at least between the late 1970s and the Great Recession. The combination of relatively steady and then declining employment, and rising output, indicates that, over the long term, labor productivity has risen faster in the manufacturing sector than in the broader economy.

■ *Teresa C. Fort is Associate Professor of Business Administration, Tuck School of Business at Dartmouth, Hanover, New Hampshire. Justin R. Pierce is a Principal Economist, Board of Governors of the Federal Reserve System, Washington, DC. Peter K. Schott is Juan Trippe Professor of International Economics, Yale School of Management, New Haven, Connecticut. Fort is a Faculty Research Fellow and Schott is a Research Associate, both at the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are teresa.fort@tuck.dartmouth.edu, justin.r.pierce@frb.gov, and peter.schott@yale.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.47

doi=10.1257/jep.32.2.47

Figure 1
US Employment and Value Added within and outside Manufacturing



Source: Monthly employment data are from the US Bureau of Labor Statistics. Annual manufacturing real value added data are from NBER-CES Manufacturing Industry Database (Becker, Gray, and Marakov 2013). Annual real GDP data are from US Bureau of Economic Analysis. Non-manufacturing value added is real GDP less manufacturing real value added.
Note: Shading corresponds to NBER-dated recessions.

For a variety of reasons, including the perception that workers in manufacturing receive comparatively high wages conditional on education (Langdon and Lehrman 2012; Ebenstein, Harrison, McMillan, and Phillips 2014), these trends have stirred intense discussion among both policymakers and academics. This debate can be summarized broadly as a dispute between views that emphasize the relative importance of trade versus technology. The trade-based explanation contends that import competition has reduced US manufacturing employment by inducing labor-intensive, low-labor-productivity industries to move abroad. The technology view argues that the decline in manufacturing employment stems from innovations in production techniques, such as automation, that have dramatically increased output per worker. If consumers spend a constant share of their expenditure on manufactured goods, then an increase in labor productivity means fewer workers are needed to meet demand for those goods.

Discussions about the decline in US manufacturing employment often culminate in a request to decompose the decrease into the part that is due to trade and the part that is due to technology. Our view is that providing a definitive accounting of the amount of employment change attributable to either factor is extraordinarily difficult for two reasons. First, identifying the numerous changes in tariff and nontariff barriers that have occurred over the last few decades, let alone the wide range of technologies that have been adopted, is a daunting task.[1] Second and more importantly, even if one could identify all of these changes, it is difficult to see how their intertwined impacts on employment could be teased apart. As an example, consider an anecdote from a recent *Wall Street Journal* article (reported in Michaels 2017), which takes place around the time of an important US trade liberalization with China discussed below:

> When Drew Greenblatt bought Marlin Steel Wire Products LLC, a small Baltimore maker of wire baskets for bagel shops, he knew nothing about robotics. That was 1998, and workers made products manually using 1950s equipment. ... Pushed near insolvency by Chinese competition in 2001, he started investing in automation. Since then, Marlin has spent $5.5 million on modern equipment. Its revenue, staff and wages have surged and it now exports to China and Mexico.

Are changes in Marlin's employment and output driven by the availability of robots or by increased Chinese competition? What about employment and output at other producers of steel wire products, who face increased competition from both China

[1] For example, even while ad valorem tariff rates have trended downward over time, and regional trade agreements have proliferated, implementation and repeal of contingent protection measures like antidumping and countervailing duties remains frequent and widespread (Bown 2016). These temporary barriers have been linked to relative declines in physical productivity and increased prices among protected manufacturing plants (Pierce 2011). Identification of the numerous technological innovations introduced during this period, including computerization, electronic communication, computer-aided design and manufacturing, just-in-time inventory management, and enterprise resource planning, is similarly difficult.

and from Marlin? These questions are even more difficult to answer if the availability of robots is itself influenced by trade liberalization—for example, by robot manufacturers' ability to source intermediate inputs from China.

In this paper, we provide a brief overview of recent efforts to answer such questions before turning to relatively unexplored dimensions of US microdata for further input. These data allow us to examine changes in US manufacturing employment across industries, firms, and regions, and thereby offer four new perspectives on how US manufacturing has evolved over the last several decades. We find that while employment changes along these dimensions provide support for both trade- and technology-based explanations, they also highlight the difficulties of cleanly separating one force from the other. Toward that end, we discuss how further analysis of the data we use might yield sharper insights.

Our first perspective examines how the overall growth of US manufacturing employment, and value added, varies by sector. We find that some sectors—such as transportation equipment—exhibit increases in output even as employment is falling, a potentially clear indication of technology adoption. On the other hand, it is not hard to find examples of sectors, such as apparel, characterized by simultaneous increases in import penetration and reductions in both employment and output. Furthermore, the set of sectors experiencing declines in both employment and output increases after 2000.

Our second perspective analyzes employment loss along firm and establishment margins of adjustment. One of our more striking findings—given conventional expectations about how creative destruction due to trade and technology likely manifest themselves—is that net firm death accounts for just 25 percent of the overall decline in US manufacturing employment between 1977 and 2012. On the other hand, we find a large role for net plant exit within incumbent firms, perhaps because adopting new technologies or adapting to import competition entails high fixed costs that continuing firms are better able to absorb and which are easier to implement by shuttering outmoded plants and opening new ones.

Our third perspective breaks down the aggregate change in US manufacturing employment between 1977 and 2012 along regional margins of adjustment. We find a steady reallocation of manufacturing employment away from the north and east towards the south and west until 2000, when employment starts falling in all regions. The earlier transition may reflect "domestic offshoring," which refers to a movement from higher- to lower-wage US regions in an era before foreign offshoring was cost-effective.

Our final perspective takes a wider view of manufacturing firms by examining their non-manufacturing activities. We find that the non-manufacturing employment of manufacturing firms increases until 2000—primarily via the addition of new non-manufacturing establishments—before leveling off. About one-third of this growth is in professional services, a trend that may represent an evolution of US manufacturing firms into "neuro-facturers" that increasingly provide intellectual services rather than physical goods (Leamer 2009). Prominent examples include Pitney Bowes, which has abandoned the production of postage meters to offer logistics services; IBM, which

increasingly offers data solutions rather than mainframes; and Apple, which designs the iPhone in the United States but uses offshore contractors for assembly.

Some of the Evidence thus Far

The last three decades have witnessed dramatic changes in both trade and technology. We provide a sense of some of these changes in Figure 2, which plots US manufacturing firms' use of two specific forms of technology—computers and electronic networks—at five-year intervals from 1977 to 2012. As indicated in the figure, the share of firms purchasing computers in the noted years increases through the 1990s, with a large jump in the early 2000s. Data tracking use of electronic networks to control or coordinate shipments are available starting in 2002, and exhibit an analogous increase in adoption during the 2000s.[2]

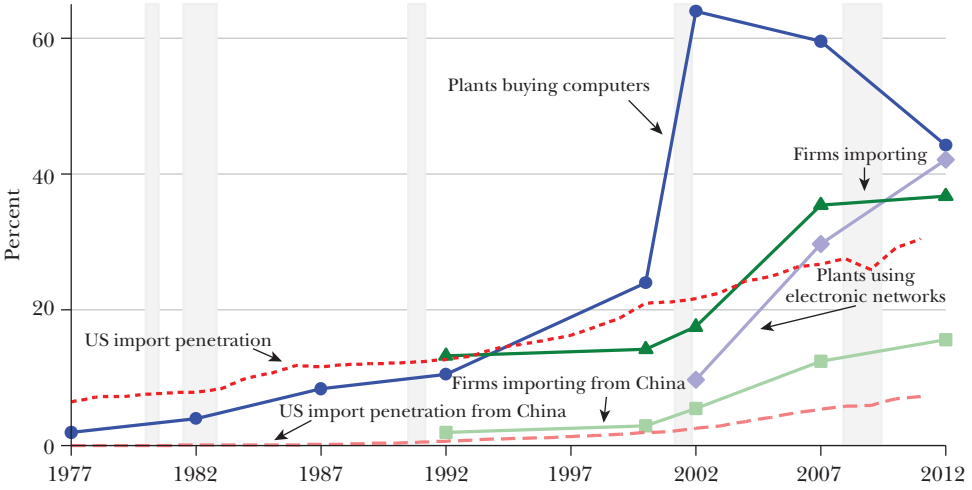
Figure 2 also reports several dimensions of trade activity. First, starting in 1992, we report the share of manufacturing firms that imports from any country as well as the share that imports from China, by census year. Here, as with our indicators of technology use, we see increases in the early 2000s. Second, we display annual measures of import penetration and import penetration from China. These series are defined as manufacturing imports (or manufacturing imports from China) divided by the sum of domestic manufacturing real shipments plus manufacturing imports less manufacturing exports, all in real terms. Import penetration from all sources is rising over time, with a pronounced upward shift after the 1981 recession and relatively rapid growth during the 1990s. Chinese import penetration rises relatively slowly in the 1990s before picking up in the 2000s.[3] A key message of Figure 2 is that both technology adoption and importing, including by US producers, generally rise over the sample period, sometimes simultaneously.

Researchers have adopted several approaches to identify effects of trade "shocks" on employment. Perhaps the narrowest definition of a trade shock is a change in trade policy, such as a reduction in import tariffs that leads to increased trade flows. Broader definitions include the impact of other factors, such as transport or communication costs, or foreign capital accumulation, that alter comparative advantage and the terms of trade. A complication associated with identifying such trade shocks is that they can be induced by technology shocks; for example, a trading partner's productivity growth may be driven by its adoption of new technologies or production techniques. Examining the US steel industry, Oster (1982) shows that large US producers were relatively slow in adopting new blast-furnace technologies

[2]As discussed in Fort (2017), plants' use of electronic networks to control or coordinate shipments involves not just using the internet or other networks, but also integrating electronic communication in the production process. Computer purchase data are not available in 1997, so we supplement the Census of Manufactures data with information from the 2000 Annual Survey of Manufactures.

[3]Online Appendix Figure A.1 displays the levels of overall US imports, exports, manufacturing value added, and manufacturing absorption (value added plus imports minus exports).

Figure 2
Technology Adoption and Importing in US Manufacturing Sector, 1977–2012



Source: Data on share of firms purchasing computers and using electronic networks are from the Census of Manufactures (1977, 1982, 1987, 1992, 2002, 2007 and 2012) and the Annual Survey of Manufactures (2000). Data on firm-level importing are from the Longitudinal Firm Trade Transactions Database. Electronic Networks include the internet and electronic data interchanges (Fort 2017). Import penetration data are from the NBER-CES Manufacturing Industry Database (Becker, Gray, and Marakov 2013), Feenstra (1996), and Schott (2008).
Note: "Plants buying computers" is the percent of US manufacturing plants purchasing computers during the year. "Plants using electronic networks" is the percent of US manufacturing plants using electronic networks to control or coordinate shipments. "Firms importing" is the percent of US manufacturing firms importing from any country. "Firms importing from China" is the percent of US manufacturing firms importing from China. "US import penetration" is manufacturing imports into the United States divided by real manufacturing domestic absorption in percentage terms, where domestic absorption is the sum of US manufacturing shipments and imports less manufacturing exports. "US import penetration from China" is defined analogously but restricts the numerator to imports from China. For penetration, industries are concorded from 1972 SIC codes to 1987 SIC codes in 1995, and from 1987 SIC codes to NAICS codes in 1997. Shading corresponds to NBER-dated recessions.

during the 1970s, a factor which may have contributed to the rise in steel imports from their faster-adopting Japanese rivals.

A growing empirical literature uses specific trade liberalizations to investigate whether US manufacturing employment or wages drop disproportionately in industries with greater exposure to changes in policy. Hakobyan and McLaren (2016), for example, use industry variation in US tariff reductions due to the North American Free Trade Agreement (NAFTA) to document a negative wage effect of NAFTA on less-educated workers between 1990 and 2000. Focusing on the following decade, Pierce and Schott (2016) show that the post-2000 decline in US manufacturing employment is relatively larger for industries exposed to the granting of Permanent Normal Trade Relations to China in October 2000. This nontraditional trade liberalization eliminated the possibility of sudden, substantial spikes in US tariffs on

many Chinese imports, thereby removing a significant deterrent to greater integration of the two economies that had been in place since the 1980s.

Research into the broader set of shocks that might alter US terms of trade makes use of changes in imports to identify reallocation. These papers devote considerable effort to excluding variation in imports driven by nontrade factors, such as secular declines in demand or common technology shocks. Bernard, Jensen, and Schott (2006), for example, find that US manufacturing plant survival and employment between 1977 and 1997 are negatively associated with increasing import penetration from low-wage countries. To identify a causal effect of trade, they use changes in US import tariffs and *ad valorem* trade costs over their sample period as instruments for import penetration. Autor, Dorn, Hanson, and Song (2014) and Acemoglu, Autor, Dorn, Hanson, and Price (2015) show that workers in industries with higher growth in Chinese imports experience increased unemployment between 1992 and 2007. In these papers, Chinese import growth in other countries is used as an instrument for its growth in the United States. The identifying assumption is that Chinese exports to these other countries are driven by productivity growth in China, and not by changes in demand or technology outside of China that might also affect US manufacturing employment.

A related body of work exploits spatial variation in the distribution of manufacturing industries across the United States. Acemoglu, Autor, Dorn, Hanson, and Price (2015) show that industries with higher growth in Chinese imports experience larger declines in employment between 1992 and 2007, and Autor, Dorn, Hanson, and Song (2014) find that workers in such industries experience relative declines in cumulative earnings. Regions with higher initial shares of employment in exposed industries also exhibit relative declines in the provision of public goods (Feler and Senses forthcoming) and marriage rates (Autor, Dorn, and Hanson 2018), as well as relative increases in household debt (Barrot, Loualiche, Plosser, and Sauvagnat 2017) and crime (Che and Xu 2016). These consequences carry over to health: Pierce and Schott (2017) show that regions more exposed to US trade liberalization with China exhibit relative increases in "deaths of despair," including drug overdoses. This connection is reminiscent of the spike in mortality rates among high-tenure workers laid off from the steel industry in Pennsylvania during the 1980s (Sullivan and von Wachter 2009).

Studies like those noted above are often conducted using a difference-in-differences framework, which does not account for potential general equilibrium effects and thus complicates calculation of a trade shock's effect on the overall level of manufacturing employment (Muendler 2017). Quantitative models, often drawing on empirical evidence from such studies, do offer such estimates, as well as quantifications of the impact of trade on social welfare. Caliendo, Dvorkin, and Parro (2015), for example, argue that increased trade with China explains approximately one-quarter of the decline in US manufacturing employment between 2000 to 2007, and that the growth of trade with China over this period increased US welfare, though, like Galle, Rodríquez-Clare, and Yi (2017), they find that gains vary across regions. Handley and Limão 2017) find that trade

liberalization with China in the 2000s benefits consumers via increased imported product variety.

While changes in trade policy and increases in imports, particularly during the 2000s, have received considerable attention, other researchers interpret the long-run decline in the manufacturing employment share implicit in Figure 1 as driven by technology. Edwards and Lawrence (2013), for example, argue that the long post–World War II decline in the share of US employment in manufacturing occurs "irrespective of the changing developments in international trade flows, the size of the trade deficit, and other factors." A number of papers assess the role of particular technologies on manufacturing employment. Collard-Wexler and De Loecker (2015) describe the importance of the introduction of mini-mills in the US steel industry to subsequent gains in output and declines in employment. Acemoglu and Restrepo (2017) find that US regions with an industrial mix that pre-disposes them to adopting more industrial robots have also experienced relatively larger employment declines, at a rate of approximately five workers per robot. Similarly, Graetz and Michaels (2017) use cross-country and industry data to show that robot adoption relates to decreased work hours by middle- and especially low-skill workers.

Another strand of research aims to decompose the respective roles of trade and technology on employment and wages. Goos, Manning, and Salomons (2014) and Autor, Dorn, and Hanson (2015) argue that technological change has decreased the relative demand for routine tasks; the latter compares the results for computerization of routine tasks to increased Chinese import penetration in the United States and concludes that Chinese imports play a larger role in the decline of US manufacturing employment, especially after 2000. While this research uses careful measures to identify technology and trade, it remains susceptible to the possibility, highlighted in the anecdote presented in the introduction as well as theoretical work in this area (for example, Acemoglu 2002), that a new technology's invention or adoption may itself be in response to a trade shock. Bernard, Jensen, and Schott (2006), Khandelwal (2010), and Bernard, Redding, and Schott (2011) show that US firms respond to import competition in part by upgrading their product mix. Bloom, Draca, and Van Reenen (2016) find evidence of technology upgrading within and across European firms that were more exposed to Chinese imports. In the US context, Autor, Dorn, Hanson, Pisano, and Shu (2016) also find that Chinese import penetration affects the innovative activities of manufacturers, though they document a negative relationship.[4] Finally, interconnectedness is also found in the other direction. Fort (2017) and Steinwender (2018) show that innovations in communications technologies facilitate trade. When viewed as a whole, this research highlights the difficulties associated with clean identification of one force over another.

[4] In related research in labor economics, Clemens, Lewis, and Postel (2017) show that imposing restrictions on low-skill immigration induced adoption of more capital-intensive production techniques and shifts in product mix in the agricultural sector.

Reallocation of Employment and Value Added Across Industries

Examination of employment and output changes by industry provides useful context for the trends displayed in Figure 1, while also offering evidence in support of both trade- and technology-based explanations for the overall decline in US manufacturing employment since the late 1970s. Figure 3 displays log changes in real value added, employment, and import penetration for the 21 three-digit NAICS sectors that constitute manufacturing. Given the sharp drop in manufacturing employment after 2000 displayed in Figure 1, we provide separate decompositions for years before (Figure 3A) and after (Figure 3B) that year, ending the latter period before the Great Recession to avoid its impact. In each period, industries are sorted by their log change in real value added, from low to high.

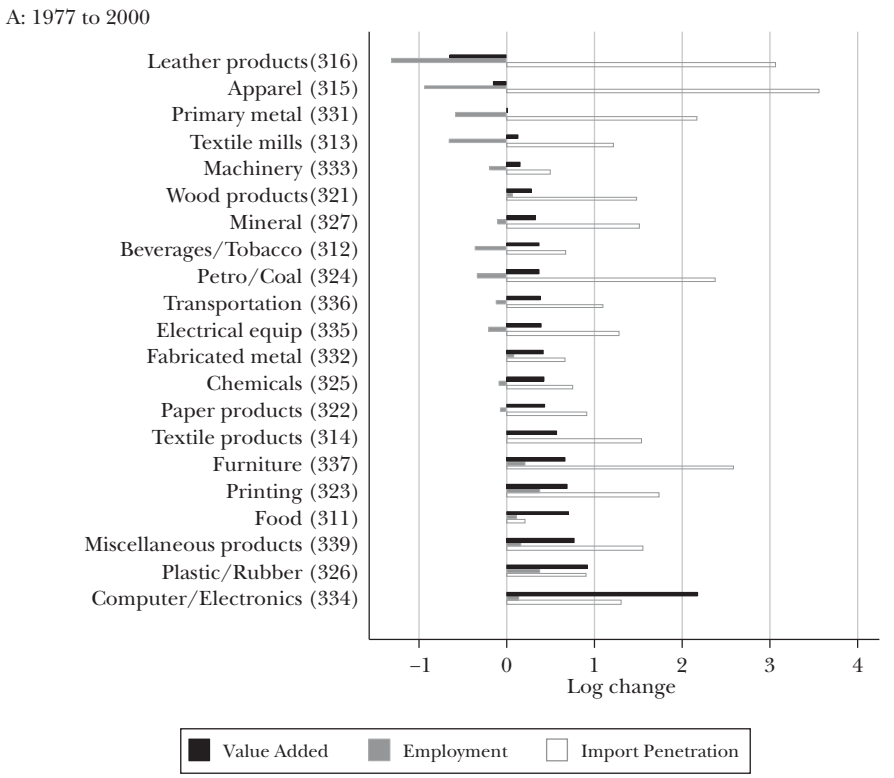
Figure 3 has three notable features with respect to identifying the influence of trade and technology. First note there are two sectors, Leather Products (316) and Apparel (315), that exhibit declines in *both* employment and value-added in both time periods. These sectors primarily encompass the production of labor-intensive goods such as clothing and footwear, commonly thought to be inconsistent with US comparative advantage. Apparel, in particular, has been subject to substantial tariff and quota reductions in the United States during the period we study (Khandelwal, Schott, and Wei 2013), and these liberalizations are reflected in the fact that it displays the largest increase in import penetration across sectors between 1977 and 2000.[5]

A second suggestive feature of Figure 3 is the increase in the number of sectors exhibiting simultaneous declines in real value added and employment in the panel B. Indeed, 52 percent of the 473 six-digit manufacturing industries that comprise manufacturing register such reductions between 2000 and 2007, versus 23 percent during the earlier time period. To the extent that this trend captures the exit of labor-intensive, low-labor-productivity firms within sectors whose products most overlap with Chinese manufacturers, this trend is consistent with the increase in Chinese import competition displayed in Figure 2 affecting US employment, and the research into trade liberalization with China discussed above. On the other hand, as Figure 2 also illustrates, the 2000s is a period when firms' use of computers and electronic networks increases. An intriguing possibility worthy of further attention, motivated by the anecdote in the introduction, is whether technology adoption during this period was hastened by trade liberalization with China.

The third noteworthy feature of Figure 3 with respect to trade and technology is the presence of sectors such as Chemicals (325), Transportation Equipment (336), and Miscellaneous Products (339; second panel only), in which value-added rises even as employment falls. These divergent outcomes, and the large growth in

[5]Reallocation may operate through occupations as well as industries, presenting another challenge to identifying the impacts of trade and technology. That is, the characteristics that make occupations susceptible to offshoring, such as routineness, also render them susceptible to automation (Ebenstein, Harrison, McMillan, and Phillips 2014; Oldenski 2014).

Figure 3
Change in Real Value Added, Employment, and Import Penetration Across Manufacturing Industries



(continued on next page)

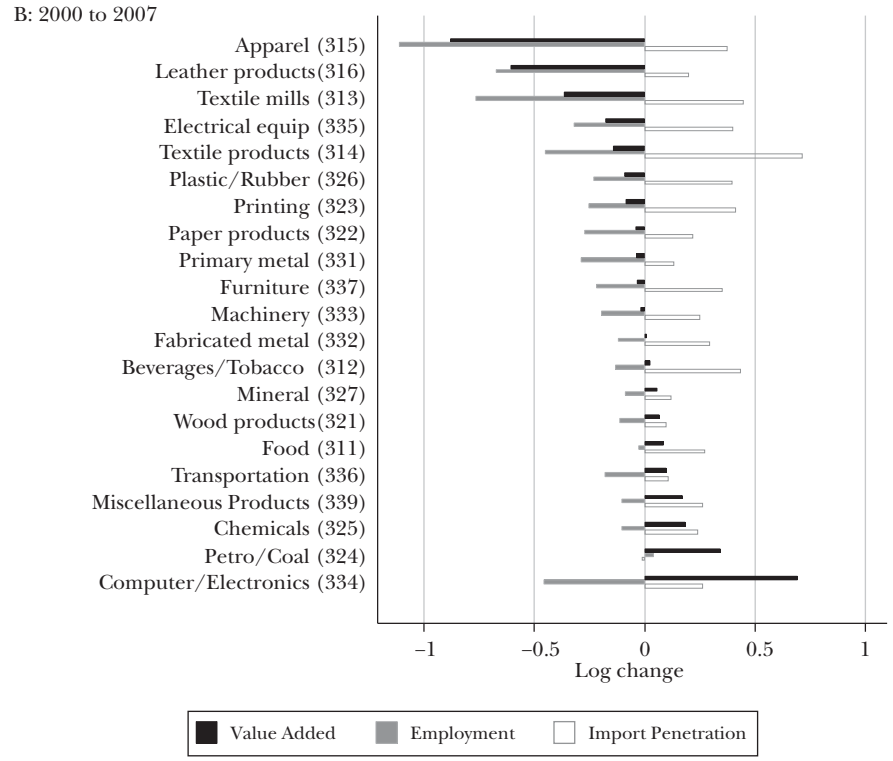
labor productivity they imply, suggest labor-saving technological change. In automobiles, for example, the replacement of workers with robots is widespread. On the other hand, to the extent that import competition induces selection away from low-labor-productivity industries within sectors, trade might also be playing a role (Schott 2003, 2004). Indeed, the industries within Miscellaneous Products with the largest loss and gain in employment between 1977 and 2000 are dolls and surgical instruments, respectively.

A particularly interesting sector exhibiting rising output along with falling employment in recent years is Computers and Electronic Products (334). As pointed out in Houseman, Kurz, Lengermann, and Mandel (2011) and suggested by its presence at the bottom of both panels of Figure 3, this sector accounts for the vast majority of real value-added growth in manufacturing over our sample period.[6] The two most influential industries within this sector in terms of aggregate real

[6]Houseman, Kurz, Lengermann, and Mandel (2011) also note that growth in manufacturing real value added may be overstated due to mismeasurement of prices for imported inputs.

Figure 3 (continued)

Change in Real Value Added, Employment, and Import Penetration Across Manufacturing Industries



Source: NBER-CES Manufacturing Industry Database (Becker et al. 2013), Feenstra (1996), Schott (2008) and authors' calculations.

Note: Real value added is deflated using shipment price deflators contained in the first source. Import penetration is real manufacturing imports divided by the sum of real shipments and real imports less real exports, in each case restricted to manufacturing industries. Industries are sorted by real value added growth over each period. Scales are different in each panel.

value-added growth are Semiconductors (334413) and Electronic Computer Manufacturing (334111). The latter has experienced significant growth in Chinese import penetration and is particularly well-known for its offshoring and outsourcing. Physical production of hard disk drives, like many other consumer electronic devices, has moved almost completely offshore during our sample period, even as their design centers remain in the United States (Igami 2018). The iPhone, in particular, is well known for being "designed in California" and assembled—using physical inputs from many countries, including the United States—in China (Folbre 2013).

The growing prevalence of such supply chains highlights a subtle but potentially important distinction between trade as import competition and trade as a technology. Although the bulk of US imports from China represent finished goods imported by

US wholesalers and retailers (Bernard, Jensen, Redding, and Schott 2010), Figure 2 reveals that a growing share of manufacturing firms import goods directly. These *direct imports* may have different consequences than *import penetration*: empirical analysis of US manufacturing firms by Antràs, Fort, and Tintelnot (2017) finds that while a firm's presence in an industry subject to increasing levels of Chinese import penetration is associated with declining firm-level employment between 1997 and 2007, increases in the value of its direct imports from China are associated with either growing or no change in employment. In their quantitative model, the authors provide a rationale for this difference, showing how greater access to foreign sourcing opportunities can allow importers to lower prices and raise output, even as non-importing firms shrink. Bernard, Fort, Smeets, and Warzynski (2018) also find that Danish firms exposed to increased import competition from China were more likely to offshore activities to Eastern Europe, which was associated with decreased domestic employment but not domestic output. Exploring the role of global value chains in the divergence between real output and employment is an important area for future research.

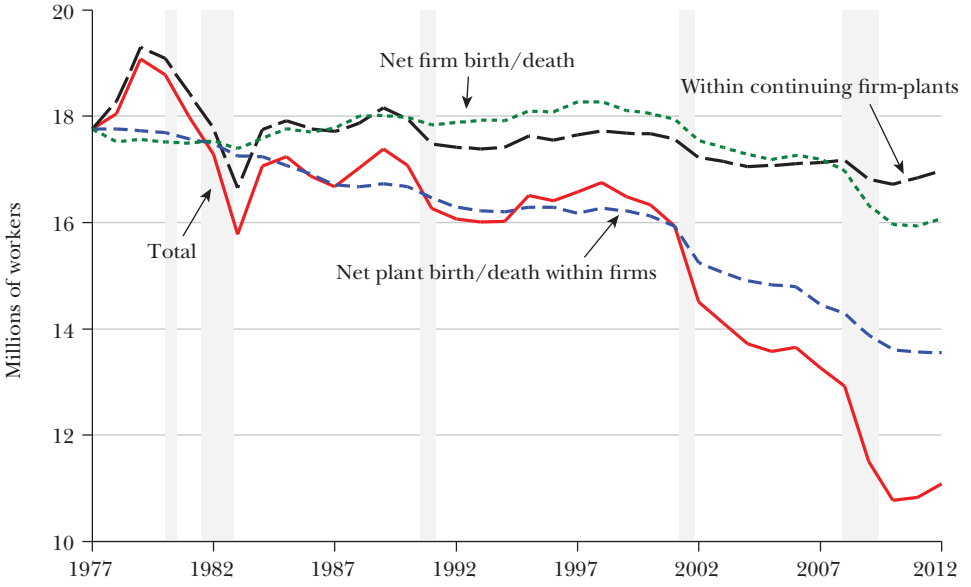
Reallocation of Employment Across and Within Firms

In this section, we dissect the overall shift in US manufacturing employment between 1977 and 2012 along firm and establishment margins of adjustment. We perform this decomposition using data from the Longitudinal Business Database (LBD) of the US Census Bureau, which links all private, nonfarm employer establishments and firms over time starting in 1977 (Jarmin and Miranda 2002). Each establishment is assigned a single industry code in each year based on its predominant activity.[7] The data make a useful distinction between an "establishment" and a "firm." An establishment denotes a single physical location where business transactions take place and for which payroll and employment records are kept, such as a manufacturing plant. In our analysis, as in official statistics, employees are grouped into industries based on the classification of the *establishment* in which they work. As a result, all employees in a manufacturing plant are classified as manufacturing employees, regardless of their occupation.

A "firm" is an organizational structure that can include one or more establishments, and therefore can span multiple industries. To capture all manufacturing employment in the Longitudinal Business Database, our decomposition includes all firms observed to have at least one manufacturing establishment at any point during

[7]We identify manufacturing plants based on an assignment of time-consistent NAICS codes developed by Fort and Klimek (2016) that ensure that the transition from SIC to NAICS does not result in spurious changes in the number of manufacturing workers based on changes in the set of activities considered "manufacturing." While the resulting manufacturing employment totals from the Longitudinal Business Database do not perfectly match the totals from the Bureau of Labor Statistics displayed in Figure 1, they are highly correlated over time. Our analysis drops records that are outside the scope of the County Business Patterns data, such as agriculture, and observations that are clearly erroneous, for example because of implausible payroll and employment numbers.

Figure 4
US Manufacturing Employment by Net Margin of Firm Adjustment



Source: Longitudinal Business Database and authors' calculations.
Note: Each line reports the change in employment along the noted net margin of firm adjustment relative to the firms and plants present in 1977. The shading corresponds to NBER-dated recessions.

the 1977 to 2012 sample period. The employment totals reported in this section are restricted solely to the manufacturing establishments at these firms; employment at their non-manufacturing establishments is analyzed later in the paper.

We examine three mutually exclusive firm margins of adjustment: changes in employment within the continuing establishments of continuing firms (also referred to as the "intensive" margin of continuing firm-plants), changes due to the birth and death of establishments within continuing firms, and changes due to the birth and death of entire firms.[8] Figure 4 illustrates the results. The solid line displays overall US manufacturing employment, showing the same pattern since 1977 as in Figure 1. The dashed lines trace out the cumulative employment in each year along the margins of adjustment, in each case relative to the firms and plants present in base year 1977. For example, the final value for the intensive margin indicates that firm-plants present in both 1977 and 2012 experience a decline in employment of approximately 0.8 million. Together, all three margins account for

[8]We follow Haltiwanger, Jarmin, and Miranda (2013) and define a firm death as occurring when all establishments of a firm exit from the Longitudinal Business Database. Analogously, firm birth occurs when all a firm's establishments are new to the LBD. While this approach avoids spurious firm birth and death due to merger and acquisition activity, future research into the extent to which these types of ownership changes are important factors in understanding manufacturing might be useful.

the 6.7 million overall decline in manufacturing employment registered by the solid line, from 17.8 to 11.1 million.

We find that most of the change in US manufacturing employment between 1977 and 2012—75 percent—takes place within firms that already existed in 1977 (consider the two lines "within continuing firm-plants" and "net plant birth/death within firms"). Most striking is the contribution of net plant birth/death within these firms, which by itself accounts for 63 percent of the overall change. Conversely, the set of firm-plants in continuous operation over the sample period is responsible for relatively little—12 percent—of the overall decline, with most of that occurring during the early 2000s.

The manner by which firms add or shed workers offers clues about their structure and transition costs, as well as the nature of the shocks they face. Consider three possibilities. If automating existing plants is relatively cost-effective, employment declines may be concentrated along the "intensive" margin—that is, within establishments of ongoing firms. If technology upgrades are more efficiently accomplished by shuttering outmoded plants in favor of new facilities, employment declines may occur via the net death of establishments within continuing firms.[9] If entrepreneurs at entering firms have an edge in creating or implementing new technologies, as argued by Christensen (1997), then resulting reductions in manufacturing employment may be driven by firm death, as outdated incumbents are pushed from the market.

Responses to increased pressures of international trade can, of course, operate along the same margins. Trade liberalization with low-wage countries might render a US firm's most labor-intensive products unprofitable. To the extent that firms are able to reallocate production away from these goods within existing facilities, globalization may manifest as declines in employment along the intensive margin. But if plants are wedded to particular products, employment loss may be driven by net plant death within continuing firms. If a broad set of firms' products is subject to increased import competition or if existing firms are unable to reallocate production within or across plants, trade competition may lead to the death of entire firms.

The fact that net firm death accounts for just 25 percent of the overall decline in US manufacturing employment between 1977 and 2012 is surprising given the magnitude of the drop in employment over this period, as well as common expectations of how creative destruction associated with trade and technology shocks likely operate. Indeed, as shown in online Appendix Figure A.3, we find that net firm *birth* accounts for the bulk of employment growth among non-manufacturing firms over the same period. On the other hand, most of the decline in employment along the net firm death margin occurs in the 2000s, which, as discussed above, may plausibly be related to import competition from China. As illustrated in online Appendix

[9] For example, Brynjolfsson and Hitt (2000) describe a medical manufacturer's experience transitioning to computer-integrated manufacturing. The firm's initial attempt to do so at an existing plant failed to generate productivity gains because current workers did not understand how to exploit the new processes. When the firm then opened a new plant with young employees, it realized such significant gains that it painted the plant windows black to prevent competitors from replicating its new techniques.

Figure A.4, we find a similar break with respect to the number of US manufacturing establishments: according to the Census Bureau's publicly available Business Dynamics Statistics (BDS), this series peaks in 1996. Overall, the small role of net firm death in the aggregate decline of US manufacturing employment suggests that incumbents may have an advantage relative to entrants.

The relatively sharp drop in employment associated with net plant death within continuing firms in the early 2000s, along with the contribution of net firm death during that period, may help rationalize the large distributional losses associated with increased import competition from China found in the literature. That is, to the extent that firm and plant closures were geographically concentrated, displaced workers may have found it more difficult to find new employment in their local labor market. On the other hand, the more-or-less constant decline of employment associated with net plant death within continuing firms prior to 2000 is consistent with firms continually replacing outmoded plants with new ones in response to a steady introduction of new technologies. To what extent do workers displaced by dying establishments find employment at new plants?

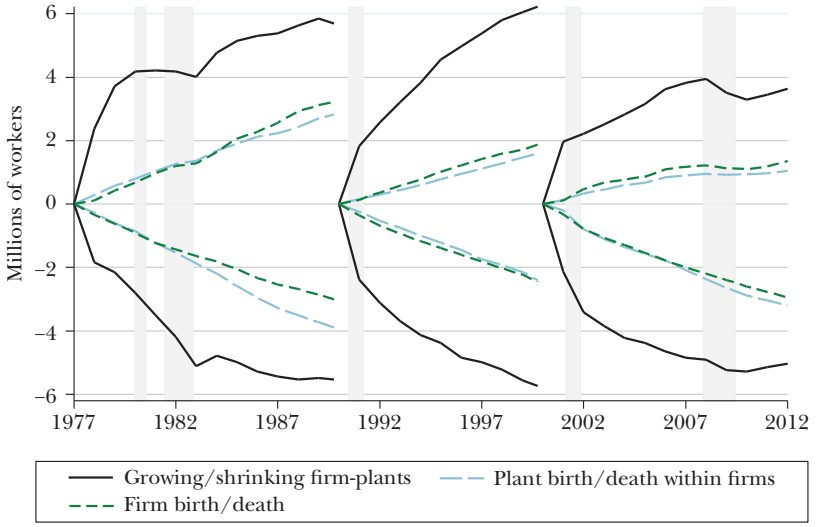
Simple descriptive regressions provide support for both trade and technology in plant turnover. For example, we find a negative correlation between the probability of a plant's death within a firm and the plant's purchases of computers. This correlation disappears after 2000, presumably due to the ubiquity of that technology, but during the 2000s we find another such correlation with respect to use of electronic networks to control or coordinate shipments.[10] In other words, there is heterogeneity within firms in terms of the establishments that adopt various technologies, and plants that do adopt these technologies have lower exit probabilities. With respect to trade, similar regressions indicate that before 2000, plant death within firms was correlated with increased import penetration in that plant's industry. After 2000, when firm death becomes a more important margin in the aggregate decline, these correlations are no longer present at the plant level, but firms facing increased import competition from China are more likely to exit.[11] One potential explanation for this result is that the firms that could re-orient themselves away from import-competing industries did so early on, either by shuttering plants or switching industries. For firms specializing in import-competing products, however, increased import penetration led to death.

The relatively small, 12 percent drop in employment among continuing firm-plants masks substantial gross flows associated with continuing firm-plants' expansion and contraction. We illustrate the magnitude of these gross flows in Figure 5, which decomposes the three net margins displayed in Figure 4 into their constituent gross job creation and destruction parts. In each case, job

[10] As discussed further in the online Appendix, these correlations are found by regressing indicator variables for plant death over years t to $t + 5$ on indicator variables for the noted activities in year t along with firm fixed effects.

[11] Unfortunately, given that trading is observed at the firm level, we are unable to examine whether plants that import are more or less likely to survive within firms over either period.

Figure 5
US Manufacturing Employment by Gross Margins of Adjustment



Source: Longitudinal Business Database and authors' calculations.
Note: Lines above zero are gross job creation margins and lines below zero are gross job destruction margins. For example, the solid line above zero displays employment growth associated with expanding plants among continuing firms, while the solid line below zero displays employment decline associated with shrinking plants at continuing firms. Employment changes along each margin are relative to the firms and establishments present in 1977, 1990, and 2000 respectively. The shading corresponds to NBER-dated recessions.

creation margins are displayed in lines above zero, while their corresponding job destruction margins are displayed in similarly patterned lines below zero. Here, to compare gross margins over time, and in contrast to Figure 4, we break the 1977 to 2012 sample period into three intervals that begin in base years 1977, 1990, and 2000. As a result, the gross margins for any year in Figure 5 are computed with respect to their nearest prior base year. For example, the final values for the gross continuing firm-plant margins indicate that firm-plants whose employment grew between 2000 and 2012 account for positive 3.6 million of the change in US manufacturing employment between 2000 and 2012, while continuing firm-plants whose employment fell accounted for negative 5.0 million.

The dominance of the intensive margin in gross employment changes represents another potentially fruitful area of study. To what extent is the adoption of new technologies, exposure to trade, or either importing or exporting associated with plant contraction? To what extent are they related to plant expansion? Large levels of job creation and destruction at continuing firms also suggest a potentially important role for technology and trade in worker reallocation. Are some workers more likely than others to shuffle among continuing plants? In the online Appendix, we show that firms' technology and trade activities are correlated with subsequent changes in their

employment and output, which is consistent with a role for both trade and technology in the reallocation of activities across continuing establishments.

Another noteworthy feature of Figure 5 is decline of all three gross job creation margins over time. These decreases are indicative of a drop in US business dynamism that has been documented across all sectors (Decker, Haltiwanger, Jarmin, and Miranda 2016). One potential explanation for this decline is a reduction in firms' responsiveness to productivity shocks due to rising adjustment frictions, such as regulatory constraints (Decker, Haltiwanger, Jarmin, and Miranda 2018), or the use of offshore rather than domestic capacity to make adjustments. Another is a reduction in competition, perhaps as a result of increasing entry barriers associated with adopting technology or adapting to globalization. De Loecker and Eeckhout (2017) document a steady rise in market power as measured by markups among US firms since the 1980s, with a sharp tick upwards in the early 2000s. A potentially intriguing area for further exploration is whether costs associated with trade or technology contribute to entry barriers. Using simple regressions of firm attributes on indicators for adoption and industry fixed effects, we find across census years—and display in the online Appendix Figure A.5—that firms purchasing computers and using electronic networks are significantly larger and have higher labor productivity than non-adopters.[12] Inspired by Acemoglu and Restreppo (2017), we find similar premia for firms that import industrial robots (Harmonized System product code 84.7950.0000) starting in 1997. These adoption premia are analogous to the size and productivity premia found for importers and exporters in the international trade literature (Bernard, Jensen, Redding, and Schott 2007). As such, they may reflect the fact that adoption of technology, like expansion into foreign markets, requires the payment of high fixed costs that only the largest, most productive firms find it optimal to incur.

Trade also may play a role in the decline of gross manufacturing job creation by pushing the US economy away from goods production and towards services. Pierce and Schott (2012b) and Asquith, Goswami, Neumark, and Rodriguez-Lopez (2017) show that during the 2000s, industries with relatively greater exposure to trade liberalization with China exhibit both suppressed job creation as well as exaggerated job destruction. Relatedly, the decline in gross manufacturing job creation along the margins of firm birth and plant birth within continuing firms may indicate that smaller, more capital-intensive firms and plants are entering at the expense of larger, more labor-intensive establishments and firms. In fact, as shown in online Appendix Figure A.4, using data from the Business Dynamics Statistics database referenced above, we find that the average number of workers per US manufacturing establishment fell 29 percent between 1977 and 2012, while the number of manufacturing establishments only begins to decline in the 1990s. Are these smaller entrants producing different goods more in line with US comparative advantage, or are they producing the same goods with a different technology?

[12] These regressions are described in greater detail in the online Appendix.

A final question related to the gross margins displayed in Figure 5 is the extent to which the decline in business dynamism in other sectors of the US economy might be related to the actions of manufacturing firms, or vice versa. Such relationships may occur through various channels, such as local labor markets or input–output linkages between manufacturing and non-manufacturing industries. Below, we show that another important dimension of such contact is the fact that manufacturing firms possess a sizable presence in non-manufacturing industries.

Reallocation of Employment across Regions

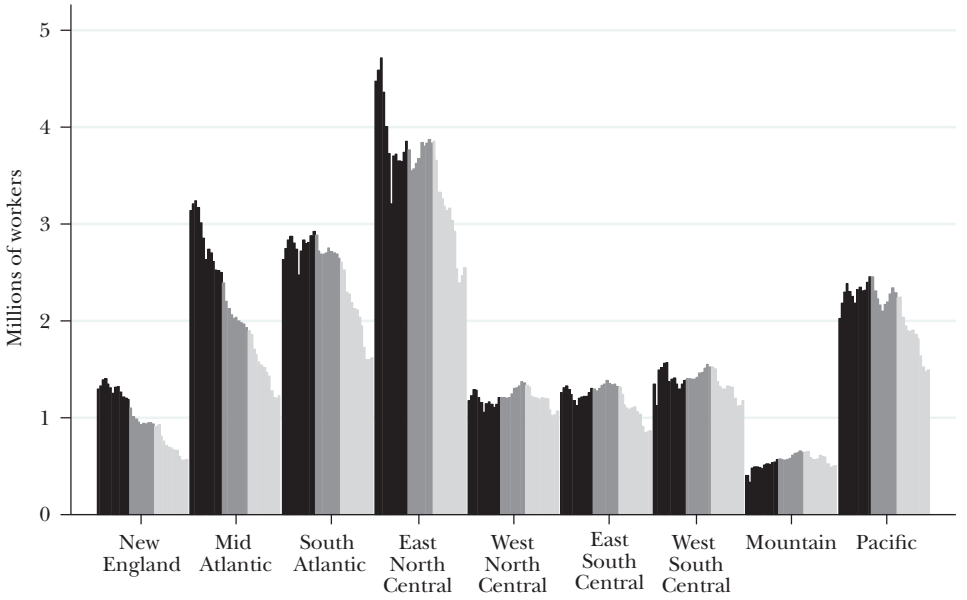
While a significant portion of the literature on both trade and technology has exploited regional variation in the distribution of manufacturing activities to identify causal impacts, plant and firm relocation within the United States remains a relatively unexplored dimension of firm adjustment to trade and technology shocks.

We find substantial reallocation of manufacturing employment across US regions over time, as well as differences in the extent to which regional declines in employment are driven by firm death versus continuing firms. Figure 6 plots US manufacturing employment from 1977 to 2012 by the nine US Census regions that comprise the United States. Each bar represents manufacturing employment in a given year and region, and bars are shaded to correspond to the three intervals used in Figure 5: 1977 to 1989 (black); 1990 to 1999 (dark grey); and 2000 to 2012 (light grey). As indicated in the figure, manufacturing employment in the New England, Mid-Atlantic, and East North Central regions declines more-or-less steadily over the sample period. In the rest of the country, by contrast, it is either relatively flat or growing until 2000, after which manufacturing employment in all regions shrinks. Indeed, between 1977 and 2000, combined manufacturing employment in the New England, Mid-Atlantic, and East North Central regions falls by 2.3 million, while the increase for all other regions as a whole is 0.8 million. After 2000, the largest decline, in percentage terms, occurs in the South Atlantic region (a drop of 38 percent).

Regions also display interesting variation in terms of the margins of firm adjustment. In results reported in online Appendix Figure A.7, we show that employment loss due to net firm death is concentrated in the New England and Mid-Atlantic regions, which together account for 16 percentage points of the overall 25 percentage point decline in US manufacturing employment attributable to that margin. The East North Central region, by contrast, stands out in terms of its disproportionate loss of employment within continuing firm-plants.

Reallocation of manufacturing activity within the United States might shed useful light on reallocation internationally. Indeed, movement of US manufacturing employment from the north and east towards the west and south up to 2000 may have been a precursor to international offshoring. Bernard, Redding, and Schott (2013), for example, show that US labor markets exhibit substantial and persistent variation in relative skill endowments and wages over this period, and that

Figure 6
US Manufacturing Employment by Census Region, 1977–2012



Source: Longitudinal Business Database and authors' calculations.
Note: Panels report manufacturing employment across years and census regions. Years from 1977 to 1989, 1990 to 1999, and 2000 to 2012 are shaded black, dark grey, and light grey, respectively. Census regions are defined as follows. New England: CT, ME, MA, NH, RI, VT. Middle Atlantic: NJ, NY, PA. East North Central: IN, IL, MI, OH, WI. West North Central: IA, KS, MN, MO, NE, ND, SD. South Atlantic: DE, DC, FL, GA, MD, NC, SC, VA, WV. East South Central: AL, KY, MS, TN. West South Central: AR, LA, OK, TX. Mountain: AZ, CO, ID, MT, UT. NV, WY. Pacific: AK, CA, HI, OR, WA.

labor markets with different relative wages tend to specialize in different groups of industries. Fort (2017) shows that US manufacturing establishments in high-wage locations are more likely to fragment production, especially domestically. Anecdotal evidence suggests firms do in fact relocate in response to variation in wages across local labor markets. Radio Corporation of America (RCA), for example, continually moved production of its most labor-intensive products west and south in search of lower wages before moving it to Mexico in the 1990s (Cowie 1999). Such activity is consistent with the Holmes (1998) finding that manufacturing employment is relatively low in more union-friendly states compared to neighboring right-to-work states. These right-to-work states are clustered in the South Atlantic, West Central, and Mountain regions, where manufacturing employment was stable or growing prior to 2000. Were such reallocations also a response to international competition? Were they facilitated by technologies that allow firms to serve customers from more remote, lower-cost labor markets? Do incumbents have an advantage in making use of such technologies?

A thornier question raised by Figure 6 is whether relocation within the United States, either within or across firms, coincides with labor-saving technology upgrades, as suggested by the long-running decline in the average number of employees per establishment referenced above? If so, how can a causal impact of technology be identified?

The Non-Manufacturing Establishments of Manufacturing Firms

Manufacturing firms can also have non-manufacturing establishments. In this section we broaden our analysis to investigate how employment at manufacturing firms' non-manufacturing establishments has evolved, and in what non-manufacturing industries they participate. As noted earlier, in this paper we define a manufacturing firm broadly to encompass any firm observed to have at least one manufacturing establishment during our 1977 and 2012 sample period. The non-manufacturing employment of manufacturing firms, therefore, is simply the sum of employment at any non-manufacturing establishments owned by a manufacturing firm. While we focus on this comprehensive set of firms in order to capture all manufacturing employment, it is important to bear in mind that this definition includes firms not traditionally thought of as manufacturers—for example, big box retailers that may encompass relatively small food preparation facilities—and that such firms might have an outsized impact on the trends in non-manufacturing employment we analyze here.

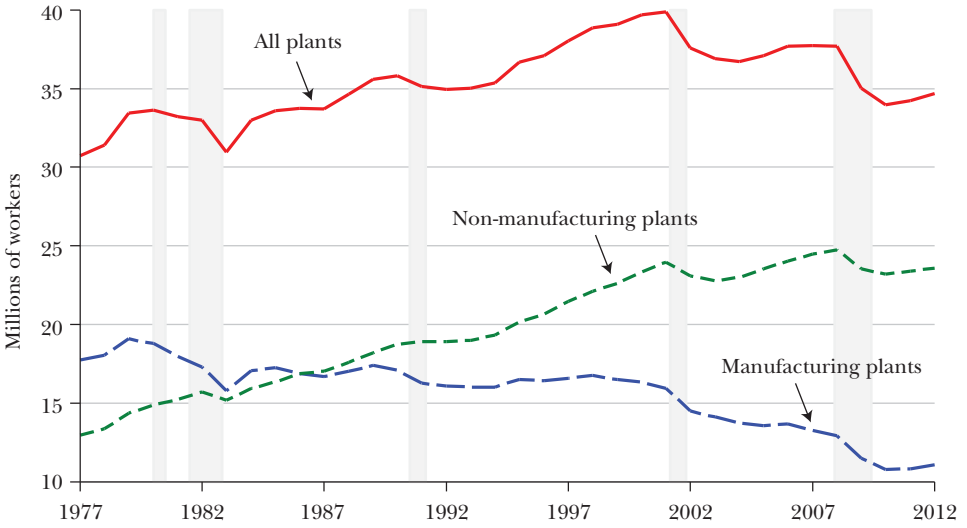
With this caveat in mind, Figure 7 displays total employment of manufacturing firms across their manufacturing versus non-manufacturing establishments. As indicated in the figure, non-manufacturing employment rises more-or-less steadily until 2000, when it levels off. As a result, total employment of manufacturing firms rises until 2000 before declining afterwards due to the sharp drop in employment at their manufacturing establishments.[13] In online Appendix Figure A.3, we show that most of the growth in manufacturing firms' non-manufacturing employment occurs via net non-manufacturing plant birth within continuing firms.

The growing share of manufacturing firm employment at non-manufacturing establishments might indicate that a growing number of workers at non-manufacturing establishments is needed to support manufacturing production, or that the scope of manufacturing firms is widening to include additional non-manufacturing activities, or simply that employment growth at firms' non-manufacturing establishments reflects the broader shift of US employment toward non-manufacturing

[13] In work not reported here, we find that the trends displayed in Figure 7 are sensitive to how manufacturing firms are defined. For example, requiring firms to have at least some threshold level of employment in manufacturing in at least one year of the sample results in flatter growth of non-manufacturing employment over the sample period. In addition, the growth of non-manufacturing employment at manufacturing firms, even with our broad definition of manufacturing firms, is slower than the growth of non-manufacturing employment at non-manufacturing firms. This differential is also worthy of further exploration.

Figure 7

Employment at Manufacturing Firms Decomposed into Employment at Manufacturing versus Non-Manufacturing Establishments, 1977–2012



Source: Longitudinal Business Database and author's calculations.
Note: Manufacturing firms are defined as any firm observed to have a manufacturing establishment during the sample period. The shading corresponds to NBER-dated recessions.

activities.[14] Further insight into these explanations comes from analysis of the particular activities occurring at non-manufacturing plants of manufacturing firms. Toward that end, we break non-manufacturing industries into three groups based on their two-digit NAICS sectors: retail (NAICS 44 to 45), professional services (NAICS 51 to 56), and all other non-manufacturing industries. Perhaps unsurprisingly, given the broad definition of manufacturing firms noted above, we find that about one-third of the overall growth in non-manufacturing employment of manufacturing firms between 1977 and 2012 is in retail, while another third falls into the "other" category.

However, 32 percent of the increase in non-manufacturing employment at manufacturing firms is driven by professional services, which captures a wide range of often skill-intensive activities: information technology (NAICS 51); finance, insurance, real estate and leasing (NAICS 52-3); engineering and other technical services (NAICS 54); headquarters services (NAICS 55); and administrative support and waste management (NAICS 56). The growing use of workers in such industries may reflect

[14] While recent research suggests that US manufacturers increasingly outsource ancillary services such as cleaning to domestic contractors (Dey, Houseman, and Polivka 2012; Berlingieri 2014; Katz and Krueger 2016), such activity would not be captured in Figure 7 as it traces non-manufacturing employment *within* manufacturing firms.

the influence of both trade and technology. For example, one action US manufacturers might take in response to growing import competition in goods is to move into neuro-facturing (Leamer 2009), either by diversifying away from goods production entirely or by making use of various communications and management technologies to focus on the engineering, design or marketing of goods rather than their physical production (Bernard and Fort 2015, 2017). Consistent with this explanation, Magyari (2017) finds that in certain cases, US manufacturing firms expanded their non-manufacturing employment in response to import competition from China.

These findings raise a number of intriguing questions. Does increasing use of design, marketing, and other management services facilitate the product differentiation and upgrading US firms undertake to compete with producers from low-wage countries? Does it help explain the rising market power of US producers documented in De Loecker and Eeckhout (2017)? Do US manufacturing firms expand their service activities in the same geographic areas in which they used to produce physical goods? As illustrated in Appendix Figure A.6, though 46 percent of manufacturing firms' nonmanufacturing employment growth takes place in the western half of the United States, the South Atlantic exhibits the fastest pace of non-manufacturing employment growth by manufacturing firms, at 27 percent. Further analysis of the broader scope of US manufacturing firms' activities across both geographic and regional dimensions seems promising.

Conclusion

The decline in US manufacturing jobs and concerns over the competitiveness of US manufacturers in a global market place have sparked considerable commentary and research in recent years, including several articles in this journal, by Charles, Hurst, and Notowidigdo (2016), Baily and Bosworth (2014), Tassey (2014), and Houseman et al. (2011). A natural question arising in these discussions is whether trade or technology plays a larger role in the sector's outcomes. As we have explained, we find that question to be overly broad. It may also distract needed attention away from research into how to facilitate reallocation among displaced manufacturing workers. Given that few economists advocate for restricting either technology or trade, such research seems both timely and necessary.

Instead, we have sought to gain new perspective on the decline of US manufacturing employment by examining relatively unexplored dimensions of microdata tracking US manufacturing firms over time, and considering how patterns in those data might be explained by various mechanisms associated with trade, technology, and other forces. Here, we summarize a few of the empirical facts we report and mention some follow-up questions that are worth pursuing.

We find that 75 percent of the 6.6 million decline in manufacturing employment between 1977 and 2012 took place within continuing firms, largely through plant closures. Why is the primary adjustment *within firms*, and in the form of plant closures? What barriers to entry—regulatory or otherwise—might have dampened

firm creation or suppressed firm destruction? How do entrants' technology and production functions differ from those of incumbents and firms that have died? What are the implications of these plant closures and new production techniques for displaced workers?

Manufacturing firms' activities outside manufacturing might offer some clues for the persistence of incumbent manufacturing firms. Before 2000, the drop in manufacturing firms' manufacturing employment is more than offset by increases in non-manufacturing workers. After 2000, a sharp decline in those firms' manufacturing employment and a flattening of their non-manufacturing employment growth leads to a decrease in their total employment. Relatively high-skill professional workers—like designers and engineers—account for approximately one-third of the non-manufacturing workers added by manufacturing firms. Are incumbents firms better suited to engage in these activities? Does the greater focus of manufacturing firms on services mimic the growth in services that takes place across non-manufacturing firms, or does it point to an important role for the firm in building up capabilities that persist over time?

Finally, trade and technology can interact with different parts of manufacturing in very different ways. Manufacturing firms that adopt specific technologies, such as computers or industrial robots, are significantly different from those that do not: in particular, they are larger and more productive upon adoption. Importing is associated with different outcomes at the firm and industry levels: while exposure to greater import competition is associated with employment decline, firms increasing their use of imported goods conditional on such exposure can exhibit employment gains. Should direct use of imported goods be considered a technology?

US manufacturing has many dimensions: manufacturing and non-manufacturing establishments; overall trends of falling employment and rising value added; incumbent and non-incumbent firms; geographical movements within US regions; sunset and sunrise industries; differences in firm-level choices regarding importing inputs and use of technology; and differences across industries from import penetration and the spread of technology. Our understanding of how trade and technology affect US manufacturing must seek to be multifaceted as well.

■ *Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the US Census Bureau, the Board of Governors, or its research staff. All results have been reviewed to ensure that no confidential information is disclosed. Part of this research was conducted while Teresa Fort was a Peter B. Kenen Fellow in the International Economics Section at Princeton University. She thanks the IES for financial support. We thank the editors for comments and Jim Davis for his exceptional help with the disclosure review process.*

References

Acemoglu, Daron. 2002. "Directed Technical Change." Review of Economic Studies 69(4): 781–809.

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2015. "Import Competition and the Great U.S. Employment Sag of the 2000s." *Journal of Labor Economics* 34(S1, Part 2): S141–98.

Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper 23285.

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot. 2017. "The Margins of Global Sourcing: Theory and Evidence from US Firms." *American Economic Review* 107(9): 2514–64.

Asquith, Brian J., Sanjana Goswami, David Neumark, and Antonio Rodriguez-Lopez. 2017. "U.S. Job Flows and the China Shock." NBER Working Paper 24080.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Autor, David, David Dorn, and Gordon Hanson. 2015. "Untangling Trade and Technology: Evidence from Local Labour Market." *Economic Journal* 125(584): 621–46.

Autor, David H., David Dorn, Gordon H. Hanson, and Jae Song. 2014. "Trade Adjustment: Worker Level Evidence." *Quarterly Journal of Economics* 129(4): 1799–1860.

Autor, David H., David Dorn, and Gordon H. Hanson. 2018. "When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men." NBER Working Paper 23173.

Autor, David H., David Dorn, Gordon H. Hanson, **Gary Pisano, and Pian Shu.** 2016. "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents." NBER Working Paper 22879.

Baily, Martin Neil, and Barry P. Bosworth. 2014. "US Manufacturing: Understanding Its Past and Its Potential Future." *Journal of Economic Perspectives* 28(1): 3–26.

Barrot, Jean-Noel, Erik Loualiche, Matthew Plosser, and Julien Sauvagnat. 2017. "Import Competition and Household Debt." Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2808981.

Becker, Randy, Wayne B. Gray, and Jordan Marvakov. 2013. NBER-CES Manufacturing Industry Database. Available at http://www.nber.org/nberces/.

Berlingieri, Guiseppe. 2014. "Outsourcing and the Rise in Services." CEP Discussion Papers CEPDP1199, Centre for Economic Performance.

Bernard, Andrew B., and Teresa C. Fort. 2015. "Factoryless Goods Producing Firms." *American Economic Review* 105(5): 518–23.

Bernard, Andrew B., and Teresa C. Fort. 2017. "Factoryless Goods Producers in the US." Chap. 5 in *The Factory-Free Economy: Outsourcing, Servitization, and the Future of Industry*, edited by Lionel Fontagné and Ann Harrison. Oxford University Press.

Bernard, Andrew B., Teresa C. Fort, Valerie Smeets, Frederic Warzynski. 2018. "Heterogeneous Globalization: Offshoring and Reoganization." http://faculty.tuck.dartmouth.edu/images/uploads/faculty/teresa-fort/BFSW_Offshoring_Reorg_03-18-18.pdf.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2010. "Wholesalers and Retailers in US Trade." *American Economic Review* 100(2): 408–413.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2006. "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of U.S. Manufacturing Plants." *Journal of International Economics* 68(1): 219–37.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives* 21(3): 105–130.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2011. "Multiproduct Firms and Trade Liberalization." *Quarterly Journal of Economics* 126(3): 1271–1318.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2013. "Testing for Relative Factor Price Equality with Unobserved Differences in Factor Quality or Productivity." *American Economic Journal: Microeconomics* 5(2): 135–63.

Bernard, Andrew B., Valerie Smeets, and Frederic Warzynski. 2017. "Rethinking Deindustrialization." *Economic Policy* 32(89): 5–38.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2016. "Trade Induced Technical Change: The Impact of Chinese Imports on Innovation, IT, and Productivity." *Review of Economic Studies* 83(1): 87–117.

Bown, Chad. 2016. Global Antidumping Database. World Bank.

Brynjolfsson, Erik, and Lorin M. Hitt. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives* 14(4): 23–48.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2015. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the

China Trade Shock." NBER Working Paper 21149.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30(2): 179–200.

Che, Yi, and Xun, Xu. 2016. "The China Syndrome in US: Import Competition, Crime, and Government Transfer." Mimeo, University of Munich.

Christensen, Clayton M. 1997. *The Innovator's Dilemma*. New York: HarperBusiness.

Clemens, Michael A., Ethan G. Lewis, and Hannah M. Postel. 2017. "Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion." NBER Working Paper 23125.

Collard-Wexler, Alan, and Jan De Loecker. 2015. "Reallocation and Technology: Evidence from the U.S. Steel Industry." *American Economic Review* 105(1): 131–71.

Cowie, Jefferson. 1999. *Capital Moves: RCA's Seventy-Year Quest for Cheap Labor*. Ithaca, NY: Cornell University Press.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S." *European Economic Review* 86:4–23.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." *European Economic Review* 86:4–23.

De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." Unpublished paper.

Dey, Matthew, Susan N. Houseman, and Anne E. Polivka. 2012. "Manufacturers' Outsourcing to Staffing Services." *Industrial & Labor Relations Review* 65(3): 533–59.

Ebenstein, Avraham, Ann Harrison, Margaret McMillan, and Shannon Phillips. 2014. "Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys." *Review of Economics and Statistics* 96(3): 581–95.

Edwards, Lawrence, and Robert Z. Lawrence. 2013. *Rising Tide: Is Growth in Emerging Economies Good for the United States?* Washington, DC: Peterson Institute for International Economics.

Feenstra, Robert C. 1996. "U.S. Imports, 1972–1994: Data and Concordances." NBER Working Paper 5515.

Feler, Leo, and Mine Z. Senses. Forthcoming. "Trade Shocks and the Provision of Local Public Goods." *American Economic Journal: Economic Policy*.

Folbre, Nancy. 2013. "Not Really Made in China (or the United States)". *New York Times*, August 19.

Fort, Teresa C. 2017. "Technology and Production Fragmentation: Domestic versus Foreign Sourcing." *Review of Economic Studies* 84(2): 650–87.

Fort, Teresa C., and Shawn D. Klimek. 2016. "The Effects of Industry Classification Changes on US Employment Composition." http://faculty.tuck.dartmouth.edu/images/uploads/faculty/teresa-fort/fort_klimek_naics.pdf.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi. 2017. "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade." NBER Working Paper 23737.

Goos, Maarten, Alan Manning, and Anna Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104(8): 2509–26.

Graetz, Georg, and Guy Michaels. 2017. "Is Modern Technology Responsible for Jobless Recoveries?" *American Economic Review* 107(5): 168–73.

Jarmin, Ron, and Javier Miranda. 2002. "The Longitudinal Business Database." Center for Economic Studies Discussion Paper CES-WP-02-17. https://www.census.gov/ces/dataproducts/datasets/lbd.html.

Hakobyan, Shushanik, and John McLaren. 2016. "Looking for Local Labor Market Effects of NAFTA." *Review of Economics and Statistics* 98(4): 728–41.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics* 95(2): 347–61.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review* 107(9): 2731–83.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy* (106)4: 667–705.

Houseman, Susan, Christopher Kurz, Paul Lengermann, and Benjamin Mandel. 2011. "Offshoring Bias in U.S. Manufacturing." *Journal of Economic Perspectives* 25(2): 111–32.

Igami, Mitsuru. 2018. "Industry Dynamics of Offshoring: The Case of Hard Disk Drives." *American Economic Journal: Microeconomics* 10(1): 67–101.

Katz, Lawrence F., and Alan B. Krueger. 2016. "The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015." NBER Working Paper 22667.

Khandelwal, Amit. 2010. "The Long and Short (of) Quality Ladders." *Review of Economic Studies*

77(4): 1450–76.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei. 2013. "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters." *American Economic Review* 103(6): 2169–95.

Langdon, David, and Rebecca Lehrman. 2012. "The Benefits of Manufacturing Jobs." ESA Issue Brief No. 01-12. U.S. Department of Commerce, Economics and Statistics Administration, available at: http://esa.doc.gov/sites/default/files/1thebenefitsofmanufacturingjobsfinal5912.pdf.

Lawrence, Robert Z., and Lawrence Edwards. 2013. "US Employment Deindustrialization: Insights from History and the International Experience." Policy Brief PB13-27, Peterson Institute for International Economics, October.

Leamer, Edward E. 2009. "Heckscher–Ohlin Models for the Post-Industrial Age." Unpublished paper, UCLA.

Magyari, Ildikó. 2017. "Firm Reorganization, Chinese Imports, and US Manufacturing Employment." Working Paper 17-58, Center for Economic Studies, US Census Bureau.

Michaels, Daniel. 2017. "Foreign Robots Invade American Factory Floors." *Wall Street Journal*, March 26.

Muendler, Marc-Andreas. 2017. "Trade, Technology, and Prosperity: An Account of Evidence from a Labor-Market Perspective." WTO Staff Working Paper ERSD-2017-15, World Trade Organization.

Oldenski, Lindsay. 2014. "Offshoring and the Polarization of the U.S. Labor Market." *Industrial and Labor Relations Review* 67(3, suppl.): 734–61.

Oster, Sharon. 1982. "The Diffusion of Innovation among Steel Firms: The Basic Oxygen Furnace." *Bell Journal of Economics* 13(1): 45–56.

Pierce, Justin R. 2011. "Plant-Level Responses to Antidumping Duties: Evidence from U.S. Manufacturers." *Journal of International Economics* 85(2): 222–33.

Pierce, Justin R., and Peter K. Schott. 2012a. "Concording U.S. Harmonized System Codes over Time." *Journal of Official Statistics* 28(1): 53-68.

Pierce, Justin R., and Peter K. Schott. 2012b. "The Surprisingly Swift Decline of U.S. Manufacturing Employment." NBER Working Paper 18655.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of U.S. Manufacturing Employment." *American Economic Review* 106(7): 1632–62.

Pierce, Justin R., and Peter K. Schott. 2017. "Trade Liberalization and Mortality: Evidence from U.S. Counties." NBER Working Paper 22849.

Schott, Peter K. 2003. "One Size Fits All? Heckscher–Ohlin Specialization in Global Production." *American Economic Review* 93(2): 686–708.

Schott, Peter K. 2004. "Across-Product versus Within-Product Specialization in International Trade." *Quarterly Journal of Economics* 119(2): 647–78.

Schott, Peter K. 2008. "The Relative Sophistication of Chinese Exports." *Economic Policy* 23(53): 5–49.

Steinwender, Claudia. 2018. "Real Effects of Information Frictions: When the States and the Kingdom Became United." *American Economic Review* 108(3): 657–96.

Sullivan, Daniel, and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data." *Quarterly Journal of Economics* 124(3): 1265–1306.

Tassey, Gregory. 2014. "Competing in Advanced Manufacturing: The Need for Improved Growth Models and Policies." *Journal of Economic Perspectives* 28(1): 27–48.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 73–90

What Do Trade Agreements Really Do?

Dani Rodrik

The Booth School of Business at the University of Chicago asked its panel of economics experts—made up of leading professors of economics around the country—to respond to two statements on international trade in its March 2012 survey (at http://www.igmchicago.org/surveys/free-trade). The first statement focused on attitudes towards the general concept of free trade: "Freer trade improves productive efficiency and offers consumers better choices, and in the long run these gains are much larger than any effects on employment." The second statement honed in specifically on the North American Free Trade Agreement (NAFTA): "On average, citizens of the U.S. have been better off with the North American Free Trade Agreement than they would have been if the trade rules for the U.S., Canada and Mexico prior to NAFTA had remained in place." The experts could choose among a range of options, from "strongly agree" to "strongly disagree."

There was near-unanimous support for the first statement on free trade. Of the 37 economists who answered, 35 picked "strongly agree" or "agree." Two answered "uncertain" and none disagreed. The second question on NAFTA produced a virtually identical response. Once again, no one disagreed and only two economists picked "uncertain." The only difference was that there was one less vote for "strongly agree" (reducing the tally for this option from 11 to 10) and one more vote for "agree" (raising the tally from 24 to 25).

■ *Dani Rodrik is the Ford Foundation Professor of International Political Economy, John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts. His email address is dani_rodrik@harvard.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.73

doi=10.1257/jep.32.2.73

The consensus in favor of the general statement supporting free trade is not a surprise. Economists disagree about a lot of things, but the superiority of free trade over protection is not controversial. The principle of comparative advantage and the case for the gains from trade are crown jewels of the economics profession, so the nearly unanimous support for free trade in principle is understandable. But the almost identical level of enthusiasm expressed for the North American Free Trade Agreement—that is, for a text that runs into nearly 2,000 pages, negotiated by three governments under pressures from lobbies and special interests, and shaped by a mix of political, economic, and foreign policy objectives—is more curious.

The economists must have been aware that trade agreements, like free trade itself, create winners and losers. But how did they weight the gains and losses to reach a judgment that US citizens would be better off "on average"? Did it not matter who gained and lost, whether they were rich or poor to begin with, or whether the gains and losses would be diffuse or concentrated? What if the likely redistribution was large compared to the efficiency gains? What did they assume about the likely compensation for the losers, or did it not matter at all? And would their evaluation be any different if they knew that recent research suggests NAFTA produced minute net efficiency gains for the US economy while severely depressing wages of those groups and communities most directly affected by Mexican competition?[1]

Perhaps the experts viewed distributional questions as secondary in view of the overall gains from trade. After all, opening up to trade is analogous to technological progress. In both cases, the economic pie expands while some groups are left behind. We did not ban automobiles or light bulbs because coachmen and candle-makers would lose their jobs. So why restrict trade? As the experts in this survey contemplated whether US citizens would be better off "on average" as a result of NAFTA, it seems plausible that they viewed questions about the practical details or the distributional questions of NAFTA as secondary in view of the overall gains from trade.

This tendency to view trade agreements as an example of efficiency-enhancing policies that may nevertheless leave some people behind would be more justifiable if recent trade agreements were simply about eliminating restrictions on trade such as import tariffs and quotas. In fact, the label "free trade agreements" does not do a very good job of describing what recent proposed agreements like the Trans-Pacific Partnership (TPP), the Trans-Atlantic Trade and Investment Partnership (TTIP), and numerous other regional and bilateral trade agreements actually do.

[1] Caliendo and Parro (2015) is the most sophisticated evaluation to date of NAFTA's overall economic effects. These authors develop a multisector, multicountry Ricardian model with intermediate inputs and productive heterogeneity within sectors. They conclude that NAFTA increased US "welfare" by 0.08 percent (that is, by less than one tenth of 1 percent). Moreover half of this gain came not from an increase in efficiency but from an improvement in the US terms of trade (that is, at the expense of other countries, mainly Mexico). As for the distributional impacts, they have been recently estimated by Hakobyan and McLaren (2016). These authors find very sharp adverse effects for certain groups of workers. High school dropouts working in industries that were heavily protected by tariffs on Mexican exports prior to NAFTA experienced a drop in wage growth of as much as 17 percentage points relative to wage growth in unaffected industries.

Contemporary trade agreements go much beyond traditional trade restrictions at the border. They cover regulatory standards, health and safety rules, investment, banking and finance, intellectual property, labor, the environment, and many other subjects. They reach well beyond national borders and seek deep integration among nations rather than shallow integration, to use Lawrence's (1996) helpful distinction. According to one tabulation, 76 percent of existing preferential trade agreements covered at least some aspect of investment (such as free capital mobility) by 2011; 61 percent covered intellectual property rights protection; and 46 percent covered environmental regulations (Limão 2016).

To illustrate the changing nature of trade agreements, compare US trade agreements with two small nations, Israel and Singapore, signed two decades apart. The US–Israel Free Trade Agreement, which went into force in 1985, was the first bilateral trade agreement the US concluded in the postwar period. It is quite a short agreement—less than 8,000 words in length. It contains 22 articles and three annexes, the bulk of which are devoted to free-trade issues such as tariffs, agricultural restrictions, import licensing, and rules of origin. The US–Singapore Free Trade Agreement went into effect in 2004 and is nearly ten times as long, taking up 70,000 words. It contains 20 chapters (each with many articles), more than a dozen annexes, and multiple side letters. Of its 20 chapters, only seven cover conventional trade topics. Other chapters deal with behind-the-border topics such anti-competitive business conduct, electronic commerce, labor, the environment, investment rules, financial services, and intellectual property rights. Intellectual property rights take up a third of a page (and 81 words) in the US–Israel agreement. They occupy 23 pages (and 8,737 words) plus two side letters in the US–Singapore agreement.

Taking these new features into account requires economists to rethink their default attitudes toward trade agreements, and the politics behind them. This paper offers a starting point toward the reconsideration that is needed. I will argue that economists' conflation of free trade with trade agreements is rooted in an implicit political economy perspective that views import-competing interests as the most powerful and dominant architect of trade policy. Under this perspective, protectionists on the import side are the main villain of the story. Trade agreements, when successfully ratified, serve to counter their influence and get us closer to a welfare optimum by reducing the protectionism (or harmful regulations) that these special interests desire. In particular, they prevent beggar-thy-neighbor and beggar-thyself policies that would result in the absence of trade agreements. In achieving these ends, governments may be assisted by *other* special interests—those with a stake in expanding exports and market access abroad. But the latter play an essentially useful role, since they are merely a counterweight to the protectionist lobbies.

There is an alternative political economy perspective, one that reverses the presumption about which set of special interests hold the upper hand in trade policy. In this view, trade agreements are shaped largely by rent-seeking, self-interested behavior on the export side. Rather than reining in protectionists, trade agreements empower another set of special interests and politically well-connected firms, such

as international banks, pharmaceutical companies, and multinational corporations. Such agreements may result in freer, mutually beneficial trade, through exchange of market access. But they are as likely to produce welfare-reducing, or purely redistributive outcomes under the guise of free trade.

When trade agreements were largely about import tariffs and quotas—that is before the 1980s—the second scenario may not have been particularly likely. But with trade agreements increasingly focusing on domestic rules and regulations, we can no longer say the same. Taking these new features into account requires us to cast trade agreements, and the politics behind them, in quite a different light.

Free Trade versus Free Trade Agreements

Basic trade theory suggests that free trade is the optimal policy for an economy, provided compensatory policies can be implemented and adverse interactions with market failures can be addressed through complementary policies. The only exception is that a large country may be able to manipulate its terms of trade at the expense of its trade partners, using an "optimal tariff." The latter motive provides a rationale for countries to enter into trade agreements, preventing mutually harmful trade protectionism.

Economists have long known that real-world trade agreements are difficult to understand from the lens of "optimal tariff" theory. And as trade agreements have evolved and gone beyond import tariffs and quotas into regulatory rules and harmonization (patent rules, health and safety regulations, labor standards, investor courts, and so on), they have become harder and harder to fit into received economic theory.

International agreements in such new areas produce economic consequences that are far more ambiguous than is the case of lowering traditional border barriers. They may well generate increases in the volume of trade and cross-border investment. Nevertheless their welfare and efficiency impacts are fundamentally uncertain. Here, I will sketch the issues that arise in four areas that have become common in modern trade agreements: trade-related intellectual property rights, rules about cross-border capital flows, investor-state dispute settlement procedures, and harmonization of regulatory standards.

Consider first patents and copyrights (so-called "trade-related intellectual property rights" or TRIPs). TRIPs entered the lexicon of trade during the Uruguay Round of multilateral trade negotiations, which were completed in 1994. The United States has pushed for progressively tighter rules (called TRIPs-plus) in subsequent regional and bilateral trade agreements. Typically TRIPs pit advanced countries against developing countries, with the former demanding stronger and lengthier monopoly restrictions for their firms in the latter's markets. Freer trade is supposed to be win-win, with both parties benefiting. But in TRIPs, the advanced countries' gains are largely the developing countries' losses. Consumers in the developing nations pay higher prices for pharmaceuticals and other research-intensive

products and the advanced countries' firms reap higher monopoly rents. One needs to assume an implausibly high elasticity of global innovation to developing countries' patents to compensate for what is in effect a pure transfer of rents from poor to rich countries.[2] That is why many ardent proponents of free trade were opposed to the incorporation of TRIPs in the Uruguay Round (for example, Bhagwati, Krishna, and Panagariya 2014).

Nonetheless, TRIPs rules have not been dropped, and in fact expand with each new free trade agreement. Thanks to subsequent trade agreements, intellectual property protection has become broader and stronger, and much of the flexibility afforded to individual countries under the original World Trade Organization agreement has been eliminated (Sell 2011).

Second, consider restrictions on nations' ability to manage cross-border capital flows. Starting with its bilateral trade agreements with Singapore and Chile in 2003, the US government has sought and obtained agreements that enforce open capital accounts as a rule. These agreements make it difficult for signatories to manage cross-border capital flows, including in short-term financial instruments. In many recent US trade agreements, such restrictions apply even in times of macroeconomic and financial crisis. This has raised eyebrows even at the International Monetary Fund (Siegel 2013).

Paradoxically, capital account liberalization became a norm in trade agreements just as professional opinion among economists was becoming more skeptical about the wisdom of free capital flows. The frequency and severity of financial crises associated with financial globalization have led many experts to believe that direct restrictions on the capital account have a second-best role to complement prudential regulation and, possibly, to provide temporary breathing space during moments of extreme financial stress. The International Monetary Fund itself, once at the vanguard of the push for capital-account liberalization, has officially revised its stance on capital controls. It now acknowledges a useful role for them where more direct remedies for underlying macroeconomic and financial imbalances are not available. Yet investment and financial services provisions in many free-trade agreements run blithely against this new consensus among economists.

A third area where trade agreements include provisions of questionable merit are the so-called "investor-state dispute settlement" (ISDS) procedures. These provisions have been imported into trade agreements from bilateral investment treaties. They are an anomaly in that they enable foreign investors, and they alone, to sue host governments in special arbitration tribunals and to seek monetary damages for regulatory, tax, and other policy changes that reduce their profits. Foreign investors (and their governments) see ISDS as protection against expropriation, but in

[2] See Diwan and Rodrik (1991) for an attempt to justify TRIPS from the standpoint of developing nations. The history of this paper is of some interest. It was written while I was visiting the World Bank as a junior researcher and at a time when developing nations were strenuously objecting to the US push for TRIPS in the Uruguay Round. The paper was motivated by a challenge that came down to us from the then chief economist of the World Bank. Wouldn't it be nice if someone could make a positive economic case for TRIPS for the developing nations? It turned out someone could.

practice arbitration tribunals interpret the protections provided more broadly than under, say, domestic US law (Johnson, Sachs, and Sachs 2015).

Developing countries traditionally have signed on to investor–state dispute settlement procedures in the expectation that they would compensate for their weak legal regimes and help attract direct foreign investment. But ISDS also suffers from its own problems: it operates outside accepted legal regimes, gives arbitrators too much power, does not follow or set precedents, and allows no appeal. Whatever the merits of ISDS for developing nations, it is more difficult to justify its inclusion in trade agreements among advanced countries with well-functioning legal systems (like the prospective Transatlantic Trade and Investment Partnership between the United States and European countries).[3]

Finally, consider the pursuit of the harmonization of regulatory standards that lies at the center of today's trade agreements. The justification for harmonization is that eliminating regulatory differences among nations reduces the transaction costs associated with doing business across borders. Taking this line of argument one step further, proponents sometimes label regulatory standards abroad that are more demanding than those at home as "non-tariff barriers." There is little question that governments sometimes do deploy regulations to favor domestic producers over foreign ones. But these differences may also reflect dissimilar consumer preferences or divergent regulatory styles. European bans on genetically modified organisms and hormone-fed beef, for example, are rooted not in protectionist motives—the same bans apply to domestic producers as well—but in pressures from consumer groups at home. The US government, for its part, considers them as protectionist barriers, and dispute-settlement panels of the World Trade Organization have often agreed (Euractiv 2006 [updated 2012]).

For economists, the trouble is that unlike in the case of tariffs and quotas, there is no natural benchmark that allows us to judge whether a regulatory standard is excessive or protectionist. Different national assessments of risk—safety, environmental, health—and varying conceptions of how business should relate to its stakeholders—employees, suppliers, consumers, local communities—will produce different standards, none obviously superior to others.

In the language of economics, regulatory standards are public goods over which different nations have different preferences. An optimal international arrangement would trade off the benefits of expanding market integration (by reducing regulatory diversity) against the costs of excessive harmonization. But in general, we have only a hazy idea where that optimal point may lie, which in any case will vary across different policy domains. Perhaps regulators and trade negotiators do their

[3] For some statistics on the use of investor-state dispute settlement procedures, see UNCTAD (2015, chapter III). Of all concluded cases as of end-2014, 27 percent resulted in a ruling in favor of the investor. Twenty-seven percent of the cases were settled, 9 percent were discontinued, and in 2 percent of cases, the state was found in breach but no damages were awarded. In the rest of the cases (36 percent), rulings were in favor of the state. Note that even when the state "wins" in these cases, it is at most awarded its legal costs.

job properly and assess the costs and benefits appropriately, safeguarding room for diversity. Perhaps not.

Regardless, it is curious that economists tend to be nearly unanimous in their view that trade agreements are a good thing. Despite not knowing much about the details, they must believe such agreements regularly strike the right balance in all these areas of ambiguity.[4] Is it that none of these complications matter as long as the agreement is called a "free trade agreement"?

The tendency to associate "free trade agreements" all too closely with "free trade" may result from the fact that the new (and often problematic) beyond-the-border features of these agreements have not yet made their mark on the collective unconsciousness of economists. But I suspect it also results from a certain implicit, hand-waving kind of political economy analysis. In this perspective, protectionist interests are the dominant influence in the determination of trade and other policies. Hence, in the absence of trade agreements, barriers to trade are too high and there is too little trade. Trade agreements are in turn a mechanism through which protectionist interests can be neutralized. The specific details of the agreement do not matter much as long as trade-creating interests are empowered to offset the otherwise dominant protectionist influences. In other words, trade agreements must move us in a desirable direction because they are a counterweight to protectionists.

This inference is valid as long as the argument's premise is correct—namely, that trade agreements on balance empower the special interests more closely aligned with good economic performance. But what if they empower the wrong special interests instead—the investors, banks, and multinational enterprises seeking to increase rents at the expense of the general interest?

When trade agreements are mostly about tariffs and quotas, there is an easy way to tell the difference. The presence of high tariffs before the agreement and tariff reduction as a result of the agreement provide prima facie evidence that protectionists were the dominant influence before the agreement and that they were countervailed through the agreement. But this intuition does not carry over to trade agreements on domestic rules, regulations, and standards because we do not readily know where the efficient benchmark is. A trade agreement captured by an alternative set of special interests may make things worse just as easily as it makes them better. Such an agreement can move us away from the efficient outcome, even if it takes the guise of a free trade agreement and expands the volume of trade and investment.

There is plenty of anecdotal evidence of rent-seeking by firms that favor trade agreements. But to put this evidence in context, let us first examine why countries sign trade agreements in the first place.

[4]An additional area of concern raised by trade specialists early on in the context of NAFTA was the design of the rules of origin, the regulations that determine whether a good imported by one country receives duty-free treatment within the free trade area. Krueger (1993) and others worried that restrictive rules of origin would essentially extend the more protectionist country's tariffs to the other partners. (This concern does not arise in customs unions where countries adopt a common external tariff.)

The Logic of Trade Agreements

When economists teach gains from trade, they emphasize that free trade is good for each nation on its own. (What it means to say "good for the nation" in the presence of losers as well as gainers is, of course, a thorny issue, but I will leave that aside, in keeping with the standard treatment.[5]) Ricardo's (1817) demonstration of the principle of comparative advantage—free trade expands a nation's consumption possibilities frontier even if it has an absolute productivity advantage in producing every good—remains one of our profession's most significant intellectual achievements. A direct implication is that countries should want to have free trade regardless of what their trade partners do. Responding to another country's protectionism by raising one's own trade barriers is tantamount to cutting off the nose to spite the face.

If this insight were the end of the story, the presence (and proliferation) of trade agreements would be a mystifying puzzle. What is the point of signing agreements with other countries to do what is in your national interest in the first place? A possible answer was provided early on by Harry Johnson (1953). Countries that are "large" in world markets have the incentive to exploit their market power. An import tariff restricts home demand for other countries' exports and drives down the world price of the imported good. A Nash equilibrium among large countries would be inefficient, as each country would be imposing its own, positive "optimal" tariffs. Correspondingly, a trade agreement that enforced free trade could leave all the countries better off.

Even if the logic of this argument is accepted, the question remains of why a *formal* trade agreement is needed, such as the World Trade Organization or NAFTA. After all, a free-trade equilibrium can be achieved through cooperation in a repeated interaction game. In addition, one can ask whether a formal agreement on its own can prevent opportunistic behavior on the part of sovereign nations. Nonetheless, the motive to manipulate the terms-of-trade provides a valid economic motive for countries to commit themselves to free trade by signing on to trade agreements.[6]

However, this theory does not sit well with the fact that actual policymakers do not seem very concerned about the terms of trade when they negotiate trade agreements. They tend to care more about the volume of trade: nations like it when their exports grow, but not so much when their imports expand. Effectively, nations trade market access: more of your imports in return for more of my exports. Moreover, these preferences do not seem to be grounded in the effects that trade volumes have on world market prices. It is true that home policies that lower import demand tend to reduce world market prices of imports, and hence improve the terms of

[5] However, Driskill (2012) takes the profession to task, correctly, for sweeping distribution under the rug when discussing the "welfare gains" from trade.

[6] See Grossman (2016) for an exposition of the Johnson argument and the subsequent literature. Bagwell and Staiger (2002) have been the most consistent and prolific defenders of this perspective on trade agreements.

trade. But on the export side, general government practice consists of boosting export supply, through export subsidies, credits, and other assistance, rather than reducing it. This has the effect of lowering export prices on world markets, and hence worsening the terms of trade.

Also, if trade agreements are really about curbing terms-of-trade manipulation, what do we make of the prohibition on export subsidies in the World Trade Organization? When a government resorts to export subsidies, it worsens its own terms of trade and confers economic benefits on other nations. If it does so nevertheless, it must be for noneconomic or special-interest reasons. Regardless, there would be no reason for trade agreements to prohibit their use. As Grossman (2016) notes, "the literature offers no compelling reason why trade agreements should outlaw export subsidies in a trading environment characterized by perfectly competitive markets."[7]

Trade policy practitioners seem to worry little about international terms-of-trade spillovers. Instead, they tend to justify trade agreements by reference to the politics of trade policy at home: Trade agreements are what enable governments to say "no" to domestic import-competing interests. Absent trade agreements, this argument goes, governments are too easily tempted to do the easy thing and provide import protection when faced with short-term political pressures (for example, Bown 2016).

A number of academic papers conceptualize this argument in the form of a time-inconsistency problem (for example, Staiger and Tabellini 1987; Maggi and Rodriguez-Clare 1998). In this framework, the government knows that free trade is the best policy in the long run. But it faces short-term political pressures to respond to organized interest groups. Forward-looking workers and capitalists understand the difference between the government's short-run and long-run incentives and behave accordingly. In particular, they make their investment decisions so as to ensure the government provides them with trade protection. In these settings, trade agreements are a commitment device for governments to withstand political pressure from future protectionists. As Grossman (2016) notes, we may question whether there is not an easier way of purchasing such commitment than negotiating very complicated deals with multiple partners over many years. Nevertheless, the view that trade agreements serve to neutralize protectionist special interests is very widely held.

This commitment or lock-in argument is analogous to the familiar case for policy delegation in other areas with dynamic inconsistency, such as monetary policy (justifying an independent central bank) or business regulation (justifying autonomous regulatory agencies). In any of these settings, the validity of the policy conclusion depends critically on the specification of the game that is being played between the government and special interests.

When there is a genuine time consistency problem, everyone is better off with pre-commitment or delegation (save, possibly for the lobbyists and special interests). When protectionists show up at the government's door, the government says:

[7] Under imperfect competition, countries may have an economic incentive to use export subsidies to shift excess profits from foreign firms to domestic firms. See Grossman (2016, section 3) for a discussion.

"Sorry, I'd love to help you out, but the trade agreement will not let me do it." This is the good kind of delegation and external discipline.

Now consider a different setting. Here, the government fears not its future self, but its future *opponents*: the opposition party (or parties). The latter may have different views on economic policy, and if victorious in the next election, the opposition may well choose to shift course. In this situation, an incumbent government enters an international agreement to tie the hands of its opponents. From the standpoint of social welfare, this strategy has much less to recommend itself. The future government may have better or worse ideas about government policy, and it is not clear that restricting what it can do in the future is a win-win outcome. This government too will present its case in traditional delegation terms. But what it is really doing is to ensure the permanence of partisan policies.[8]

Now suppose further that the current government is captured by special interests—but by exporter lobbies instead of import-competing lobbies. In this case, the government's objectives are explicitly redistributive, to transfer rents from the rest of society to a special interest. But unlike in the usual model, the rent-seekers are not the traditional protectionists. They are pharmaceutical companies seeking tighter patent rules, financial institutions that want to limit ability of countries to manage capital flows, or multinational companies that seek special tribunals to enforce claims against host governments. In this setting, trade agreements serve to empower special interests, rather than rein them in.

Whose Interests Do Trade Agreements Serve?

With traditional trade agreements, which focused on reducing tariff and nontariff barriers to trade, it was relatively easy to figure out which of these different models approximated reality better. Consider for example the GATT (General Agreement on Tariffs and Trade) rounds of multilateral trade negotiations before the World Trade Organization was established in 1995. Tariff levels were high after World War II, and negotiations were largely about bringing them down. Few other issues were discussed beyond tariffs and other explicit barriers at the border. The fact that tariffs were high to begin with is prima facie evidence that protectionist interests had previously held the upper hand in the political equilibrium. The fact that trade agreements succeeded in lowering tariffs is evidence that such agreements served to counteract those protectionist interests. In other words, the trade-agreements-as-political-commitment story worked pretty well. It suggests that these agreements were moving the economies of the negotiating parties broadly in the right direction.

[8] Of course, if trade deals are the outcome of partisan politics, there will be pressure in the future to renege on them, once political power changes hands. But this is no different than in the standard time-inconsistency case, where short-term incentives always militate in favor of reneging on trade agreements. In both cases, the argument relies on the presence of costs that render international agreements hard to reverse.

With post-1995 trade agreements, matters are no longer so simple. Tariffs and explicit barriers to trade have dropped considerably, and many new areas of negotiation have opened up in which there is typically no efficient "free-trade" benchmark analogous to the role that zero duties play in the context of tariffs. Do Vietnam's capital-account regulations, say, or patent rules serve the country's economic development well or poorly? Are European Union food safety regulations closely aligned with European consumers' risk preferences or do they privilege producer interests too much? Does US jurisprudence provide adequate protections for foreign investors, or not? To be sure, domestic regulations and product standards can be enacted for protectionist purposes—simply to keep competing imports out. But they can be also used to serve developmental, social, or other deserving goals.

If countries have gotten the balance wrong in these and other areas, can we be at all sure that trade agreements such as the Trans-Pacific Partnership or the Transatlantic Trade and Investment Partnership will move their policies closer to the social optimum—and not further away? Can the dispute settlement process provided by such trade agreements draw the appropriate distinctions in practice between pure protectionism and legitimate regulatory divergence?[9]

It is hard to provide a definitive answer to these questions. What is clear is that we cannot simply look at whether agreements are trade-creating or not and evaluate them on that basis. It is all too easy to come up with examples where too much regulatory harmonization in the name of reducing transaction costs to trade leaves at least one of the negotiating parties worse off. The case of tightening intellectual property rights in developing countries, mentioned earlier, is a prominent example. Erosion of consumer protections in high-standard countries may likewise expand trade, but it will not leave importing countries better off. It is similarly easy to see that agreements that privilege investors or corporations over other interests (like labor or the environment) can end up producing largely redistributive consequences with few efficiency gains. That fear is widespread among opponents of investor courts.

Potential trade-offs arise in all of these areas: regulatory harmonization may spur trade, but it could also prevent regulations from reflecting domestic preferences. A proper negotiating process would take both sides of the ledger into account. The texts of trade agreements pay plenty of lip service to economic and social goals beyond trade. However, these are fundamentally *trade* deals. They are not negotiations on public health, regulatory experimentation, promoting structural change and industrialization in developing nations, or protecting labor standards in the advanced economies. It would not be surprising if the process were captured by

[9]See Sykes (2017) for some of the difficulties. Howse and Tuerk (2001) provide an early discussion of WTO jurisprudence, drawing attention to the risk that WTO rules can be used to challenge domestic regulations aimed at addressing serious health risks. The specific case they discuss is a case brought by Canada against France's ban on asbestos in construction materials. Even though the WTO panel ultimately ruled against Canada, it accepted the basis of the claim, namely that asbestos and non-asbestos materials were "like products" and that therefore France had discriminated against Canadian imports.

trade interests. Nor should it be unexpected that the success of the agreements is typically gauged by the volume of trade they create.

We could gain further insight into specific outcomes by looking at the actual process through which trade agreements are negotiated. However, such negotiations are typically secret—a feature that draws the ire of labor, public-interest groups, and many politicians. During the Trans-Pacific Partnership negotiations, for example, only two copies of the text were made available in special reading rooms to US congressmen and their staff with special security clearances (Bradner 2015). And even these readers could be prosecuted if they revealed the contents to the public.

The ostensible reason for secrecy is to facilitate the back-and-forth dealing needed to produce compromise. But from the perspective of broader social welfare, secrecy is a mixed blessing. It may promote quicker bargains. But it also tends to bias the results against interests not present in the negotiation (Kucik and Pelc 2016). Business is rarely far from the actual negotiations. In fact, it is commonplace for business lobbyists to wait just outside the negotiation room and influence the outcome in real time (for example, see the account of NAFTA negotiations in Smith 2015).

One of the better-known and most instructive cases is the story of trade-related intellectual property rights, or TRIPs. The inclusion of TRIPs in the 1994 agreement that established the World Trade Organization was a landmark event. As Devereaux, Lawrence, and Watkins (2006, p. 42) write, "[a]fter seven years of negotiating, industries that rely on copyrights, patents, and trademarks received more protection than anyone had believed possible at the outset of the talks." Business interests had been pushing since at least the 1970s to get the US government to enforce patent and copyright protections abroad. The conventional international forum for discussion of such issues was the World Intellectual Property Organization (WIPO). However, US firms regarded WIPO as an ineffective UN agency dominated by developing countries. A coalition made up of agrochemical companies like Monsanto, trademark-based companies such as Levi-Strauss and Samsonite, pharmaceutical companies like Pfizer, and computer companies such as IBM effectively redefined TRIPs as a trade issue. They managed to engineer what political scientists call "forum shifting," moving the focus of international negotiation from WIPO to what would eventually become the World Trade Organization. US firms coordinated with their counterparts in Europe and Japan to develop minimum standards on which they would agree. These standards would in turn significantly shape the final agreement that emerged from the Uruguay Round (Devereaux, Lawrence, and Watkins 2006; Sell 2011).

The shift in forums from the World Intellectual Property Organization to the World Trade Organization was a brilliant strategic move for business. It ensured that commercial considerations would dominate and outweigh other goals, such as implications for economic development and public health. But TRIPs was only the beginning. Following their success with using the Uruguay Round to pursue their goals for protection of intellectual property, pharmaceutical and other companies engaged in what Sell (2011) calls "vertical" forum shifting—that is, pushing for and obtaining further protections in specific free trade agreements. The United States

had much greater bargaining leverage vis-à-vis individual developing nations in bilateral or regional trade agreements. Once a precedent had been set in the WTO talks, it was now possible to go considerably beyond TRIPs. Pharmaceutical companies were able to obtain test data exclusivity (preventing generics providers from using the same data in their own licensing applications), a prohibition on parallel imports (of original products, but by other than the patent-holders), severe restrictions on compulsory licensing requirements by host governments, and automatic patent term extensions (Sell 2011). The Trans-Pacific Partnership (TPP), the latest of these agreements, has such broad protection of intellectual property that the head of the World Health Organization has spoken out against it, blaming interference by "powerful economic operators" (Germanos 2015).[10]

The influence of special interests is rarely exercised through the naked application of power—do this, or else! Instead, these groups get their way by convincing policymakers and the broader public that certain of their goals also further the public interest. The success of TRIPs had to do in no small part with the framing of the issue in terms that gave it broad legitimacy and appeal. Thus, what might have been more accurately called monopoly rents were transformed into "property rights." Then firms abroad who imitated and reverse-engineered technologies of the more advanced countries became engaged in "piracy," even though this is a time-honored practice by technologically lagging countries, including today's developed countries in the past. As one prominent example, many of Boston's original textile mill owners "stole" their designs from Lancashire, taking elaborate steps to evade British intellectual property rights protections (Morris 2012). With some hand-waving, preserving the monopoly of US film studios, pharmaceutical companies, and fashion houses turned into a fundamental issue of free trade.

Pro-trade business interests are known to have played a significant role in the expansion of trade agreements into other new areas beyond intellectual property. For example, the push to include services in multilateral trade negotiations took place at the behest of American firms. Services differ from trade in goods insofar they often require changes in domestic regulations. Financial services is a good example. As Marchetti and Mavroidis (2011, p. 692) write, "it was the US financial

[10]Even though US pharmaceutical companies would have gotten a better deal under the Trans-Pacific Partnership than they ever had, they were dissatisfied at the end because they could not get the other countries to agree to the 12-year protection in biologics that they currently enjoy in the United States. They got a minimum of five years instead, with three additional years under special regulatory safeguards. Mark Grayson, a spokesman for PhRMA, which represents top pharmaceutical companies, is quoted as saying: "They were supposed to come back with U.S. law on [intellectual property] rights, and they didn't, and our board is very disappointed" (Ferris 2015). For another perspective, Branstetter (2016) concludes that TPP struck a reasonable balance between incentives to innovation and access to medicines. He praises the agreement for exporting the US regulatory model (specifically, the Drug Price Competition and Patent Term Restoration Act of 1984, often known Hatch–Waxman Act) to the Asia-Pacific context. But given that the parties to the trade agreement are at very different stages of development, it is not clear that a similar model is suitable for all, rather than being a mechanism for rent shifting from less-developed to more-developed countries. Moreover, empirical research fails to find strong effects on innovation from more restrictive patents (Boldrin and Levine 2012; Moser 2013; Sakakibara and Branstetter 1999; Branstetter 2004).

services sectors that first argued systematically in favor of a trade round that would include a chapter on liberalization of trade in services." A key role was played by the Coalition of Service Industries, a trade group representing US service industries, which focused its energies on the right of establishment of financial and insurance companies in foreign countries: "The CSI gathered data, organized conferences, engaged in extensive public lecturing, and heavily lobbied the US government to this effect" (p. 693). The heads of Citibank and American Express each headed key advisory groups organized by the US Trade Representative in the run-up to the Uruguay Round agreement in 1994. American Express was especially active, with its executives building up a domestic lobby, establishing links with other service-industry lobbies around the world, and exerting influence on US policy through direct participation in negotiations with other countries (Yoffie 1990, cited in Marchetti and Mavroidis 2011).

Interestingly, this lobbying to expand markets abroad in services has not done much to nullify service protectionism in the United States itself—unlike in the traditional account of what trade agreements do. For example, one of the most blatant forms of US protectionism is the Merchant Marine Act of 1920 (often known as the Jones Act), which prevents foreign ships from serving domestic US shipping lines. The objective of the law is to maintain a strong US shipbuilding industry and merchant marine, ostensibly for national security purposes. But the protectionist intent and consequences are clear (Grennes 2017). It has remained untouched in all the trade agreements the United States has negotiated, including the Trans-Pacific Partnership.

As trade agreements move into these new areas, the role of business lobbies changes as well. Governments have to rely on knowledge and expertise from businesses to negotiate complex regulatory changes. Hence, business lobbies become partners and collaborators for the trade negotiators: they help define the issue, provide information and expertise, and mobilize support from other business groups transnationally. As Woll and Artigas (2007, p. 131) put it, "[u]nlike the exchange model assumed in the traditional economic models, firms do not just exchange votes or money to lobby against regulation. Rather, they offer expertise and political support in exchange for access to the elaboration of specific stakes." Business lobbies also become much more intimately involved in the actual trade negotiations, sometimes forming a larger part of the delegation than the actual government representatives (Woll and Artigas 2007).

A rough idea of who actually lobbies for trade agreements can be obtained from data collected by the Sunlight Foundation for the Trans-Pacific Partnership. Their analysis is based on public lobbying reports issued by corporations and industry associations, and whether the TPP is mentioned by name in those reports (as reported by Drutman 2014). Perhaps unsurprisingly, pharmaceutical manufacturing firms and PhRMA (the industry association) dominate the list. Others that stand out are auto manufacturers, milk and dairy producers, textiles and fabrics firms, information technology firms, and the entertainment industry. Labor unions such as United Steelworkers and AFL-CIO, which are traditionally

associated with protectionist motives, tend to lag behind these industry-based groups.

Business interests exert influence also through their presence in the various trade advisory committees that are set up in the course of trade negotiations. Such committees are in principle made up of a wide range of all the stakeholders, including labor groups and environmental nongovernment organizations who may have a negative view of conventional trade agreements. But business representatives and trade associations are by far the dominant group, making up more than 80 percent of the membership of such committees during the negotiation of Trans-Pacific Partnership (as reported in Ingraham 2014; Ingraham and Schneider 2014).

Systematic studies of how interest groups on different sides of trade agreements shape the negotiations are rare, given the lack of transparency of the process. However, one analysis of Swedish lobbies takes advantage of the fact that Sweden has a far-reaching freedom-of-information clause in its constitution, which enabled Rönnbäck (2015) to access all the documents behind trade policy formulation in the country during the Uruguay Round. As Rönnbäck points out, the commonly maintained assumption in the literature on the political economy of trade is that the process is influenced overwhelmingly by import-competing, protectionist interests. Trade agreements are signed *despite* these interests, not because of them. But Rönnbäck found that the approach pursued by the Swedish government in the trade negotiations was not only in line with special-interest lobbying, but it was largely shaped by it. The interest groups that played the determining role in the consultative process were in favor of expanding trade. But the interests of these groups were not in tariffs per se, which were already low. Instead their demand was "to broaden the scope of the agenda of the GATT, by including issues such as trade in services, investment measures and public tender agreements" (p. 286). In other words, industry lobbies pushed for deep integration measures beyond the standard free-trade policies.

Rönnbäck's (2015) study also documents how trade negotiations can help special interests coordinate across national borders. Apparently Swedish businesses initially did not show much awareness—or interest in—intellectual property rights. But as the United States pushed harder on TRIPs in the negotiations, the issue rose in prominence among Swedish interest groups. As Rönnbäck (p. 287) puts it: "It seems as if the interest groups only realized the potential for economic rents that the trade negotiations could offer quite slowly, as the negotiations progressed. As soon as the Swedish interest groups realized this potential, however, they did not hesitate to act and make demands on the government." As individual corporations such as Astra as well as the Pharmaceutical Industry Association took up the cause, Sweden's government followed suit. By the late 1980s, any doubts the Swedish government may have harbored about the wisdom of including TRIPs in the Uruguay Round seems to have vanished. In the years leading to the final agreement, the intellectual property issue came to be "described as among the most important for the Swedish interests" (p. 288).

Finally, the influence of special interests also shows up in the dog that does not bark: potential areas of negotiation with high social returns that are left out of the trade agenda. One such area that touches directly the interests of large firms is global tax-and-subsidy competition. In a world with mobile capital, governments are tempted to offer better terms to globally mobile corporations in order to compete for investment. This results in a sub-optimal Nash equilibrium with larger transfers to corporations and their shareholders than is globally desirable. In practice, the effects show up in two areas: investment subsidies (in the form of tax holidays and other sweeteners) and reductions in corporate tax rates. In view of the obvious cross-border externalities, enacting global disciplines on tax-and-subsidy competition would make excellent economic sense. Yet trade agreements never touch on this issue. They are replete with restrictions on what home governments can do to impose obligations on foreign investors. But they do not prevent these governments from wasting tax dollars and enriching corporations in a harmful race to the bottom.

Ammunition to the Barbarians?

When I recently gave a talk arguing that economists underplay some of the adverse consequences of advanced globalization, an economist in the audience took me to task: Don't you worry, he asked, that your arguments will be used (or abused) by populists and protectionists to further their own interests? It is a reaction that reminds me of a response from a distinguished economist more than two decades ago to my 1997 monograph *Has Globalization Gone Too Far?* All your arguments are fine, he told me, but they will give "ammunition to the barbarians."

The objection is instructive insofar as it lays bare the implicit political economy understanding with which economists tend to approach public discussions of trade policy. In this perspective, the serious threats to sensible trade policy nearly always come from the import protectionists, and trade agreements mainly offset the influence of the protectionists. But as trade agreements have evolved and gone beyond import tariffs and quotas into regulatory rules and harmonization—intellectual property, health and safety rules, labor standards, investment measures, investor-state dispute settlement procedures, and others—they have become harder and harder to fit into received economic theory. Why do many economists presume that it is more dangerous to express skepticism in public about these rules than it is to cheerlead? In other words, why do they think that there are barbarians only on one side of the issue?

I have presented an alternative perspective in this paper. Rather than neutralizing the protectionists, trade agreements may empower a different set of rent-seeking interests and politically well-connected firms—international banks, pharmaceutical companies, and multinational firms. They may serve to internationalize the influence of these powerful domestic interests. Trade agreements could still result in freer, mutually beneficial trade, through exchange of market access. They could result in the global upgrading of regulations and standards, for labor,

say, or the environment. But they could also produce purely redistributive outcomes under the guise of "freer trade." As trade agreements become less about tariffs and nontariff barriers at the border and more about domestic rules and regulations, economists might do well to worry more about the latter possibility. They may even adopt a stance of rebuttable prejudice against these new-type trade deals—a prejudice against these deals, which should be overturned only with demonstrable evidence of their benefits.

■ *I am grateful to Robert Lawrence, Gene Grossman, and the editors for helpful comments that improved the paper, and to Vedant Bahl for research assistance.*

References

Bagwell, Kyle, and Robert W. Staiger. 2002. *The Economics of the World Trading System.* Cambridge, MA: MIT Press.

Bhagwati, Jagdish N., Pravin Krishna, and Arvind Panagariya. 2014. "The World Trade System: Trends and Challenges." May. Unpublished paper, Department of Economics, Columbia University.

Boldrin, Michele, and David K. Levine. 2012. "The Case Against Patents." Working Paper 2012-035A, Economic Research, Federal Reserve Bank of St. Louis, September.

Bradner, Eric. 2015. "How Secretive is the Trans-Pacific Partnership?" *CNN*, June 12. http://www.cnn.com/2015/06/11/politics/trade-deal-secrecy-tpp/index.html.

Branstetter, Lee G. 2004. "Do Stronger Patents Induce More Local Innovation?" *Journal of International Economic Law* 7(2): 359–70.

Branstetter, Lee. 2016. "TPP and the Conflict over Drugs: Incentives for Innovation versus Access to Medicines." Chap. 13 in *Assessing the Trans-Pacific Partnership, Vol. 2: Innovations in Trading Rules,* edited by Jeffrey J. Schott and Cathleen Cimino-Isaacs. PIIE Briefing 16-4. Washington, DC: Peterson Institute for International Economics, Washington.

Bown, Chad P. 2016. "The Truth about Trade Agreements and Why We Need Them." Peterson Institute for International Economics, Washington, DC, November 21. https://piie.com/commentary/op-eds/truth-about-trade-agreements-and-why-we-need-them.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82(1): 1–44.

Devereaux, Charan, Robert Z. Lawrence, and Michael D. Watkins. 2006. *Case Studies in US Trade Negotiation*, Vol. 1: *Making the Rules.* Washington, DC: Institute for International Economics.

Diwan, Ishac, and Dani Rodrik. 1991. "Patents, Appropriate Technology, and North–South Trade." *Journal of International Economics* 30(1–2): 27–47.

Driskill, Robert. 2012. "Deconstructing the Argument for Free Trade: A Case Study of the Role of Economists in Policy Debates." *Economics and Philosophy* 28(1): 1–30.

Drutman, Lee. 2014. "How Big Pharma (and Others) Began Lobbying on the Trans-Pacific Partnership before You ever Heard of It." Sunlight Foundation, March 13, https://sunlightfoundation.com/2014/03/13/tpp-lobby.

Euractiv. 2006 [updated 2012]. "EU GMO Ban Was Illegal, WTO Rules." May 12, 2006 [updated: Nov 5, 2012]. https://www.euractiv.com/section/trade-society/news/eu-gmo-ban-was-illegal-wto-rules/.

Ferris, Sarah. 2015. "Pharma Flap Imperils President's Trade Deal." *The Hill*, October 20. http://thehill.com/policy/healthcare/lobbying/257386-pharma-flap-imperils-presidents-trade-deal.

Germanos, Andrea. 2015. "WHO Head Joins Chorus Denouncing TPP's Blow to Public Health, Boon to Big Pharma." *Common Dreams*, November 12. https://www.commondreams.org/

news/2015/11/12/who-head-joins-chorus-denouncing-tpps-blow-public-health-boon-big-pharma.

Grennes, Thomas. 2017. "An Economic Analysis of the Jones Act." May 2, Research Paper/Study, Mercatus Center, George Mason University.

Grossman, Gene M. 2016. "The Purpose of Trade Agreements." Chap. 7 of *Handbook of Commercial Policy*, edited by Kyle Bagwell and Robert Staiger. Amsterdam: Elsevier.

Hakobyan, Shushanik, and John McLaren. 2016. "Looking for Local Labor Market Effects of NAFTA." *Review of Economics and Statistics* 98(4): 728–41.

Howse, Robert, and Elisabeth Tuerk. 2001. "The WTO Impact on Internal Regulations—A Case Study of the Canada–EC Asbestos Dispute." Chap. 12 of *The EU and the WTO: Legal and Constitutional Issues*, edited by Gráinne de Búrca and Joanne Scott. Oxford and Portland: Hart Publishing.

Ingraham, Christopher. 2014. "Interactive: How Companies Wield Off-the-Record Influence on Obama's Trade Policy." *Washington Post*, February 28.

Ingraham, Christopher, and Howard Schneider. 2014. "Industry Voices Dominate the Trade Advisory System." *Washington Post*, February 27.

Johnson, Harry G. 1953. "Optimum Tariffs and Retaliation." *Review of Economic Studies* 21(2): 142–53.

Johnson, Lise, Lisa Sachs, and Jeffrey Sachs. 2015. "Investor-State Dispute Settlement, Public Interest and U.S. Domestic Law." CCSI Policy, Columbia Center for Sustainable Investment, Columbia University, May.

Krueger, Anne O. 1993. "Free Trade Agreements as Protectionist Devices: Rules of Origin." NBER Working Paper 4352, April.

Kucik, Jeffrey, and Krzysztof J. Pelc. 2016. "Measuring the Cost of Privacy: A Look at the Distributional Effects of Private Bargaining." *British Journal of Political Science* 46(4): 861–89.

Lawrence, Robert Z. 1996. *Regionalism, Multilateralism, and Deeper Integration*. Washington, DC: Brookings Institution.

Limão, Nuno. 2016. "Preferential Trade Agreements." NBER Working Paper 22138, March.

Maggi, Giovanni, and Andres Rodriguez-Clare. 1998. "The Value of Trade Agreements in the Presence of Political Pressures." *Journal of Political Economy* 106(3): 574–601.

Marchetti, Juan A., and Petros C. Mavroidis. 2011. "The Genesis of the GATS (General Agreement on Trade in Services)." *European Journal of International Law* 22(3): 689–721.

Morris, Charles, R. 2012. "We Were Pirates, Too." *Foreign Policy*, December 6.

Moser, Petra. 2013. "Patents and Innovation: Evidence from Economic History." *Journal of Economic Perspectives* 27(1): 23–44.

Ricardo, David. 1817. *The Principles of Political Economy and Taxation*. London: John Murray.

Rönnbäck, Klas. 2015. "Interest-group Lobbying for Free Trade: An Empirical Case Study of International Trade Policy Formation." *Journal of International Trade & Economic Development* 24(2): 281–93.

Sakakibara, Mariko, and Lee Branstetter. 1999. "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms." NBER Working Paper 7066, April.

Sell, Susan K. 2011. "TRIPs Was Never Enough: Vertical Forum Shifting, FTAS, ACTA, and TPP." *Journal of Intellectual Property Law* 18(2): 447–78.

Siegel, Deborah E. 2013. "Capital Account Regulations and the Trading System." Chap. 6 in *Capital Account Regulations and the Trading System: A Compatibility Review*, edited by Kevin P. Gallagher. Pardee Center Task Force Report, Boston University, March.

Smith, Tacey Vanek. 2015. "Remembering NAFTA Gives Insight Into Why Trade Deals Are Kept Secret." *NPR*, June 19. http://www.npr.org/2015/06/19/415809462/remembering-nafta-gives-insight-into-why-trade-deals-are-kept-secret.

Staiger, Robert W., and Guido Tabellini. 1987. "Discretionary Trade Policy and Excessive Protection." *American Economic Review* 77(5): 823–37.

Sykes, Alan O. 2017. "Regulatory Consistency Requirements in International Trade." *Arizona State Law Journal* 49(Special Issue): 821–66.

United Nations Conference on Trade and Development (UNCTAD). 2015. *World Investment Report 2015: Reforming International Investment Governance*. Geneva, Switzerland: UNCTAD

Woll, Cornelia, and Alvaro Artigas. 2007. "When Trade Liberalization Turns into Regulatory Reform: The Impact on Business–Government Relations in International Trade Politics." *Regulation & Governance* 1(2): 121–38.

Yoffie, David B. 1990. "Trade in Services and American Express." Case 23 in *International Trade and Competition: Cases and Notes in Strategy and Management*, 1st edition, by David B. Yoffie, 367–86. McGraw-Hill.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 91–114

Modeling Risk Aversion in Economics

Ted O'Donoghue and Jason Somerville

Economists have a shared preconception that, for the most part, people dislike risk. We typically assume that a person who is offered a choice between a risky lottery versus a sure payment equal to the expected value of that lottery will choose the latter. Similarly, when a person compares two lotteries with equal expected values, we assume that the person will choose the lottery with less risk. When an individual compares two lotteries where one has a higher expected value but also more risk, we assume that the person's choice will depend on the extent of risk aversion—for example, if risk aversion is small enough, the person will choose the lottery with higher expected value and more risk.

This risk-aversion intuition is a key driver in many prominent economic applications. Risk aversion creates a demand for insurance, which gives rise to a large economics literature on health insurance, unemployment insurance, property insurance, flood insurance, and so forth. Risk aversion plays a central role in financial investment, driving the key trade-off between risk and return in the pricing of financial assets. Risk aversion is relevant in principal–agent models, and is the source of the incentives–insurance trade-off that commonly arises in such models. Risk aversion is also important in life-cycle models as people face risk concerning employment, income, asset returns, health, and so forth.

To capture the risk-aversion intuition, the standard approach in economics has been to utilize the model of expected utility, in which risk aversion derives from

■ *Ted O'Donoghue is Zubrow Professor of Economics and Jason Somerville is a PhD student in economics, both at Cornell University, Ithaca, New York. Their email addresses are edo1@cornell.edu and js2946@cornell.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.91

doi=10.1257/jep.32.2.91

diminishing marginal utility for wealth (or diminishing marginal utility for aggregate consumption). The expected utility model is extremely tractable and portable into a wide variety of applications and has been used to derive many important insights. Moreover, expected utility permits a measure of the degree of risk aversion, and thus makes it possible to obtain quantitative estimates of risk aversion, or to develop (and estimate) models with heterogeneity in the degree of risk aversion.

But over the years, economists and psychologists have identified various issues with expected utility as a descriptive model of choice. In this journal, Rabin and Thaler (2001) highlight how the structure of expected utility generates a calibration problem in dealing with stakes of different sizes, and as a result expected utility does not permit seemingly plausible preferences—for example, it does not permit people to exhibit noticeable but modest risk aversion for both small and moderate stakes. Perhaps more importantly, researchers have pointed out predictions of expected utility that do not seem to accord with empirical observation, and over the past two decades a number of economists have pursued alternative ways to model risk aversion.

Our goal in this article is to urge economists to take seriously the research agenda of developing and assessing different ways to model risk aversion. We proceed in three main steps. First, whereas many economists seem to take expected utility with diminishing marginal utility for wealth to be synonymous with risk aversion, we highlight that the basic intuition of risk aversion that drives many results in economics is not intimately tied to expected utility. Second, we describe a few alternative models that can also capture the basic intuition of risk aversion. Finally, we discuss that, while expected utility and the alternative models might all capture the basic intuition of risk aversion, the alternative models can generate additional, more nuanced implications not shared with expected utility, that in some cases seem to be borne out by data. As we will highlight, these alternative models also are not perfect, and further research is needed to identify even better approaches.

Expected Utility and Risk Aversion

An option that involves risk can be described by a lottery, which is a list of possible outcomes along with the probability associated with each outcome. A choice between risky options can thus be thought of as a choice between lotteries, and to model how people make such choices, we need a model of how people evaluate and compare lotteries. Expected utility is exactly such a model.

According to expected utility, a person has a utility function that assigns a "utility" to each outcome. The overall evaluation of a lottery is then a weighted average of the utility from each possible outcome, where the weight attached to each utility is simply the probability of that outcome occurring. In other words, the person is assumed to choose the option that yields the largest expectation of utility.

The concept of *risk aversion* is typically applied when a person is choosing between lotteries where the outcomes are expressed in monetary amounts. For

instance, a person might face a choice whether to accept or reject a 50:50 gamble to lose \$10 or win \$10. A common definition of risk aversion is that, for any lottery, a person prefers a sure payment equal to the expected value of the lottery to facing the lottery itself. Under this definition, because a 50:50 gamble to lose \$10 or win \$10 has an expected value of \$0, a risk-averse person would reject this lottery.

Expected utility yields a simple and elegant explanation for risk aversion: under expected utility, a person is risk-averse—as defined in the prior paragraph—if and only if the utility function over monetary wealth is concave. In other words, risk aversion derives from diminishing marginal utility for monetary wealth. Expected utility therefore attributes the decision to reject the 50:50 gamble to lose \$10 or win \$10 to the idea that the utility decline from having one's wealth be \$10 smaller is larger than the utility increase from having one's wealth be \$10 larger.

Expected utility becomes especially useful when, unlike in the example above, a person is choosing between lotteries with different expected values. For instance, now consider a 50:50 gamble to lose \$10 or win \$12, which has an expected value of \$1. The decision whether to accept or reject this lottery involves a trade-off: accepting the gamble means taking on risk, but it also means a higher expected value. Expected utility yields a way to resolve this trade-off. Specifically, it permits a measure of the degree of risk aversion such that the person will accept the gamble if risk aversion is small enough, and otherwise the person will reject the gamble.[1]

In applications, economists often use a specific functional form for the utility function over wealth. One prominent functional form is the constant relative risk aversion utility function, in which there is a single parameter, ρ, that captures the degree of a person's risk aversion.[2] In the example above, a person with prior wealth \$50,000 would accept the 50:50 gamble to lose \$10 or win \$12 as long as ρ is smaller than 831. To help illustrate ideas, we often describe the implications of the constant relative risk aversion utility function in the examples below.

More generally, the existence of a measure of a person's risk aversion makes the model quite powerful. It permits analyses that use data on observed choices to estimate an individual's risk aversion or the distribution of risk aversion in a population. Perhaps more importantly, it also makes the model quite portable in that, once one has estimated a person's degree of risk aversion in one domain—for example, with the constant relative risk aversion utility function and an estimate of a person's ρ—the model makes predictions for how that person would behave in other domains.

To further highlight the basic intuition of risk aversion, and how it is not intimately tied to the expected utility model, we next describe how this intuition operates in three classic textbook domains: insurance, financial assets, and agency.

[1] Two common measures are the coefficient of absolute risk aversion and the coefficient of relative risk aversion, both defined by Pratt (1964) and Arrow (1965).

[2] The functional form is $u(x) = x^{1-\rho}/(1-\rho)$ for $\rho \neq 1$ and $u(x) = \ln x$ for $\rho = 1$.

Insurance

Suppose an individual is exposed to a potential loss of $10,000 that has a 5 percent chance of occurring, but also has the option to purchase full insurance—that is, an insurance policy that pays out $10,000 in the event that the loss occurs—for a premium of $600. This person is choosing between two lotteries. If the person does not buy insurance, then this person faces a lottery in which there is a 5 percent chance that wealth will decline by $10,000, and a 95 percent chance that wealth is unaffected—and thus, in expectation, wealth will decline by $500. If the person purchases insurance, then wealth declines by $600 with certainty.

In this domain, the intuition of risk aversion implies that a person who dislikes risk should be willing to give up some expected value in order to reduce risk and thus the willingness to pay for insurance should be larger than the actuarially fair price. In the example above, the intuition of risk aversion implies the person should be willing to pay more than $500 for full insurance. Moreover, a person who is more averse to risk will be willing to pay more for insurance. Whether this individual is willing to pay the premium of $600 will depend on whether the individual is sufficiently risk-averse.

Extending this example, suppose the person can choose how much insurance to purchase where insurance is priced linearly at a rate of 6 cents per dollar of insurance. In other words, this person can insure against the full loss for $600, insure against half the loss for $300, insure against a quarter of the loss for $150, and so forth. The individual now faces a choice not just between two lotteries, but rather between an array of lotteries with different coverage levels. This decision involves a trade-off: purchasing more insurance reduces risk, but it also reduces expected value. The intuition of risk aversion does not say where a person should end up, but it does suggest that an individual who is more risk-averse should choose a higher coverage level.

The basic conclusions above follow from the simple intuition of risk aversion, and, importantly, they do not require the expected utility model. The value of applying expected utility is that it can yield more quantitative statements—for example, with the constant relative risk aversion utility function, a person with $\rho = 5$ and prior wealth $50,000 would choose to cover $8,126 of the potential $10,000 loss. More importantly, expected utility permits analysis of many more nuanced questions, such as the nature of optimal risk sharing, the role of deductibles or coinsurance in combating moral hazard, and so forth. It also permits structural empirical analyses—for instance, in the context of auto insurance, Cohen and Einav (2007) apply expected utility to study heterogeneity in risk aversion and the implications of this heterogeneity for insurance pricing.

Financial Investment

Suppose an individual must decide how to divide personal wealth between two assets, a risk-free asset (like bonds) and a risky asset (like stocks). The risk-free asset has price normalized to $1 and pays out $1 per share (a return of zero). The risky asset also has price of $1, but its payout is uncertain: there is a 90 percent chance that it pays out $1.05 per share, and there is a 10 percent chance that it pays out

$0.90 per share. Hence, the risky asset involves a larger expected return than the risk-free asset (3.5 versus 0 percent), but also carries the possibility of doing worse.

In deciding how to divide personal wealth, the individual faces a trade-off: investing more in the risky asset yields a higher expected final wealth (or, equivalently, a higher expected return), but it also creates more risk. While the intuition of risk aversion does not say where any given person should end up, it does suggest that the more risk-averse a person is, the less that person should invest in the risky asset.

Extending this example, now suppose the price of the risky asset is endogenous. In particular, suppose the risky asset has a fixed supply of shares and the payout per share is as above. However, the price adjusts such that the demand for those shares is equal to the supply of those shares. This demand might come from a homogeneous population in which everyone invests the same amount in the risky asset, or it might come from a heterogeneous population in which people invest different amounts in the risky asset. The intuition of risk aversion implies that individuals will hold the risky asset only if it generates a higher expected value than the risk-free asset, and therefore implies that risky assets should pay a higher expected return than risk-free assets. Moreover, the risk-aversion intuition implies that if the population becomes more risk-averse, the expected return on the risky asset must increase (or, equivalently, the price of the risky asset must decline).

As for the basic conclusions in the insurance domain, the basic conclusions above follow from the simple intuition of risk aversion, and they do not require the expected utility model. Again, expected utility can yield more quantitative statements. For example, with the constant relative risk aversion utility function, a person with $\rho = 20$ and prior wealth $50,000 who faces an exogenous price of $1 per share of the risky asset will invest $24,746 in the risky asset. If prices were endogenous and there were a homogeneous population with $\rho = 20$ and 20,000 shares of the risky asset per person, then the equilibrium price for the risky asset would be $1.009 per share, yielding an expected return of 2.6 percent. If this population became more risk-averse in the sense that everyone had $\rho = 40$, then the demand for the risky asset would decrease leading to a new equilibrium price of $0.968 per share, yielding an expected return of 7.0 percent.

However, the real value of applying expected utility to study financial investment is that it permits much more complex analyses. As a classic example, the consumption capital asset pricing model builds from expected utility—in particular, from diminishing marginal utility from aggregate consumption. The model implies that assets whose returns are positively correlated with aggregate consumption will be less valued than assets whose returns are negatively correlated with aggregate consumption, and that the strength of this effect depends on the degree of risk aversion (for an overview of the consumption capital asset pricing model, see Breeden, Litzenberger, and Jia 2015).

Principal–Agent Relationships

Consider a principal–agent paradigm in which a risk-neutral principal hires a risk-averse agent to complete a task. This framework can be used to describe a

number of interesting economic situations, such as the relationship between a real estate agent and a buyer, an elected official and a group of voters, a firm and an employee, or the managers of a company and its stockholders.

As a concrete example, consider a tenant–farmer (the agent) who might enter into a contract with a landlord (the principal). The tenant exerts effort that affects the farm's crop yield, although the crop yield also depends on other random forces, such as the weather and the appearance of pests. The landlord benefits from selling the crops, and then the landlord compensates the tenant. Because the landlord cannot observe the tenant's effort, and can only observe the realized crop yield, the landlord's compensation to the tenant can be conditional only on the observed crop yield. The tenant has the outside option of getting work as day laborer with a certain wage, and will only accept the contract offered by the landlord if it will make the tenant at least as well off.

To incentivize effort, the landlord must make the tenant's compensation depend on the realized crop yield—otherwise the tenant will exert no effort. However, because the crop yield depends on more than just effort, this form of contract imposes risk on the agent. The intuition of risk aversion implies that, when signing the contract, the agent will demand to be compensated for taking on this risk. As a result, the optimal contract from the principal's perspective yields smaller profits for the principal than would occur if the principal could observe and contract directly on effort. Intuitively, if the landlord could directly contract on effort, then he could demand the efficient level of effort while simultaneously imposing no risk on the tenant. In contrast, when the landlord can contract only based on crop yield, the landlord must provide extra compensation to the farmer for bearing risk.

Some form of this incentives–insurance trade-off is at the heart of many principal–agent analyses. Yet again, the basic conclusions above follow from the simple intuition of risk aversion, and, importantly, they do not require the expected utility model.

Decoupling Risk Aversion from Expected Utility

In each of the three domains above, we have emphasized how many of the basic conclusions often attributed to expected utility in fact follow in a straightforward way from the simple intuition of risk aversion. Expected utility has its value—beyond capturing the simple intuition of risk aversion—only when one moves towards more quantitative analyses or more complex, nuanced predictions.

As we shall soon see, however, other models can also capture the simple intuition of risk aversion, and thus would yield the same basic conclusions in the three domains above. However, when we move toward more quantitative analyses or more complex, nuanced predictions, these alternative models often yield different conclusions. Moreover, in many instances, these different conclusions seem more in line with the empirical facts than the conclusions derived from expected utility. Later in the paper, we revisit the three domains above in order to demonstrate these points. But before turning to alternative models, we highlight an important flaw in the expected utility model.

The Calibration Problem

Rabin (2000) and, in this journal, Rabin and Thaler (2001) point out a serious flaw with the expected utility model: the structure of the model rules out seemingly plausible preferences due to a calibration problem in dealing with stakes of different sizes—for example, the model can have problems simultaneously explaining choices over small stakes and choices over moderate stakes.

To illustrate the issue, consider the earlier example in which a person with a constant relative risk aversion utility function and prior wealth \$50,000 would accept the 50:50 gamble to lose \$10 or win \$12 only if ρ is smaller than 831. Suppose we observed such a person reject this bet, implying ρ is larger than 831. Any such ρ would imply the person would also reject a 50:50 gamble to lose \$100 or win a huge positive amount like \$1 trillion—in fact, the model implies that a person with this wealth and risk aversion would reject the 50:50 bet that risks a loss of \$100 no matter how large the potential positive gain.

This flaw does not rely on the functional form assumption for the utility function. Rabin (2000) proves a theorem that implies that, for any functional form, if a person rejects a 50:50 gamble to lose \$10 or win \$12 for a range of prior wealths, the expected utility model puts strong restrictions on how that person must behave for larger stakes. This calibration problem derives from the fact that, using the expected utility model, risk aversion is attributed to diminishing marginal utility over wealth. Hence, if one exhibits noticeable and significant risk aversion over small stakes, it implies significant local curvature of the utility function. If that noticeable risk aversion over small stakes applies over a range of prior wealth, it adds up to a huge amount of curvature over larger stakes.

Hence, the structure of the expected utility model puts serious constraints on individual preferences. For instance, expected utility—with no restrictions on the utility function—implies that a person could not have the following preferences over 50:50 gambles:

1) Reject a 50:50 gamble to lose \$10 or win \$12 for any prior wealth between \$49,000 and \$60,000.
2) Accept a 50:50 gamble to lose \$100 or win \$9,000 for prior wealth \$50,000.

Whether a person has such preferences ought to be an empirical question—and it surely seems plausible that some people might. Under expected utility, however, it is not an empirical question, because the model implies that this pattern of preferences simply is not permitted.

For some applications, the expected utility calibration problem might not be a major issue. Most notably, in applications that focus on a single choice for each individual, or that focus on multiple choices for each individual that all involve similar-sized stakes, the calibration problem may not be much of a problem. However, in applications where individuals make multiple choices over different-sized stakes, the calibration problem can be important. Moreover, the calibration

problem further implies that one must be extremely cautious when taking quantitative estimates of risk aversion derived from one environment and using those estimates to predict choices or conduct a quantitative welfare analysis in another environment.

We next turn our attention to some alternative models that can also capture the basic intuition of risk aversion and that also have advantages relative to expected utility. One advantage will be that these models do not suffer from the same calibration problem. But a more important advantage is that the alternatives sometimes better capture the empirical facts than expected utility, even in applications where the calibration problem is unlikely to be an issue.

Alternative Models of Risk Aversion

Alternative models of risk aversion, often motivated by the psychology and behavioral economics literature, are starting to provide new insights and empirical content to classic domains of risk. Much of our discussion will focus on two alternative models of risk aversion that were integral components in Kahenman and Tversky's (1979) prospect theory: loss aversion and probability weighting.[3] Loss aversion has been the most extensively applied alternative model of risk attitudes. The potential value of probability weighting has only recently been recognized by economists. We also briefly discuss a third alternative—context-dependence and salience—that is starting to gain some traction. As we'll see, these models maintain the basic structure of expected utility, but introduce alternative sources of risk aversion. While we discuss each model independently, the best approach to modeling risk preferences might incorporate multiple sources of risk aversion, perhaps including diminishing marginal utility of wealth.

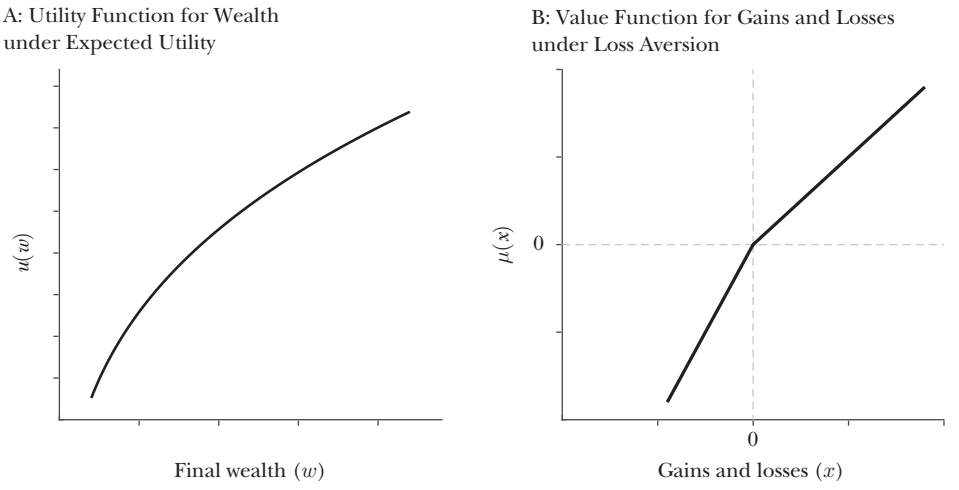
Loss Aversion

The model of loss aversion has much the same structure as expected utility, except that expected utility's utility function for wealth (depicted in Figure 1A) is replaced by a value function for gains and losses (depicted in Figure 1B).

The model of loss aversion involves two key features. First, instead of thinking in terms of final wealth, a person evaluates outcomes in terms of gains and losses relative to some reference point. To illustrate, consider a person with prior wealth $50,000 who has the option to accept or reject a 50:50 gamble to lose $10 or win $12. According to expected utility, the person compares having the utility of $50,000 with certainty versus a 50:50 chance having the utility of $49,990 or the utility of $50,012. Under loss aversion with a reference point equal to prior wealth, the person instead

[3] In this journal, Barberis (2013) provides a perspective on how Kahneman and Tversky's (1979) prospect theory had been used in economics over the three decades since it was written. Although we echo some themes of that article, we focus more explicitly on risk aversion.

Figure 1
Expected Utility versus Loss Aversion as a Basis for Risk Aversion



Source: Authors.
Note: The figure on the left depicts a utility function for wealth that might be used under expected utility. The figure on the right depicts a value function for gains and losses that might be used under loss aversion.

compares having the value of $0 with certainty versus a 50:50 chance of having the value of –$10 or the value of +$12.

The second key feature is that losses loom larger than gains: the negative value from a loss is larger (in magnitude) than the positive value from an equal-sized gain. Figure 1B depicts a simple, two-part-linear functional form for the value function that captures this feature. With this functional form, there is a single parameter, λ, that reflects the degree of loss aversion. Formally it indexes the relative slope in the loss domain versus the gain domain, so that $\lambda = 1$ implies no loss aversion and $\lambda > 1$ implies loss aversion, with larger λ implying more loss aversion. When a person evaluates options that involve both gains and losses, the "kink" in the value function between losses and gains will generate risk aversion.[4]

A major issue with loss aversion is the question of what determines the reference point around which gains and losses are defined. Kahneman and Tversky (1979) primarily assume that the reference point is prior wealth. However, applications of loss aversion have often posited different reference points. To illustrate the importance of what one assumes, consider a person with a two-part-linear value function who faces a choice between obtaining $10 with certainty versus a 50:50 chance to obtain $0 or $22.

[4] Kahneman and Tversky (1979) further assume that the value function exhibits diminishing sensitivity to the magnitude of gains and losses, which implies concavity in the domain of gains and convexity in the domain of losses. However, most applications of loss aversion in economics have adopted the two-part-linear functional form.

With a reference point of prior wealth, these two choices involve outcomes entirely in the domain of gains, and because the value function is linear in that domain, the person would be risk-neutral and thus choose the risky option. However, suppose instead that, when presented with this choice, the person starts focusing on the possibility of obtaining \$10 with certainty and uses this as the reference point. With this framing, the comparison becomes obtaining \$0 with certainty versus a 50:50 chance to lose \$10 or gain \$12. Now the kink becomes relevant, and if loss aversion is large enough (λ is large enough), the person would choose the certain gamble.

Of course, in the example above, both assumptions about the reference point are a bit arbitrary, making this an important degree of freedom in applications of the model. Motivated by this concern, Kőszegi and Rabin (2006, 2007, 2009) develop a model of loss aversion with endogenous reference points, wherein the reference point is determined by one's expectations about outcomes, and those expectations are determined by one's choice. Kőszegi and Rabin in fact posit several variants of how to think about endogenous expectations, depending on the nature of the situation.[5] Over the past decade, the literature has been heavily influenced by the Kőszegi–Rabin approach, and our discussion of loss aversion will primarily focus on the value of this approach.

Before we conclude this initial discussion of loss aversion, we highlight two further features. First, models of loss aversion need not suffer a calibration problem. As discussed above, under expected utility, if someone would reject a 50:50 gamble to lose \$10 or win \$12 for a range of prior wealths, it implies significant local curvature of the utility function (for example, that in Figure 1A) over a range of wealth. Under loss aversion, in contrast, the same preference might only imply significant curvature of the value function (for example, that in Figure 1B) over the domain from –\$10 to +\$12. For instance, this would be the case if the reference point were prior wealth, and it can also be the case under the Kőszegi–Rabin model of loss aversion with endogenous reference points.

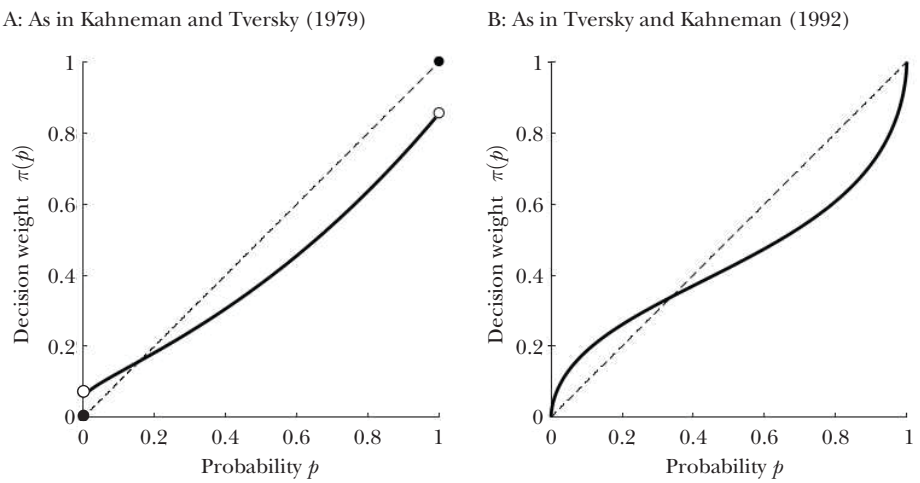
Second, note that the two-part-linear functional form in Figure 1B implies proportional risk preferences (under a reference point of prior wealth and under endogenous reference points). For example, if one rejects a 50:50 gamble to lose \$10 or win \$12 but accepts a 50:50 gamble to lose \$10 or win \$20, then one would reject a 50:50 gamble to lose \$100 or win \$120 but accept a 50:50 gamble to lose \$100 or win \$200. Models of loss aversion permit deviations from proportional risk preferences if the value function deviates from two-part linearity.

Probability Weighting

Under expected utility, the utility associated with each outcome is weighted by the probability of that outcome occurring. The basic idea of probability weighting

[5] See Kőszegi and Rabin (2007) for a description of two variants: "preferred personal equilibrium" and "choice-acclimating personal equilibrium." The latter shares some features with older models of "disappointment aversion" (Bell 1985; Loomes and Sugden 1986; Gul 1991). These distinctions will not be important for the discussion in this article.

Figure 2
Probability Weighting Functions as a Basis for Risk Aversion



Source: Authors.
Note: The figure on the left depicts the function that was suggested by Kahneman and Tversky (1979). The figure on the right depicts the one-parameter function that was proposed by Tversky and Kahneman (1992).

is that individuals might use decision weights that differ from the probabilities in systematic ways. To formalize this idea, one specifies a model for how probabilities are transformed into decision weights.[6]

Figure 2 illustrates two possibilities for a "probability weighting function" to use when making these transformations. Based on their original evidence, Kahneman and Tversky (1979) suggest a function similar to Figure 2A. Based on additional evidence, Tversky and Kahneman (1992) propose the function in Figure 2B, which also eliminates the discontinuity at the endpoints. Lattimore, Baker, and Witte (1992), Prelec (1998), and Gonzalez and Wu (1999) also suggest functional forms similar to that in Figure 2B. In each case, the horizontal axis shows the actual probability of an event, while the vertical axis shows the decision weight assigned to each probability. The 45-degree line thus corresponds to the decision weights under expected utility, and deviations from that line represent under- or overweighting of the objective probabilities.

Consider the implications of the probability weighting functions in Figure 2 for binary gambles. Specifically, consider a gamble with a probability p of getting \$10 and a probability $1 - p$ of getting a \$100. For gambles with small p (say, less than 0.2),

[6]Edwards (1954, p. 398) urged researchers to "think of a weighting function applied to the scale of objective probabilities which weights these objective probabilities according to their ability to control behavior."

the $10 receives a decision weight larger than its probability while the $100 receives a decision weight smaller than its probability. As a result, this gamble will look less attractive than it would to an expected utility maximizer—in other words, the probability weighting would generate a source of risk aversion. In contrast, for gambles with large p (say, greater than 0.8), exactly the opposite holds, and probability weighting would generate a source of risk seeking. Hence, unlike both diminishing marginal utility for wealth (as in expected utility) and loss aversion, probability weighting has somewhat more nuanced predictions, sometimes predicting risk aversion and sometimes predicting risk seeking.

While early formulations applied the probability weighting function to the probability of each outcome, this approach generates violations of dominance. Thus, it is now typical to use the rank-dependent approach proposed by Quiggin (1982) in which the probability weighting function is applied to the cumulative probability of each outcome. While this approach preserves the predictions above for binary gambles, it generates additional predictions for gambles with more than two outcomes. Specifically, rank-dependent probability weighting leads a person to overweight tail events and to underweight intermediate events.

Probability weighting has only started to be used more in economic applications, and issues remain to be worked out. For instance, the implications of probability weighting for 50:50 gambles is not entirely clear—and more to the point, different variants of probability weighting can predict risk aversion or risk seeking for such gambles. Perhaps more importantly, there is relatively limited discussion or consensus about the psychological principles that underlie probability weighting (for one discussion, see Burns, Chiu, and Wu 2010). However, as economists start to appreciate the potential value of probability weighting—as in the applications that we discuss below—we expect we will gain a deeper understanding.

Context-Dependence and Salience

The idea of context-dependence and salience is that the context or environment of a choice leads people to pay more attention to certain features of the choice situation. While this basic idea has a long tradition in psychology, it has proved elusive to formalize, especially in a way that can be used in economic applications. However, some promising models have been proposed in recent years, and this direction is worth additional exploration.

Context-dependence and salience can be thought of as providing a psychological mechanism for probability weighting. For example, in the salience model of Bordalo, Gennaioli, and Shleifer (2012), outcomes are weighted according to how much they differ from the average in a given state of the world, with more extreme payout states garnering more attention. When evaluating a lottery in which one has a 90 percent chance of winning $20 and a 10 percent chance of losing $100, there are two clearly defined states: the good state in which you win $20 if you take the bet versus nothing if you decline, and the bad state in which you lose $100 if you take the bet versus nothing if you decline. As the bad state involves more extreme outcomes, it attracts attention, resulting in the over-weighting of this low probability event.

The more general intuition is that in contexts where the downside of lotteries is salient, individuals will overweight those outcomes and exhibit risk aversion. Conversely, in contexts where the upside is salient, individuals may exhibit risk-seeking behavior. More importantly, context-dependence and salience can generate predictions that diverge from standard probability weighting, especially for gambles with more than two outcomes. In particular, because the context might include all options in a choice set, decision weights for one option might be influenced by other options in the choice set. As a result, unlike rank-dependent probability weighting where only extreme outcomes are overweighted, context-dependence and salience could lead to an overweighting of intermediate outcomes if attention is drawn there.

Of course, the big question—and a potentially big degree of freedom for researchers using this approach—is what determines salience. Several answers have been proposed. For example, Bordalo, Gennaioli, and Shleifer (2012) suggest that a feature will be more salient the more it differs from its average value in the choice set. Kőszegi and Szeidl (2013) suggest that a feature will be more salient if its range in the choice set is larger, while Bushong, Rabin, and Schwartzstein (2017) suggest that a feature will be more salient if its range in the choice set is smaller.[7] More work is needed in this area. Nevertheless, this approach has the potential to provide a more sophisticated and nuanced perspective, as we highlight in the next section.

Applications of Alternative Models

Having outlined three alternative models that can also generate risk aversion, we now revisit the three domains of insurance, financial investment, and principal–agent problems. For each domain, we describe the extent to which the alternative models do—and do not—capture the basic risk-aversion intuition for the simple behaviors discussed earlier. We then consider more nuanced behaviors to highlight how these alternative models can generate predictions different from expected utility that might accord better with empirical observations.

Reconsidering Insurance

Again, suppose an individual is exposed to a potential loss of $10,000 that has a 5 percent chance of occurring, with an option to purchase full insurance for a premium of $600. According to the simple intuition, a sufficiently risk-averse person might purchase this insurance.

[7] Bordalo, Gennaioli, and Shleifer (2012) explicitly explore the implications of their model of salience for risk preferences in the context of lotteries. Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartstein (2017) do not explicitly discuss how their models apply to preferences over lotteries, though the intuition of range-based salience could also be applied in this domain.

In fact, loss aversion with a reference point equal to prior wealth does not generate risk aversion in this domain. With a reference point of prior wealth, all outcomes would (weakly) involve losses, and thus loss aversion—which is about comparing losses to gains—would become irrelevant.[8] However, loss aversion with an endogenous reference point (as in Kőszegi and Rabin 2006, 2007, 2009) nicely captures the intuition of risk aversion in this domain. When the reference point is endogenous, it might be determined by the choice to insure or by the choice not to insure. In either case, the comparison will involve gains and losses. Specifically, the outcome of –$10,000 will always be viewed as a loss relative to the outcomes of –$600 and $0, and the outcome of $0 will always be viewed as a gain relative to the outcomes of –$600 and –$10,000. Hence, the kink in the value function between gains and losses will generate risk aversion, and the person's willingness to pay for insurance will be larger than the actuarially fair price. Moreover, the larger is loss aversion (the larger is λ), the larger will be the willingness to pay, and thus the person would buy the full insurance if loss aversion is large enough.

Probability weighting can also generate risk aversion in this domain, but it need not do so. For our specific example, probability weighting with the most common functional forms—as in Figure 2—nicely captures the risk-aversion intuition. In particular, the 5 percent chance of the loss is overweighted, while the 95 percent chance of no loss is underweighted, which together will generate risk aversion and thus a willingness to pay for insurance that is larger than the actuarially fair price. Moreover, the stronger is this overweighting, the larger will be the willingness to pay.

But matters become more complicated with probability weighting as we consider insurance against events with different probabilities. Most notably, if the probability of a loss is relatively high—say, 80 percent—the probability weighting functions in Figure 2 imply that the loss event would be underweighted. In that case, probability weighting would predict risk-seeking behavior and a willingness to pay that is below the actuarially fair premium. For instance, if a person faced an 80 percent chance of a $1,000 loss, the person would not be willing to pay $800 for full insurance. While we do not know of empirical evidence on the demand for insurance as a function of the probability of a loss, we note that many of the salient forms of real-world insurance tend to be for smaller-probability events.

In principle, models of context-dependence and salience can also generate risk aversion in this domain. If the context makes the loss event salient, the resulting added attention would generate risk aversion and thus a willingness to pay for insurance that is larger than the actuarially fair price. In addition, factors that increase the salience of the loss event would increase the willingness to pay. Conversely, if instead the context makes the no-loss event salient, it would produce risk-seeking behavior. Because there is not yet a clear sense of what determines salience, more work is required to generate a clear prediction even for this simple example.

[8] If as in Figure 1B the value function is linear in the domain of losses, the person would be risk-neutral in insurance decisions. If instead there is diminishing sensitivity and the value function is convex in the domain of losses, the person would actually be risk seeking in insurance decisions.

Beyond being able to generate risk aversion in our basic insurance example, the alternative models have additional implications for more nuanced behaviors. For instance, researchers have found an applied version of the expected utility calibration problem in the insurance domain, and the alternative models can help to address it. Using micro-data on property insurance decisions, Sydnor (2010) finds that, when interpreted with expected utility, the amount that many customers pay to reduce their deductible from $1,000 to $500 implies extremely large and implausible levels of diminishing marginal utility for wealth. He then demonstrates how plausible levels of both loss aversion and probability weighting can be consistent with the level of risk aversion in his data.

For some behaviors, the alternative models make predictions about risk aversion that differ from the predictions of expected utility. For instance, suppose we expand our example above such that the insurance policy includes a deductible. How does one's willingness to pay for the insurance depend on the size of the deductible? The structure of expected utility implies that one's marginal willingness to pay to reduce a potential loss is larger the larger is that loss. For example, one's willingness to pay to reduce one's deductible from $750 to $500 is larger than one's willingness to pay to reduce one's deductible from $500 to $250, which in turn is larger than one's willingness to pay to reduce one's deductible from $250 to $0. In contrast, both loss aversion with an endogenous reference point and probability weighting imply that one's marginal willingness to pay to reduce a potential loss is independent of the size of the loss. Hence, one's willingness to pay to reduce one's deductible from $750 to $500 is the same as one's willingness to pay to reduce one's deductible from $500 to $250 and the same as one's willingness to pay to reduce one's deductible from $250 to $0.[9]

In fact, these different predictions for how the willingness to pay reacts to the magnitude of risk can be used in empirical work to distinguish risk aversion stemming from diminishing marginal utility for wealth from risk aversion stemming from these alternative sources. For instance, Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2013) use data on household deductible choices for home and auto insurance to estimate a structural model of risk preferences that permits both diminishing marginal utility for wealth and "probability distortions," which they define to encompass both probability weighting and loss aversion. Taking advantage of the feature above, they estimate that the vast majority of risk aversion in their data is attributed to probability distortions.[10]

It is also instructive to return to our discussion of how much to insure given a linear price of 6 cents per dollar of insurance. Earlier, we highlighted how the simple intuition of risk aversion implies that a more risk-averse person should buy more insurance. The structure of expected utility additionally implies that, in fact,

[9]For simple binary (loss/no-loss) insurance, this pattern holds for preferred personal equilibrium, choice-acclimating personal equilibrium, and some variants of disappointment aversion.

[10]See Barseghyan, Molinari, O'Donoghue, and Teitelbaum (forthcoming) for a comprehensive discussion of estimating risk preferences in field settings.

the person should never fully insure no matter how large the individual's risk aversion is. This result follows from the fact that, in the expected utility model, starting from full insurance there is essentially no cost to taking on a small amount of risk. Hence, when insurance is actuarially unfair, it is always worthwhile to take on a little risk to get an increase in expected value. In contrast, under either loss aversion with an endogenous reference point or probability weighting, taking on a small amount of risk can be strictly costly, and thus it is possible to strictly prefer full insurance, even at an actuarially unfair rate.[11]

We have not discussed any specific predictions of context-dependent models here, because these models require further development. That said, we believe such models might eventually prove quite useful in the domain of insurance, especially for more complicated insurance products where the details of the insurance product could have a significant impact on what becomes salient and draws the attention of the decision maker.

Reconsidering Financial Investment

As before, suppose people must decide how to divide their wealth between two assets, a risk-free asset and a risky asset, where the risky asset involves a larger expected return than the risk-free asset but also carries the possibility of doing worse. According to the basic intuition of risk aversion, the more risk-averse a person is, the less that person should invest in the risky asset. Moreover, endogenous pricing of the risky asset implies that, with a risk-averse population, the price should adjust such that the risky asset has a larger expected return than the risk-free asset, and the more risk-averse is the population, the higher will be the equilibrium expected return on the risky asset.

Once again, loss aversion with an endogenous reference point nicely captures the intuition of risk aversion in this domain. Analogous to the insurance example, with the reference point determined by one's choice, the outcome when the risky asset pays a positive return will be viewed as a gain, while the outcome when the risky asset pays a negative return will be viewed as a loss. Hence, the kink in the value function between gains and losses will generate risk aversion, and the more loss-averse a person is (the larger is λ), the less likely the person becomes to invest in the risky asset. In addition, if the price of the risky asset is endogenous, loss aversion yields that the risky asset should have a larger expected return than the risk-free asset, and the more loss-averse is the population, the higher will be the expected return on the risky asset.

Under probability weighting, the implications for financial investment are more nuanced, depending on which states are underweighted and which are overweighted. For our specific example, the probability weighting functions in Figure 2 imply that the 90 percent chance that the risky asset yields a positive return should be underweighted, while the 10 percent chance that it yields a negative return

[11] This difference is related to first-order versus second-order risk aversion as discussed by Segal and Spivak (1990).

should be overweighted. The result is that the person would indeed display risk aversion, and thus probability weighting would nicely capture the intuition of risk aversion. More generally, though, probability weighting need not generate risk aversion—for example, if the risky asset had a small probability of a large positive return and a large probability of a small negative return.

Beyond being able to generate risk aversion in our basic financial-investment example, the alternative models have additional implications. This domain includes perhaps the most famous expected utility calibration problem: the equity-premium puzzle (Mehra and Prescott 1985), which is surveyed more recently in Mehra (2008) and in this journal by DeLong and Magin (2009). While diminishing marginal utility for wealth predicts that stocks should yield a greater return than bonds, the sheer magnitude of the differential that is observed empirically cannot be rationalized with reasonable levels of risk aversion. Benartzi and Thaler (1995) and Barberis, Huang, and Santos (2001) demonstrate how a model of loss aversion can be consistent with the observed historical equity premium. Note, however, that both of these papers make a rather different assumption about the coding of gains and losses: they assume that investors experience gain/loss utility each time they check their portfolio. This highlights the degree of freedom to make assumptions about how people define gains and losses.

One dimension on which the alternative models of risk aversion can make very different predictions from expected utility is participation in stock markets. Under expected utility, a well-known result is that anyone who is saving wealth for the future should invest at least some of that wealth in risky assets. As mentioned earlier, starting from no risk (investing everything in the risk-free asset), the cost of taking on a little risk (investing some in the risky asset) is second order, while the benefit from the higher expected return on risky assets is first order. In contrast, under loss aversion with an endogenous reference point or under probability weighting, the cost of taking on a little risk is first order, and as a result it can be optimal for a person to strictly prefer investing all wealth in the risk-free asset. Empirically, a large fraction of individuals do not participate in stock markets (for instance, Bertaut 1998).

Barberis and Huang (2008) develop in detail the implications of probability weighting for the pricing of skewed assets. For assets with positively skewed returns, the small chance of a large payoff is overweighted, inducing risk-seeking behavior. Investors are therefore willing to accept a negative excess return to invest in these assets. Barberis and Huang (2008) argue that this intuition offers a unified explanation for a number of puzzles in the finance literature: stocks that recently went through an initial public offering, options that are far from their strike price, distressed assets, and private equity all display positively skewed returns, and hence are prone to pricing "anomalies" if these phenomena are viewed from the perspective of expected utility. Recently, some of these predictions have received empirical support (for instance, Boyer, Mitton, and Vorkink 2010; Green and Hwang 2012).

Finally, we mention the potential for models of context-dependence and salience in this domain. Most of the work on such models does not yet seriously address financial market implications. However, Bordalo, Gennaioli, and Shleifer

(2013) emphasize how salience can provide a psychological grounding for a preference for skewness. They further describe how it can provide an intuitive account of the growth-value puzzle, which refers to the empirical finding that stocks with low market prices relative to fundamentals ("value stocks") command an above-average return, while stocks with high prices relative to fundamentals ("growth stocks") earn below average returns (Fama and French 1992). Bordalo et al. suggest that the unlikely upside of a growth company becoming a market leader attracts attention, while the small possibility of a value stock going bankrupt also attracts attention. They also suggest that salience offers a unique account for the empirical finding that risk premiums vary over time (Campbell and Shiller 1988). In particular, Bordalo et al. suggest that in booms, the upside of the market is salient, prompting risk-seeking behavior and overvaluation, whereas in busts, the downside weighs on investors' minds, causing them to become risk-averse and undervalue the market. These speculations are intriguing, but clearly more work needs to be done.

Reconsidering Principal–Agent Relationships

In the principal–agent paradigm in which a risk-neutral principal hires a risk-averse agent to complete a task—as with a landlord and a tenant-farmer—alternative models of risk aversion can easily generate the same general trade-off between incentives and insurance that arises from expected utility. Imposing risk on the tenant remains necessary to generate incentives. An agent who dislikes that risk, for whatever reason, will demand to be compensated.

Much as for the domains of insurance and financial investment, loss aversion with an endogenous reference point would clearly generate such risk aversion. Probability weighting could also generate it, depending on the specific probabilities. If the states with low crop yield—and therefore low compensation—are overweighted, the tenant will demand to be compensated for the risk. If instead those states are underweighted, then the landlord can in fact take advantage of the tenant's probability weighting to pay an expected compensation that is less than the agent's reservation wage.

Yet again, the potential value of alternative models lies in whether they make better predictions for more complex behaviors. Indeed, attempts to take the basic principal–agent model based on expected utility to data have not been especially successful: factors that one might expect to moderate the strength of incentives—such as the noisiness of the relationship between effort and output, the marginal return to effort, and the degree of risk aversion of agents—seem to have little predictive content. As a result, the agency literature has turned its focus in other directions that rely less on risk aversion as a limit on incentives: for example, to concerns about multi-tasking and the inefficient allocation of effort across types of tasks (for overviews, see Prendergast 1999, and in this journal, Gibbons 1998). While such analyses have been successful, some recent work suggests that there might be value to turning to behavioral approaches to contract theory, as reviewed in depth by Kőszegi (2014). Alternative models of risk aversion are but one of a range of perspectives that have been offered in this literature.

One example of the potential gains from such an approach revolves around what is sometimes known as the "paradox of simple contracts." A classic result in agency theory (based on expected utility) is that the optimal contract should make use of any observable that reveals information about the agent's otherwise unobserved choice of effort (Hölmstrom 1979). For instance, in our landlord-tenant example, if there are many possible realizations of crop yield, and if increased effort shifts the distribution of crop yields towards higher amounts, then the optimal contract should involve a compensation level that fully depends on crop yield. In other words, the theory predicts that contracts should be complex. In practice, however, we often see very simple contracts—for example, the dependence on crop yield might be simplified to a small number of possible compensations.

Herweg, Müller, and Weinschenk (2010) demonstrate that, in fact, loss aversion with an endogenous reference point can make simple contracts optimal. Specifically, they rely on a feature of the Kőszegi and Rabin (2006, 2007, 2009) approach wherein the reference point can in fact be a reference lottery, and the person compares the realized outcome to each outcome that was possible in the reference lottery. With this structure, the more risk there is in the reference lottery, the more disutility there is from loss aversion. Herweg et al. show that, even when many different outcomes are possible, the optimal contract can in fact be a binary contract in which the agent gets a high wage if output is large enough, and otherwise gets a low wage. Intuitively, there must be at least two wage levels to generate incentives for effort, but adding additional wage levels only creates the possibility of additional disutility from loss aversion.

Another example relates to firms' use of stock options. In this journal, Hall and Murphy (2003) note that the prevalence of stock options in compensation packages is puzzling from the perspective of expected utility because standard risk aversion implies that employees should value options below their market price, making them an expensive way of providing incentives. Using data on compensation packages for 598 chief executive officers, Dittmann and Maug (2007) show that to (partially) account for observed stock option holdings, agents need to exhibit very low risk aversion. However, such low risk aversion also predicts a negative base salary. Spalt (2013) argues that probability weighting offers an intuitive explanation for why firms issue so many stock options. Specifically, stock options typically involve a small chance of a large return, and if agents overweight this possibility, stock options become a cheap method for firms to incentivize their workers.

Finally, recent work on context-dependence and salience might also shed light on why strongly incentivized contracts are so unpopular. Bushong, Rabin, and Schwartzstein (2017) argue that agents might be less likely to exert effort when incentives create significant income uncertainty. Intuitively, by generating a wide range of potential incomes, such incentives have the perverse effect of making effort especially salient, causing agents to underweight these monetary incentives. While this result depends on the specific assumptions about salience, it highlights the potential value context-dependence risk preferences offer.

Discussion

This article has had three main goals: 1) to highlight that the basic intuition of risk aversion driving many results in economics is not intimately tied to expected utility, 2) to describe some alternative models that can also capture the basic intuition of risk aversion, and 3) to discuss how, for more complex behaviors, these alternative models might better explain some observed phenomena (than does expected utility). Much work remains to be done, and we conclude by discussing some broader issues related to this agenda.

Rabin (2013) emphasizes the importance of "portable models" that can easily be applied in a broad set of economic applications. It is also important that models be tractable so that they can be extended from simpler to more complex settings. Expected utility fares well on both dimensions: it is a simple and straightforward model with few degrees of freedom. Alternative models of risk aversion currently fare less well.

Early models of loss aversion were fairly simple and tractable, but they weren't entirely portable. In particular, in each application, one had to—or, more to the point, one was permitted to—make application-specific assumptions about the determinants of gains and losses. The Kőszegi and Rabin (2006, 2007, 2009) approach to loss aversion with an endogenous reference point attempted to reduce this degree of freedom by imposing that the reference point is fully determined by one's expectations about outcomes. But this approach comes at the expense of some tractability, and there is still some flexibility in what one assumes about the source of expectations. Moreover, there might be reference points that are unrelated to expectations. For instance, people might also define gains and losses relative to past outcomes, as in the dynamic model of job search in DellaVigna, Lindner, Reizer, and Schmieder (2017). Or people might define gains and losses relative to certain focal outcomes, as in the finding of Rees-Jones (forthcoming) that a zero balance due seems to be a focal reference point for tax filers. A more systematic understanding of when these various reference points are appropriate is necessary for richer, more portable, characterizations of loss aversion.

In principle, probability weighting appears to be a portable and somewhat tractable model. However, it runs into issues in applications where one must simplify the state space: for example, when one assumes for tractability a coarser set of outcomes than might really be relevant. While such simplifications are relatively innocuous in the expected utility model, decisions on how to simplify the state space can have a big impact with nonlinear probability weighting. Furthermore, the psychology of probability weighting is poorly understood.

Models of context-dependence and salience may provide a better foundation for when, and why, probabilities are re-weighted. In their current form, however, there is no agreed-upon definition of salience or what aspects of the choice environment grab attention. As an additional layer of complexity, expectations, previous choice sets, or options available elsewhere may be an integral part of the "context" for the current choice. While we are optimistic that such models might yield new

insights into the nuances of risk preferences, much empirical evidence and theoretical endeavor is required.

Alternative models of risk aversion require us to rethink how we conduct welfare analysis and make policy recommendations. In simple terms, we have already highlighted how alternative models can yield different behavioral predictions, which is of course relevant for welfare analysis. Indeed, these models suggest new policy levers that one might have thought unimportant when viewed through the lens of expected utility—for instance, policies that require bundling (or unbundling) of risks in ways that change the perception of gains and losses, the relevant probabilities, or the broader context could have significant impacts on behavior. Beyond the behavioral predictions, a perhaps even more important issue arises: the psychology of the alternative models suggests that the model that describes people's behavior might not be the metric we ought to use for welfare analysis. Such a distinction between "decision utility" and "experienced utility" was first discussed by Kahneman (1994), and there continues to be a debate about how best to approach this distinction (for some thoughtful discussions, see Kahneman and Thaler 2006; Kőszegi and Rabin 2008; Bernheim and Rangel 2009).

An important direction for future research is to apply alternative models of risk aversion in dynamic models of risky choice. Some prominent situations of risk aversion have an intertemporal dimension, including savings and consumption problems, dynamic labor supply decisions, and health decisions. So far, there has been limited progress in taking alternative models to dynamic settings. A notable exception is Kőszegi and Rabin (2009), who define loss aversion over changes in beliefs regarding both current and future consumption. A number of novel insights emerge from these risk attitudes, and Pagel (2017, forthcoming) shows that they combine to offer a unified explanation for a number of seemingly disparate puzzles. But these papers are just the first steps in this area.

Real-world risk aversion is clearly not as straightforward as expected utility suggests. Perhaps most notably, we don't even always observe risk aversion, as in some situations individuals systematically exhibit risk-seeking behavior. At horse races and in casinos, people actively make bets in a domain where virtually all bets have a negative expected payoff. In experiments, people are often risk seeking when considering simple binary gambles with a moderate probability of a loss, or with only a small probability of a gain. If we want a portable model that can explain behavior across domains, simple expected utility will not work. Additional sources of risk aversion (or risk seeking) need to be used instead of, or in conjunction with, diminishing marginal utility of wealth. The alternative models discussed here might not ultimately prove to be the best models for studying risk aversion, but they are useful steps in what we hope will be an ongoing search.

■ *We thank Ori Heffetz, Gordon Hanson, Enrico Moretti, and Timothy Taylor for excellent comments and suggestions.*

References

Arrow, Kenneth J. 1965. *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnssonin Säätiö. (Reprinted in 1971 in *Essays in the Theory of Risk-Bearing*, Markham Publishing, Chicago, pp. 90–109.)

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27(1): 173–196.

Barberis, Nicholas, and Ming Huang. 2008. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." *American Economic Review* 98(5): 2066–2100.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics* 116(1): 1–53.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. "The Nature of Risk Preferences: Evidence from Insurance Choices." *American Economic Review* 103(6): 2499–2529.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. Forthcoming. "Estimating Risk Preferences in the Field." *Journal of Economic Literature*.

Bell, David E. 1985. "Disappointment in Decision Making under Uncertainty." *Operations Research* 33(1): 1–27.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics* 110(1): 73–92.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124(1): 51–104.

Bertaut, Carol C. 1998. "Stockholding Behavior of U.S. Households: Evidence from the 1983–1989 Survey of Consumer Finances." *Review of Economics and Statistics* 80(2): 263–75.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics* 127(3): 1243–85.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Asset Prices." *American Economic Review* 103(3): 623–28.

Boyer, Brian, Todd Mitton, and Keith Vorkink. 2010. "Expected Idiosyncratic Skewness." *Review of Financial Studies* 23(1): 169–202.

Breeden, Douglas T., Robert H. Litzenberger, and Tingyan Jia. 2015. "Consumption-Based Asset Pricing, Part 1: Classic Theory and Tests, Measurement Issues, and Limited Participation." *Annual Review of Financial Economics* 7(1): 35–83.

Burns, Zach, Andrew Chiu, and George Wu. 2010. "Overweighting of Small Probabilities." *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons.

Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein. 2017. "A Model of Relative Thinking." Working Paper, Harvard Business School. http://www.hbs.edu/faculty/Publication%20Files/RelativeThinking_6a962ed2-9254-43d6-984d-6781ffe8a78e.pdf.

Campbell, John Y., and Robert J Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1(3): 195–228.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review* 97(3): 745–88.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder. 2017. "Reference-Dependent Job Search: Evidence from Hungary." *Quarterly Journal of Economics* 132(4): 1969–2018.

DeLong, J. Bradford, and Konstantin Magin. 2009. "The U.S. Equity Return Premium: Past, Present, and Future." *Journal of Economic Perspectives* 23(1): 193–208.

Dittmann, Ingolf, and Ernst Maug. 2007. "Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 62(1): 303–343.

Edwards, Ward. 1954. "The Theory of Decision Making." *Psychological Bulletin* 51(4): 380–417.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47(2): 427–65.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12(4): 115–132.

Gonzalez, Richard, and George Wu. 1999. "On the Shape of the Probability Weighting Function." *Cognitive Psychology* 38(1): 129–66.

Green, T. Clifton, and Byoung-Hyoun Hwang. 2012. "Initial Public Offerings as Lotteries: Skewness Preference and First-Day Returns." *Management Science* 58(2): 432–44.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59(3): 667–86.

Hall, Brian J., and Kevin J. Murphy. 2003. "The Trouble with Stock Options." *Journal of Economic Perspectives* 17(3): 49–70.

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk. 2010. "Binary Payment Schemes: Moral Hazard and Loss Aversion." *American*

Economic Review 100(5): 2451–77.

Hölmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10(1): 74–91.

Kahneman, Daniel. 1994. "New Challenges to the Rationality Assumption." *Journal of Institutional and Theoretical Economics* 150(1): 18–36.

Kahneman, Daniel, and Richard H. Thaler. 2006. "Anomalies: Utility Maximization and Experienced Utility." *Journal of Economic Perspectives* 20(1): 221–34.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–92.

Kőszegi, Botond. 2014. "Behavioral Contract Theory." *Journal of Economic Literature* 52(4): 1075–1118.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics* 121(4): 1133–65.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review* 97(4): 1047–73.

Kőszegi, Botond, and Matthew Rabin. 2008. "Choices, Situations, and Happiness." *Journal of Public Economics* 92(8–9): 1821–32.

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99(3): 909–36.

Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics* 128(1): 53–104.

Lattimore, Pamela- K., Joanna R. Baker, and Ann D. Witte. 1992. "The Influence of Probability on Risky Choice: A Parametric Examination." *Journal of Economic Behavior & Organization* 17(3): 377–400.

Loomes, Graham, and Robert Sugden. 1986. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Review of Economic Studies* 53(2): 271–82.

Mehra, Rajnish. 2008. *Handbook of the Equity Risk Premium.* Elsevier.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15(2): 145–61.

Pagel, Michaela. 2017. "Expectations-Based Reference-Dependent Life-Cycle Consumption." *Review of Economic Studies* 84(2): 885–934.

Pagel, Michaela. Forthcoming. "A News-Utility Theory for Inattention and Delegation in Portfolio Choice." *Econometrica***.**

Pratt, John W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32(1/2): 122–36.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica* 66(3): 497–527.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37(1): 7–63.

Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior & Organization* 3(4): 323–43.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68(5): 1281–92.

Rabin, Matthew. 2013. "An Approach to Incorporating Psychology into Economics." *American Economic Review* 103(3): 617–22.

Rabin, Matthew, and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives* 15(1): 219–32.

Rees-Jones, Alex. Forthcoming. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies***.**

Segal, Uzi, and Avia Spivak. 1990. "First Order versus Second Order Risk Aversion." *Journal of Economic Theory* 51(1): 111–25.

Spalt, Oliver G. 2013. "Probability Weighting and Employee Stock Options." *Journal of Financial and Quantitative Analysis* 48(4): 1085–1118.

Sydnor, Justin. 2010. "(Over)insuring Modest Risks." *American Economic Journal: Applied Economics* 2(4): 177–99.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5(4): 297–323.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 115–134

On the Relationship between Cognitive Ability and Risk Preference

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde

Many decisions of individuals involve a combination of internal preferences and mental processes related to cognitive ability. As Frederick (2005) argued in this journal, "there is no good reason for ignoring the possibility that general intelligence or various more specific cognitive abilities are important *causal* determinants of decision making." Since then, a number of empirical studies have focused on the relationship between cognitive ability and decision-making in different contexts. This paper will focus on the relationship between cognitive ability and decision-making under risk and uncertainty. Taken as a whole, this research indicates that cognitive ability is associated with risk-taking behavior in various contexts and life domains, including incentivized choices between lotteries in controlled environments, behavior in nonexperimental settings, and self-reported tendency to take risks.

We begin by clarifying some important distinctions between concepts and measurement of risk preference and cognitive ability. In particular, complexity and possible confusions arise because observed measures of risk preference and cognitive ability are used to represent the latent characteristics of these concepts. We discuss the substantial (and somewhat implausible) range of assumptions that need

■ *Thomas Dohmen is Professor of Applied Microeconomics and Armin Falk is Director of the Behavior and Inequality Research Institute, both at the University of Bonn, Bonn, Germany. David Huffman is Professor of Economics at the University of Pittsburgh, Pittsburgh, Pennsylvania. Uwe Sunde is Professor of Economics at the University of Munich, Munich, Germany. Their email addresses are t.dohmen@uni-bonn.de, Armin.Falk@briq-institute.org, huffmand@pitt.edu, and uwe.sunde@econ.lmu.de.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.115 doi=10.1257/jep.32.2.115

to be satisfied in order to be able to interpret a correlation between measures of risk preference and cognitive ability as a relationship between latent risk preference and latent cognitive ability. Drawing causal inferences from such relationships raises additional challenges.

We go on to argue that it is nevertheless important and valuable to study whether cognitive ability is related to measured risk preference (see also Dohmen, Falk, Huffman, and Sunde 2010). Risk preference is typically measured by risky behavior (actual or self-reported). If risky behavior varies systematically with cognitive ability, this may reinforce or counteract the impact of cognitive ability on life outcomes, depending on the nature of the correlation. If there is a relationship, it also becomes important to control for cognitive ability when relating life outcomes to standard revealed preference measures of risk preference. If cognitive ability has a causal impact on measured risk preference, it is important to understand the mechanism, and some intriguing policy implications arise.

We then take stock of what is known empirically on the connections between cognitive ability and measured risk preferences, looking at studies using real-world risky behavior, experimental measures of risky choice, and self-reported measures of willingness to take risks. One pattern that emerges frequently in these studies is that cognitive ability tends to be positively correlated with avoidance of harmful risky situations, but it tends to be negatively correlated with risk aversion in advantageous situations. This suggests that the relationship between cognitive ability and risk taking has a reinforcing effect on economic outcomes. There is also intriguing emerging evidence that measured risk preference is particularly strongly related to certain facets of cognitive ability, those that facilitate quantitative problem solving, with implications for understanding mechanisms and possibly for better targeting policy interventions.

We conclude by discussing perspectives for future research, in particular the scope for the development of richer sets of elicitation instruments and measurement across a wider range of concepts. We also consider progress in neuroscience, but conclude that at present that field still seems relatively far from allowing definitive conclusions about latent risk preference and cognitive ability. Nevertheless, the existing empirical evidence suggests that interventions to influence cognitive ability, should they be possible, might have spillovers on risky choice.

A Conceptualization of Risk Preference and Cognitive Ability: Measurement and Causality

In economic theory, risk preference has traditionally been conceptualized as a primitive of the decision model that affects the way in which individuals make risky trade-offs. This trait is general in the sense that it is relevant for risky choices in all contexts and domains, whether the choices are about financial assets, car driving, or health. Importantly, this concept of risk preference as a primitive is not directly observable, but rather is a latent trait. The typical way of obtaining empirical measures of this trait is to observe the decisions of individuals who face

a particular, well-defined trade-off between choice options that are associated with different riskiness. Under the assumption that all other factors that affect the trade-off are controlled for and that the subject perceives the trade-off exactly as intended by the researcher, the subject's choice reveals latent risk preference. For example, risk preference is often measured in experiments based on a choice between two financial lotteries that differ with respect to risk. The key assumptions behind this set-up are that the individual understands probabilities and the expected values of the options being offered, and that other factors that may affect risky choice besides the latent preference (for example, wealth), can be controlled for adequately. Another crucial assumption is that the inferred preference ranking is invariant to the context or framing of the choice, so that it can be generalized to risky choice in other contexts. Similar assumptions are needed to infer latent risk preference from observing choices over risky life outcomes, or from self-reports of risk attitudes.

If latent risk preference were systematically related to what is often assumed to be another primitive of economic models—cognitive ability—this would have important implications for economics. While both are typically assumed (implicitly or explicitly) to be orthogonal, it is conceivable that these traits are in fact systematically related. If so, this would have important consequences for the appropriate specification of models, theoretical and empirical, that include both traits. Various mechanisms have been proposed through which a relationship might exist. In a "two-system model," if latent risk preference is partly driven by the emotional, impulsive system-one, but higher cognitive ability entails greater control of decisions by the deliberative, calculating system-two, then there might be a causal impact of cognitive ability on latent risk preference (Benjamin, Brown, and Shapiro 2013). In this case, latent risk preference would be, at least partly, endogenous to cognitive ability.

Alternatively, causality could potentially go the other direction, with latent risk preference playing a role in development of partly endogenous cognitive ability. For instance, Cunha, Heckman, and Schennach (2010) formulate and estimate a model of the accumulation of cognitive and noncognitive skills where the evolution of skills depends on family environment, genetic factors, parental skills, investments, and shocks. In their dynamic model, changes in the vector of skills depend on the stock of skills at the beginning of a given period. Cognitive skills might thus be affected by noncognitive skills, and potentially preferences. They also emphasize that causality in early childhood development may be bi-directional, due to feedback processes. Heckman, Pinto, and Savelyev (2013) show that skill formation is affected by the Perry Preschool program, in which children aged three and four were exposed to a learning environment that promotes social competency, planning, and organization. Likewise, Kosse, Deckers, Schildberg-Hörisch, and Falk (2016) show that participation in a mentor program affects the formation of prosociality. It is conceivable that individuals partly choose their environment by sorting into particular types of situations, and this sorting decision could be related to risk preference.

Latent risk preference could also play a role later in life in choices about investment in education, with the latter fostering improved cognitive ability. Another source for a relationship between cognitive ability and latent risk preference could be coevolution: that is, certain combinations of these two traits might be beneficial. For example, perhaps evolutionary pressures might have created a tendency for low cognitive ability to be paired with cautious tendencies (Dohmen et al. 2010). Likewise, intergenerational transmission of traits (for example, Dohmen, Falk, Huffman, and Sunde 2012) might imply a correlation of these traits in the population (Boyer 2006).

One challenge to investigating a potential relationship between latent risk preference and cognitive ability arises precisely because risk preference is latent and the assumptions needed to infer this latent risk preference from risky behavior are fairly restrictive. If the choice situation is not held constant across individuals, and this introduces noise in measured risk preference, it may obscure any relationship between latent risk preference and cognitive ability. In addition, cognitive ability might systematically affect some of the factors that influence risky choice but are distinct from (latent) risk preference. After all, choices made on any particular task in settings of risk and uncertainty will be, at least partly, the outcome of a conscious process of mental deliberation, which in turn involves cognitive abilities such as processing of information related to probabilities and stakes, the calculation of expected values, and the valuation of different choice alternatives. Cognitive ability might also matter for the importance of other factors that may influence observed measures of risk preference, including aspects of the choice setting itself (for example, the decision frame, the domain of possible choices, stake size associated with decisions) and personal characteristics (like the level of pre-existing wealth or level of education).

Other challenges arise if the goal is to test *causal* relationships between latent risk preference and cognitive ability. A first challenge is developing an intervention that exogenously varies either latent risk preference or cognitive ability. It is not trivial to do this, in particular as the formation and operation of these primitives is not fully understood. A second challenge is ensuring that such an intervention changes exclusively one of the traits without also affecting the other (see Heckman 2005). For example, suppose that repeated exposure to risk in childhood affects latent risk preference. For this to be a valid source of exogenous variation, it is necessary that this repeated exposure does not also affect cognitive ability. It is unclear whether this will be the case. Likewise, interventions to affect cognitive ability might also influence latent risk preference. Without knowing more about the formation of latent risk preference and cognitive ability it is difficult to be sure that causality is being identified. Finally, identifying a causal effect requires the ability to measure changes in latent risk preference and cognitive ability, which is itself a challenge.

Yet another layer of complexity is added because cognitive ability is itself a latent trait, as well as multidimensional. Cognitive ability is typically measured by different types of achievement or performance tests. These tests only capture cognitive ability if other factors that might affect test performance are held constant. For

example, distractions on the day of the test, and personality traits that determine task motivation could play a role in test performance. It may be possible to control for some of these factors—for example, by measuring personality type, as we do in Dohmen, Falk, Huffman, and Sunde (2010). While some of these factors affecting test performance may be orthogonal to latent risk preference and just add noise, certain other factors such as risky or cautious test-taking strategy could conceivably be affected by latent risk preference. This raises the possibility that measured risk preference is related to performance on cognitive ability tests (with causality going from risk preference to test performance) but without indicating a relationship between latent risk preference and cognitive ability.

Moreover, cognitive ability is multidimensional. One might think of different facets of cognitive ability as lying along a continuum, ranging from exogenous primitives of the model to more endogenous and akin to acquired knowledge. This representation corresponds to the distinction between fluid and crystalized intelligence in psychology (Cattell 1971; 1987). Broadly speaking, crystallized intelligence represents knowledge and acquired skills (for example, verbal skills, numeracy, or financial literacy), whereas fluid intelligence captures the ability to solve novel problems and represents the outcome of biological factors in intellectual development, which includes capacities such as speed of processing, or memory. There are also other categorizations of cognitive abilities, which is one reason why there are many different types of tests for measuring cognitive ability (for example, Ackermann and Heggestad 1997). One test, the Wechsler Adult Intelligence Scale (WAIS), refers to a person's global capacity to act purposefully, to think rationally, and to deal effectively with his environment; different submodules capture different aspects of crystallized and fluid intelligence. Another test, Raven's Progressive Matrices, tries to measure abstract reasoning more closely related to fluid intelligence. Risk preference might be related to some facets of cognitive ability but not others, and the presence and direction of causality could depend on the facet.

For these reasons, finding a defensible way to extract a measure of latent risk preference that is distinct from cognitive ability, while understanding any potential causal relationship between the two, is very difficult. Add in the possibility that certain risky choices are structured in a way that is too fast for conscious calculation so that subconscious processes play a role, and an exquisitely difficult problem becomes even harder.

Evidence Concerning the Correlation between Cognitive Ability and Measured Risk Preference

The empirical literature contains a considerable array of evidence on how observed measures of risky choice are related to various facets of measured cognitive ability. Broadly speaking, three measurement approaches prevail: 1) observing risk-taking behavior in everyday life; 2) observing risk-taking from a menu of standardized choices (typically lottery choices) in a controlled

environment, as in an incentivized laboratory experiment; and 3) risk preference as captured by self-reported willingness to take risk. Studies of risk preference also differ with respect to the measures of cognitive ability and the facets of cognition that are considered, as well as with respect to risk-taking behavior across different contexts by respondents with different demographic and socioeconomic characteristics.

Cognition and Risk-Taking Behavior in Nonexperimental Settings

While risk-taking behavior in different contexts of life has been found to be correlated with various facets of cognition, the sign and magnitude of the correlation seems to vary across contexts and studies. With a closer look at this variation, however, a pattern emerges. Cognitive ability tends to be positively correlated with avoidance of harmful risky situations and to be negatively correlated with risk aversion in advantageous situations. Evidence on this theme emerges both from studies of behavior in risky situations, often done by psychologists and psychiatrists, and also from studies focused on economic decision-making.

Boyer (2006) reviews four strands of research in the developmental psychology literature on risk-taking behaviors in situations that involve undesirable real-world risks, such as substance use, alcohol consumption, unsafe sexual behavior, or criminal behavior. Each of the four strands considers a different class of factors that potentially influence the development of risk-taking behavior. Cognitive developmental research focuses on changes in decision-making capacities during childhood and adolescence that potentially underlie risk perception, probability judgment, and sensitivity to risk. Emotional development research examines the development of affective decision-making. Psychobiological research analyzes the neurological and biochemical bases of risk taking. Social development research investigates influences of the social environment on the emergence of risk-taking tendencies. The literature suggests that cognitive, emotional, and psychobiological development, as well as social factors, are correlated with changes in risk-taking behavior. Empirical evidence by and large indicates that engagement in precarious risky behaviors is lower for individuals with higher cognitive capacity and emotional regulation skills. It is not clear, however, whether there is a direct link between cognitive skills and risk preference. For example, the correlation could potentially reflect an effect of cognitive ability on quality of decision-making and therefore the ability to avoid harmful decisions.

Poor decision-making quality might also underlie the finding in epidemiological studies that low-cognitive-ability types are more likely to smoke, drink alcohol, or to commit a crime. Kubička, Matějček, Dytrych, and Roth (2001), for example, document that IQ (as measured during childhood by Wechsler Intelligence Scale, Children) is correlated with a lower prevalence of smoking in adulthood. Using data for 330 individuals from a 24-year-follow-up study, they estimate that the odds ratio of being a smoker at age 28–31 (versus being a nonsmoker) is twice as high for a child with an IQ of 85 compared to a child with IQ of 115. Frisell, Pawitan, and Långström (2012) share this methodological approach and relate a measure of

cognitive ability to risk-taking behavior later in life, using administrative data and a much larger sample of more than 250,000 Swedish men born between 1961 and 1975. They link data on a measure of general cognitive ability, which is contained in the Conscript Register (1980–1993) and based on the Swedish Enlistment Battery (SEB80), to violent criminal convictions from the Crime Register, 1973–2009. The raw data indicate a negative relationship between the cognitive ability score and the proportion of individuals convicted for violent crime for all three sub-samples considered: half-brothers who grew up apart, half-brothers raised together, and full brothers. Even in the group that has the smallest crime rate—brothers who were raised together—the proportion that had been convicted of violent crime is roughly 15 percentage points higher for those with the lowest SEB80 score of 1 compared to those with the highest score of 9. Regression results confirm a significant association between the cognitive ability score and crime rates for all subgroups that is robust to familial characteristics that might confound the relationship between cognitive ability and risky criminal behavior.

A large strand of literature on financial literacy and financial decision-making shows that a set of cognitive skills—such as numeracy, the ability to process economic information, and knowledge—improve the quality of decision-making under risk in the sense that they lead to better financial planning and wealth accumulation (reviewed by Hasting, Madrian, and Skimmyhorn 2013; Lusardi and Mitchell 2014). Agarwal and Mazumder (2013) relate a measure of cognitive skills that is based on the Armed Forces Qualifying Test (AFQT) score to the optimal use of credit cards for convenience transactions and to financial mistakes on a home equity loan application. They find that a one standard deviation increase in the AFQT score is associated with an 11 percentage point decrease in the likelihood of making a rate-changing mistake in the home loan application. Moreover, consumers that score one standard deviation higher on the AFQT are 24 percentage points more likely to discover the optimal balance transfer strategy in credit card use. These results are driven by arithmetic reasoning and math knowledge, two subparts of the AFQT test that assess math ability, but not by verbal ability (which is measured by paragraph comprehension and word knowledge in the AFQT).

In the context of desirable risk, better cognitive functioning seems to be associated with more risky behavior. Grinblatt, Keloharju, and Linnainmaa (2011) focus on the effect of cognitive ability on financial decision-making in a context that involves risk. They study stock market participation in the year 2000 of 158,044 Finnish males between 1953 and 1982. Merging data from several administrative data sources, they relate men's test scores from the Finnish Armed Forces (FAF) Intelligence Assessment, which contains 120 questions to elicit mathematical, verbal, and logical skills, to Finnish tax administration data and data on their daily portfolios and trades. Controlling for a large set of confounding factors, including wealth, income, marital status, children, age, asset ownership, and labor market status, they show that stock market participation monotonically increases with intelligence scores. These findings substantiate the results of a set of studies that higher cognitive ability is correlated with more risk taking in financial markets (for

example, Christelis, Jappelli, and Padula 2010; Cole, Paulson and Shastry 2014).[1] One potential explanation for these findings may be that cognitive ability improves the quality of decision-making in the sense of being able to recognize when risk taking is beneficial financially, although higher cognitive ability might also improve decision-making quality in a way that reduces the riskiness of decisions. Grinblatt, Keloharju, and Linnainmaa (2011), for example, show that high-IQ investors tend to hold less-risky portfolios and earn higher risk-adjusted returns. However, these studies on cognitive ability, financial literacy, and risky behavior in financial matters cannot tell us whether the relationship between cognitive skills and risk-taking behavior in a nonfinancial context also works through the quality of decision-making.

Focusing on risk-taking behavior in financial decisions and financial literacy, Lusardi and Mitchell (2014) discuss the difficulties in establishing whether such links are causal—as opposed to being related due to, say, endogeneity and measurement error. They review a set of studies that use instrumental variables estimation methods to address endogeneity of financial literacy, and find support for a causal impact of financial literacy on financial decision-making. An excellent example of this approach is the study by van Rooij, Lusardi, and Alessie (2011), who use the financial experiences of siblings and parents as an instrumental variable to show that financial literacy has a significant positive effect on stock market participation.

Cognition and Experimental Measures of Risk Aversion

Compared to nonexperimental settings, choice experiment settings have the advantage of greater control over the primary attributes of the risky choice options. A number of experimental studies have been conducted that use different measures of risk aversion and cognitive ability.

Experimental studies often use incentivized choices from menus of different monetary lotteries. Such menus are typically administered in the form of multiple price lists that confront subjects with a sequence of choices between lotteries. One prominent price list format proposed by Holt and Laury (2002) involves choices between two lotteries that differ in the spread between the high and the low payoff. From row to row, the potential outcomes of the lotteries are kept constant, but the probabilities of these outcomes are varied. A second type of price list format involves choices between a lottery and a safe payment. Typically the outcomes and probabilities of the lottery are not varied across rows, but the size of the safe payment is changed. A potential advantage of the latter approach is its simplicity as

[1] Christelis, Jappelli, and Padula (2010) analyze data from almost 20,000 seniors who responded to the Survey of Health, Aging and Retirement in Europe (SHARE) and find that answers to three cognitive tests—the number of animals one can name in one minute, the number of nouns (out of 10) one recalls, and a series of up to four numeracy questions—are significantly positively related with self-reported stock market participation. Cole, Paulson, and Shastry (2014) document that measures of cognition in the National Longitudinal Survey of Youth (NLSY) are related to self-reported ownership of stocks, bonds, or mutual funds.

perceived by the decision maker, as it allows for a straightforward identification of an individual's certainty equivalent of the respective lottery.

Burks, Carpenter, Goette, and Rustichini (2009) use the latter approach to elicit certainty equivalents of different lotteries among a subject pool of trucker trainees and to study the relationship of risky choice to cognitive ability. To measure cognitive ability, they use three different tests: a nonverbal IQ test (Raven's matrices), a test of the ability to plan (the Hit 15 game), and a quantitative literacy (or numeracy) test. From the lottery choices in the gain domain, they compute the coefficient of relative risk aversion. Relating it to the cognitive ability measures, they find a strong negative association between the measure of cognitive performance and risk aversion. This result stems from the fact that truckers with lower performance on the cognitive tests switch earlier in the choice table from preferring the lottery to the safe payment, which increases with each choice while the lottery is constant. They speculate that the perceived utility of the lottery is noisy, while the utility of the sure payment is perceived precisely. If the noise in perception is higher when cognitive ability is lower, and if individuals dislike what they do not perceive precisely, truckers with low cognitive ability should prefer the safe payment more frequently. Corroborating evidence for this conjecture comes from two findings: First, truckers with lower performance on the cognitive tests more frequently choose lotteries that have an expected value smaller than the safe payment alternative. Second, truckers with lower performance on the cognitive tests are more likely to make inconsistent choices, by switching multiple times in the price list. Truckers in the lowest quartile of the cognitive score distribution have a 25 percent higher likelihood of being inconsistent compared to those in the highest quartile. Other studies also indicate that higher intelligence is associated with a stronger tendency for consistent choices (for example, Frederick 2005; Oechssler, Roider, and Schmitz 2009; Benjamin, Brown, and Shapiro 2013).

Analyzing a representative sample of adults in Germany, in Dohmen, Falk, Huffman, and Sunde (2010), we use a similar incentivized choice between a lottery and a safe payment to elicit risk preferences. To measure cognitive ability, we apply two different tests: a symbol–digit correspondence test and a word fluency test (which are similar to submodules in the Wechsler Adult Intelligence Scale). While both tests are related to the speed of processing, they capture different dimensions of cognitive ability in terms of fluid and crystallized intelligence. The results point to a significantly negative correlation between risk aversion and cognitive ability even when accounting for family background, education, test-taking strategy, and personality traits. The findings also show that both distinct types of cognitive ability tests have power for explaining risky choice.

Benjamin, Brown, and Shapiro (2013) consider the correlation between risk aversion and cognitive ability among a sample of Chilean high school students, and they go on to explore possible causal channels behind this correlation. In contrast to the previous papers, they use multiple price lists offering choices between a safe payment and a sequence of lotteries with a 50/50 chance of winning. Instead of varying the safe payment, their design varies the high outcome in the lottery

across choice alternatives. In addition, they consider different lotteries with a low outcome that is zero or negative. Benjamin et al. have access to achievement test scores, taken from the quantitative and verbal sections of a national exam for university admission, as a measure of cognitive ability. Similar to the earlier studies, they find a negative correlation between risk-averse choices and cognitive ability as reflected by test scores. Their result does not appear to be driven by errors in computing expected values. To shed light on channels and potential causality, they use two interventions to manipulate cognitive resources available for decision-making: a distracting task and a requirement for explaining the reasons for decisions. Their evidence is consistent with behavior stipulated by two-system models (as mentioned earlier, an emotional/impulsive system and a deliberative/calculating system) as cognitive load induces more risk-averse choices. In the context of the conceptual discussion above, however, it is unclear whether this can be interpreted as evidence for a causal relationship between latent risk preference and cognitive ability.

In a recent study, Andersson, Holm, Tyran, and Wengström (2016) reinvestigate the tendency for individuals with low cognitive ability to be relatively more risk averse in choice experiments. They study a representative sample of Danish adults, and measure cognitive ability with a variation of Raven's Progressive Matrices as well as a cognitive reflection test. They present members of the sample with one of two sets of risky choices (multiple price lists). In both multiple price lists, subjects make ten choices between a relatively safe lottery, which is held constant across choices within a price list, and a more risky lottery, for which the spread and expected value is increased from row to row. For example, in the first multiple price list, the first choice is between a lottery with a 50:50 chance of winning 30 or 50 Danish krone (DKK), and a lottery with a 50:50 chance of winning 5 or 60 DKK. In each subsequent row, the spread and expected value of the latter riskier lottery is increased by raising the high outcome of the lottery by 10 DKK (that is, the risky lottery in the second row offers a 50:50 chance of winning 5 or 60 DKK). In both price lists, risk-averse and risk-neutral individuals initially prefer the safer lottery, but switch to preferring the riskier lottery later in the choice table.

The two multiple price lists differ in terms of how often a risk-neutral person would prefer the safe lottery before switching to the riskier one. In the first set, a risk-neutral individual would start preferring the riskier lottery in row 3, whereas in the second set the same individual would switch to favoring the riskier lottery in row 7. Andersson et al. find that individuals with lower cognitive ability are relatively more risk averse in the first set (that is, they switch relatively late from preferring the low-spread lottery to preferring the high-spread lottery). By contrast, in the second set, low cognitive ability is associated with being relatively less risk averse. Andersson et al. argue that these findings reflect a tendency for those with low cognitive ability to make random choice errors. For example, a risk-neutral person should prefer the safe lottery only in the first two choices of the first price list, but not in the other eight choices. Random choice errors could lead to risk-seeking behavior in the first two choices, but would induce risk-averse decisions in the other eight choices. In

the second multiple price list, however, random choices will tend to introduce bias towards more risk-seeking behavior.

Andersson et al. (2016) suggest that one interpretation is that the correlation between risk aversion and cognitive ability found in earlier work might be spurious. In our view their results do nicely illustrate a type of noise in measuring latent risk preference, arising from how the choice architecture can systematically influence risky choice in such measures. The results also demonstrate, however, that cognitive ability is related systematically to risky choice and risk aversion, in a way that is mediated by the nature of the choice architecture. Indeed, they replicate previous findings about the relationship between cognitive ability and risk aversion in the type of choice setting used in previous work, where there are relatively more opportunities to choose in a risk-averse than a risk-loving way. The interesting result is that this relationship can flip when the choice setting changes. Their results raise interesting questions about which type of nonexperimental choice settings might foster systematic risk aversion, or risk-seeking behavior, among those with low cognitive ability. As with the rest of this literature, the Andersson et al. (2016) findings do not in our view warrant conclusions about the relationship between latent traits of risk preference and cognitive ability, because it is not identified whether cognitive ability affects lottery choices solely through other channels than risk preference (for example, through mistakes in decision-making) or whether cognitive ability impacts latent risk preference. Nevertheless, their results indicate the difficulties of identifying the nature of the relationship between (latent and measured) risk preference and cognitive ability.

These difficulties aside, several other patterns emerge from this branch of the literature. First, when considering the results from studies with a range of different sample sizes, the findings indicate that measurement error may be an issue and effect sizes are potentially small. For example, no statistically significant relationship is observed in a number of studies that involve small sample sizes of less than 200 observations (Sousa 2010; Mather et al. 2012; Tymula et al. 2012; Taylor 2016; Pachur, Mata, and Hertwig 2017). While negative correlations between measures of risk aversion in the gain domain and measures of cognitive ability have been found in particular samples, such as trainee truckers (Burks, Carpenter, Goette, and Rustichini 2009) and college students (Benjamin, Brown, and Shapiro 2013), such correlations tend to be stronger in representative adult population samples (Dohmen et al. 2010; Andersson et al. 2016). A likely reason is that larger samples also tend to display more heterogeneity with respect to cognition, which makes it possible to find a stronger correlation.

A second pattern that emerges from the literature reveals differences in the correlation between risk taking in lotteries and cognitive ability depending on whether lotteries only entail gains or also potential losses.[2] In particular, and similar

[2]An example of a lottery with outcomes in the gain domain is: you win 100 with 50 percent chance or you win 900 with 50 percent chance. An example of a lottery in the loss domain is: you lose 100 with 50 percent chance or you lose 900 with 50 percent chance. A risk-neutral person is indifferent between the first lottery and a sure payment of 500, and indifferent between the second lottery and a sure loss of –500.

to the findings in several nonexperimental settings, experimental studies tend to find a negative correlation between risk aversion in lottery choice and various measures of cognitive ability when the lottery outcomes are in the gain domain, whereas the findings suggest a positive correlation between risk aversion and cognitive ability when the lottery outcomes involve potential losses (for example, Rustichini, DeYoung, Anderson, and Burks 2012, 2016; Burks et al. 2009; Oechssler, Roider, and Schmitz 2009; Frederick 2005).

Statistically significant negative correlations between risk aversion and cognitive ability are found for cognition measures related to numeracy (for example, Benjamin, Brown, and Shapiro 2013; Rustichini et al. 2012, 2016), but also for cognition measures such as the ability to solve Raven's Progressive Matrices (Andersson et al. 2016), the Hit 15 game (Burks et al. 2009), speed of recognition and word fluency (Dohmen et al. 2010), the Cognitive Reflection Test (Frederick 2005), as well as standardized achievement test scores (Benjamin, Brown, and Shapiro 2013). Some studies have access to different cognitive ability measures. For instance, in Dohmen et al. (2010), we find that two distinct types of cognitive ability tests both have power for explaining risky choice. Taken together, this evidence could indicate that risk aversion when measured in the gain domain is not linked to one narrow aspect of cognitive ability, but rather a broader trait that underlies the performance in a range of cognitive tasks.

The correlation between measured risk preferences and cognitive ability seems to be particularly strong when quantitative IQ tests are used for its elicitation rather than memory tests or grades. Statistically significant correlations are typically found for cognition measures or IQ tests that accentuate numeracy (for example, Beauchamp, Cesarini, and Johannesson 2017; Benjamin, Brown, and Shapiro 2013; Rustichini et al. 2012, 2016). Weaker findings emerge for tasks involving memory or school grades (for example, Booth and Katic 2013; Angrisani and Casanova 2011). This evidence might also suggest that different facets of cognitive ability play different roles. However, relatively few studies have run horseraces between different cognitive ability measures.

Evidence from Measures of Self-reported Willingness to Take Risks

Many survey studies include measures of cognitive ability and survey instruments to elicit self-reported willingness to take risks, including the National Longitudinal Study of Youth (NLSY); the German Socioeconomic Panel (SOEP); the Dutch DNB Household Survey; Understanding Society—The UK Household Longitudinal Study; and the Survey of Health, Aging and Retirement in Europe (SHARE). Typically, these surveys include a simple question asking individuals to rate their willingness to take risks on a "Likert scale" that offers a range of possible answers. The most widely used format is the question about the respondent's willingness to take risk in general. The simplicity of this general risk question or related formats has the advantage of being easy to understand, thereby limiting the problem of decision errors or noise, while delivering relevant information about risk preferences (Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner 2011). Several studies using

such self-reported measures of individuals' willingness to take risks find that the answers are positively related to measures of cognitive ability (for example, Dohmen et al. 2010; Beauchamp, Cesarini, and Johannesson 2017). Analyzing novel data from the Global Preference Survey, in Falk, Becker, Dohmen, Enke, Huffman, and Sunde (forthcoming), we show that answers to the general risk question are positively related to self-reported math skills in representative population samples of 76 countries. In 55 countries, this relationship is significant at the 1 percent level, and in only 12 countries is it not statistically significant at the 10 percent level.

Own calculations show that average willingness to take risks of adults is also related significantly to average cognitive ability at the country level, as measured by the OECD Programme for the International Assessment of Adult Competencies (PIAAC). PIAAC is designed by the OECD to measure the competencies of adults in numeracy, literacy, and problem solving. In our calculations, willingness to take risks is more pronounced in countries with higher average numeracy (correlation of −0.598) and literacy (correlation of −0.541) for the 16 countries surveyed in both the Global Preference Survey and in PIAAC.

In summary, the findings in this literature suggest that individuals with low cognitive ability view themselves as relatively unwilling to take risks. It is conceivable that this reflects errors in self-assessment, but it is not obvious that low cognitive ability should lead to errors that are systematically biased towards indicating greater risk aversion. If low cognitive ability just adds noise to self-assessments, this would work against finding a relationship between self-reported willingness to take risks and cognitive ability.

Changes in Risk-Taking Behavior and Cognition Related to Aging

A large body of evidence indicates that adults exhibit more risk-averse behavior as they grow older (as discussed by Schildberg-Hörisch in this issue). In an experimental study of 135 subjects ranging in age from 12 to 90, Tymula, Belmaker, Ruderman, Glimcher, and Levy (2013) support the finding that older individuals are more risk averse than their midlife counterparts but also find that adolescents are more risk averse too. Two papers, Falk, Becker, Dohmen, Enke, Huffman, and Sunde (forthcoming) and Mata, Josef, and Hertwig (2016), document that the relationship between age and risk attitudes is found across the globe. At the same time, a large body of psychology and neuroscience finds that performance on a wide variety of cognitive tasks declines over the life course, including processing speed (Salthouse 1996) and working memory (Van der Linden, Brédart, and Beerten 1994).

Some studies have investigated whether this pattern is related to cohort effects rather than an age profile in risk preferences. Studies based on cross-sectional data document that older cohorts are on average less willing to take risks than younger cohorts (Barsky, Juster, Kimball, and Shapiro 1997; Donkers, Melenberg, and van Soest 2001; Dohmen et al. 2011). However, studies based on longitudinal data indicate that the difference in willingness to take risks is not solely driven by a cohort effect (for example, Sahm 2012). Using data from two representative samples from Germany and the Netherlands, in Dohmen, Falk, Golsteyn, Huffman, and Sunde

(2017), we disentangle cohort, period, and age effects to investigate how risk attitudes change over the life course, and we provide evidence that individuals do become less willing to take risks as they grow older.

In light of this evidence, a recent strand of literature in economics and neuroscience has begun to relate changes in risk aversion and cognition over the lifespan. As one example, Bonsang and Dohmen (2015) use data from the Survey of Health, Aging and Retirement in Europe (SHARE) that includes both a measure of financial risk preference and measures of cognitive ability for a representative sample of individuals aged 50+ in 11 European countries. They document that the age-related cross-sectional variation in willingness to take risks is not statistically significantly different, once cognition is controlled for. Mata, Josef, Samanez-Larkin, and Hertwig (2011) conduct a meta-analysis and conclude that age-related differences in risk taking vary across tasks and decision context. Decreased learning performance among older adults seems to contribute to age-related differences in risk taking, in particular when objective probability distributions are not known.

Taken as a whole, these findings indicate that the process of cognitive decline is connected with changes in risk taking over the life course. Yet the findings do not necessarily document a causal relationship between cognition and latent risk preferences. As an alternative hypothesis, cognition and risk preferences might both be affected by simultaneous and interrelated physiological processes. Neuropsychological evidence indicates that the human brain changes not only during early childhood but also at later stages (for example, Giedd et al. 1999; Best and Miller 2010), and this is associated with executive control but also other abilities such as social cognition and empathy (Singer 2006) and potentially risk preference. One prominent example of how risky choice behavior varies with the development of the socio-emotional system in the brain is in the context of puberty (Steinberg 2008).

Perspectives for Future Research

Our discussion suggests two main avenues for future research. The first is to delve deeper into the physiological foundations of decision-making. Given the challenges that arise from the revealed preference approach to measuring latent risk preference and cognitive ability, it appears natural to turn to alternative approaches to measuring such traits, applying recently developed methodologies from neuroscience. Research in brain science builds on the idea that both latent cognitive ability and latent risk preference are rooted in the brain and seeks to use different regions of the brain as the unit of analysis. Ideally, it would be possible to infer from activity in a given brain region the degree of latent risk preference, and from another brain region the level of latent cognitive ability, and assess whether and how these are related. Even more ambitious, one could test causality by intervening directly in the brain in a way that influences exclusively either risk preference or cognitive ability (but not both). Given the current state of knowledge in neuroscience about the

function of different brain regions, and the available technology for measuring and intervening in the brain, however, we seem relatively far from being able to address the question of interest.

Neuroeconomic studies have made some progress in identifying parts of the brain and neural circuits that are activated when individuals are making a choice related to risk or uncertainty in particular decision contexts. These studies typically use functional magnetic resonance imaging (fMRI) to observe changes in brain activity (as measured by oxygen saturation of blood in certain regions of the brain) when aspects of choice alternatives (like probabilities or reward size) are varied, in order to identify brain regions that are associated with risky choice. This literature, as reviewed in Tobler and Weber (2014), indicates that risky choice is associated with activity in the valuation system, including dopamine neurons, striatum orbitofrontal cortex, medial prefrontal cortex, and posterior cingulate and parietal cortex (see also d'Acremont and Bossaerts 2008; Platt and Huettel 2008; Rushworth and Behrens 2008; Schultz et al. 2008; Schultz, O'Neill, Tobler, and Kobayashi 2011). Another approach, reviewed in Levasseur-Moreau and Fecteau (2012), has been to use noninvasive brain stimulation such as transcranial direct current stimulation (tDCS) or transcranial magnetic stimulation (TMS) to investigate whether stimulating certain brain regions can influence risk-taking behavior. However, even if a brain region is identified as being pivotal for risk-taking behavior, this brain region does not necessarily (or exclusively) represent the location of the underlying latent level of risk preference.[3]

There is also evidence that the brain regions involved in risky choice change when the study involves more complex risky choices rather than the simple lotteries that are typically used in the literature, raising questions about the location of latent risk preference in the brain (Tobler and Weber 2014). Studies with nonhuman primates and rats suggest that risk and probabilities are processed by single neurons that are distributed throughout the brain (for a review, see Burke and Tobler 2011). Because these neurons are located in brain regions that are connected, a possibility exists that risk preference is encoded in a neural system distributed throughout the brain, complicating the study of factors that might affect this predisposition. Analogously, latent cognitive ability is also potentially determined in a distributed neural system.

Clearly, analyzing the nature of the relationship between risk preference and cognition, given the limitations of the revealed preference approach and the state of our current understanding of brain structure and brain functioning, is very challenging. However, this complexity is often not spelled out in empirical studies of these topics, where observed measures of risk preference are often treated as

[3]An alternative approach to manipulating cognitive ability is to use tasks that increase or decrease the overall cognitive load needed to make decisions. This approach does not require impinging upon the brain directly. However, higher cognitive load has been shown to dampen emotional responses along with performance on cognitive tasks (DeFraine 2016). Thus, when using cognitive load to measure changes in cognitive level, the researcher may be finding out how shifts of emotion influence *risk taking*, rather than how cognitive ability does so.

unbiased proxies for latent risk preference. It would be useful for researchers to formulate more tightly focused questions that concern the relationship between measured risk preference and measures of cognition.

Thus, a second avenue for future research is to extend the existing work using measures of risk preferences and cognitive ability and address the (many) open questions that the existing literature has raised. One aspect worthy of more work is the development and use of richer sets of measures and elicitation instruments and potentially applying more data-intensive methodologies. Some recent studies have made progress in this direction and applied several of the various ways to measure risk preferences from nonexperimental, experimental, and survey approaches jointly in order to isolate a single factor that measures risk (for example, Frey, Pedroni, Mata, Rieskamp, and Hertwig 2017). This approach, which is akin to measuring a latent trait (or single manifold) of intelligence by a battery of different tests, seems to be useful in that the impact of idiosyncratic factors that confound single measures of risk preference is reduced. With equally rich information about cognitive abilities, it might be possible to gain further insights into the relation between the latent traits of risk preference and cognitive ability. This would also help with reconciling the contradictory evidence about the correlation between risk preference and cognitive ability in choice situations in the gain or in the loss domain, and it would help with the issue of decision errors.

Similarly, it could prove valuable to explore the relationship between measured risk preferences and cognitive ability across a wider range of contexts. For example, one can imagine a range of experiments that would vary how the risky scenario is described to subjects, to assess whether this influences risky choice in a manner consistent with cognitive differences. Along these lines, Rabin and Weizsäcker (2009) show that difficulties with calculating expected value seem to be a factor underlying "narrow bracketing"—that is, evaluating choices separately in choices across different pairs of lotteries. In addition, it may be fruitful to study whether individuals with low cognitive ability might be more likely to violate the transitivity assumption when choosing among different lotteries (Rustichini 2015) and whether they are less likely to act upon concepts such as stochastic dominance. One strand of this literature documents that findings of such violations and inconsistent choice behavior are more prevalent among low-cognitive-ability types (for example, Burks et al. 2009; Eckel 1999; Huck and Weizsäcker 1999; Dave, Eckel, Johnson, and Rojas 2010).

These kinds of experiments might be linked to a public policy context. For example, varying the presentation of retirement or benefit options offered by a firm to its employees, or the sequence of benefit options being offered by a government, might influence choice among individuals with lower cognitive ability. A researcher might also study interventions to improve numeracy or other cognitive abilities, and whether they lead to differences in risky lottery choices.

To sum up, this article has argued that economists and psychologists should increase their awareness about the connections between cognitive ability and revealed (measured) risk preference. Using the concept of revealed risk preference

without also including the concept of cognitive ability runs the risk of confusion, both in the theory that is used to analyze decision-making in conditions of risk and in the interpretation of empirical work on risk preference.

As these connections are gradually uncovered in future research, the answers are likely to have widespread applications to economic analysis and public policy. For example, if it was established that cognitive ability does cause risky choice, at least in a given setting, then under relatively mild assumptions about the function relating the two (like monotonicity and independence), one would be able to generalize the relationship to many different choice settings. However, it may turn out that the relationship between cognition and risk preference differs depending on the setting—and in particular, on the specific task that is being carried out in a risky environment. If this is true, it will be harder to extrapolate results from one study to another. Looking at connections between cognition and different kinds of tasks may well illuminate the specific mechanisms connecting the two. Thus, this research area should seek to map out the relationship of these constructs across a broader range of settings than has been studied so far. It may also turn out that while risk preference and cognitive ability are related, they have meaningfully distinct effects on choices and outcomes. Then, the nature of the correlation between these traits in different settings can have implications for mitigating or exacerbating inequality. We know that there is considerable individual heterogeneity in risk preferences and in behavior in risky settings. Looking at variation in measures of cognitive ability and relating it to measures of risk preference in different decision environments may help to explain this heterogeneity.

Finally, delving into the relationship between cognitive ability and risky choice seems likely to have implications for making and testing public policy. For example, it may become possible to point to specific facets of cognition that are most strongly related to risk-taking behavior. If these specific cognitive tasks can be influenced by training or by gaining practice in that setting, appropriate interventions could alter risky behaviors. Similarly, changing choice architectures or frames could affect risky choice in different ways depending on variations in cognitive ability. To the extent that risky choices are correlated with other outcomes of interest, the connections between cognitive ability and risk preference could be used to target policy interventions.

■ *The authors thank Dan Benjamin, Daniel Krähmer, Sven Rady, and Philippe Tobler for helpful discussions, and the editors Gordon Hanson, Enrico Moretti, and Timothy Taylor for their valuable comments and suggestions. We are also grateful to Radost Holler for providing excellent research assistance. Financial support by Deutsche Forschungsgemeinschaft through CRC TR 190 (Sunde) and CRC TR 224 (Dohmen and Falk) is gratefully acknowledged.*

References

Ackermann, Phillip L., and Eric D. Heggestad. 1997. "Intelligence, Personality, and Interests: Evidence for Overlapping Traits." *Psychological Bulletin* 121(2): 219–45.

Agarwal, Sumit, and Bhashkar Mazumder. 2013. "Cognitive Abilities and Household Financial Decision Making." *American Economic Journal: Applied Economics* 5(1): 193–207.

Andersson, Ola, Håkan J. Holm, Jean-Robert Tyran, and Erik Wengström. 2016. "Risk Aversion Relates to Cognitive Ability: Preferences or Noise?" *Journal of the European Economic Association* 14(5): 1129–54.

Angrisani, Marco, and Maria Casanova. 2011. "Understanding Heterogeneity in Household Portfolios: The Role of Cognitive Ability and Preference Parameters." May. http://www.webmeets.com/files/papers/SAEe/2011/54/Angrisani_Casanova_2011.pdf.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112(2): 537–79.

Beauchamp, Jonathan P., David Cesarini, and Magnus Johannesson. 2017. "The Psychometric Properties of Measures of Economic Risk Preferences." *Journal of Risk and Uncertainty* 54(3): 203–37.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro. 2013. "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11(6): 1231–55.

Best, John R., and Patricia H. Miller. 2010. "A Developmental Perspective on Executive Function." *Child Development* 81(6): 1641–60.

Bonsang, Eric, and Thomas Dohmen. 2015. "Risk Attitude and Cognitive Aging." *Journal of Economic Behavior & Organization* 112: 112–126.

Booth, Alison L., and Pamela Katic. 2013. "Cognitive Skills, Gender and Risk Preferences." *Economic Record* 89(284): 19–30.

Boyer, Ty W. 2006. "The Development of Risk-taking: A Multi-Perspective Review." *Developmental Review* 26(3): 291–345.

Burke, Christopher J., and Philippe C. Tobler. 2011. "Coding of Reward Probability and Risk by Single Neurons in Animals." *Frontiers in Neuroscience* vol. 5. https://doi.org/10.3389/fnins.2011.00121.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini. 2009. "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment." *PNAS* 106(19): 7745–50.

Cattell, Raymond B. 1971. *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin.

Cattell, Raymond B. 1987. *Intelligence: Its Structure, Growth and Action*. Amsterdam, The Netherlands: Elsevier.

Christelis, Dimitris, Tullio Jappelli, and Mario Padula. 2010. "Cognitive Abilities and Portfolio Choice." *European Economic Review* 54(1): 18–39.

Cole, Shawn, Anna Paulson, Gauri K. Shastry. 2014. "Smart Money? The Effect of Education on Financial Outcomes." *Review of Financial Studies* 27(7): 2022–51.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78(3): 883–931.

d'Acremont, Mathieu, and Peter Bossaerts. 2008. "Neurobiological Studies of Risk Assessment: A Comparison of Expected Utility and Mean-Variance Approaches." *Cognitive, Affective, & Behavioral Neuroscience* 8(4): 363–74.

Dave, Chetan, Catherine C. Eckel, Cathleen A. Johnson, and Christian Rojas. 2010. "Eliciting Risk Preferences: When is Simple Better?" *Journal of Risk and Uncertainty* 41(3): 219–43.

DeFraine, William C. 2016. "Differential Effects of Cognitive Load on Emotion: Emotion Maintenance versus Passive Experience." *Emotion* 16(4): 459–67.

Dohmen, Thomas, Armin Falk, Bart H. H. Golsteyn, David Huffman, and Uwe Sunde. 2017. "Risk Attitudes across the Life Course." *Economic Journal* 127(605): F95–F116.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3): 1238–60.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2012. "The Intergenerational Transmission of Risk and Trust Attitudes." *Review of Economic Studies* 79(2): 645–77.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522–50.

Donkers, Bas, Bertrand Melenberg, and Arthur Van Soest. 2001. "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *Journal of Risk and Uncertainty* 22(2): 165–95.

Eckel, Catherine. 1999. "Commentary on 'The Effects of Financial Incentives in Experiments:

A Review and Capital-Labor Production Framework.'" *Journal of Risk and Uncertainty* 19(1–3): 47–48.

Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. Forthcoming. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics.*

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19(4): 25–42.

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3(10): e1701381.

Frisell, Thomas, Yudi Pawitan, and Niklas N. Långström. 2012. "Is the Association between General Cognitive Ability and Violent Crime Caused by Family-Level Confounders?" *PloS ONE* 7(7): e41783.

Giedd, Jay N., Jonathan Blumenthal, Neal O. Jeffries, F. X. Castellanos, Hong Liu, Alex Zijdenbos, Tomáš Paus, Alan C. Evans, and Judith L. Rapoport. 1999. "Brain Development during Childhood and Adolescence: A Longitudinal MRI Study." *Nature Neuroscience* 2(10): 861–63.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa. 2011. "IQ and Stock Market Participation." *Journal of Finance* 66(6): 2121–64.

Hasting, Justine S., Brigitte C. Madrian, and William L. Skimmyhorn. 2013. "Financial Literacy, Financial Education, and Economic Outcomes." *Annual Review of Economics* 5: 347–73.

Heckman, James J. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35(1): 1–97.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 106(6): 2052–86.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–55.

Huck, Steffen, and Georg Weizsäcker. 1999. "Risk, Complexity, and Deviations from Expected-Value Maximization: Results of a Lottery Choice Experiment." *Journal of Economic Psychology* 20(6): 699–715.

Kosse, Fabian, Thomas Deckers, Hannah Schildberg-Hörisch, and Armin Falk. 2016. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." IZA Discussion Paper 9861, Institute for the Study of Labor (IZA).

Kubička, Ludek, Zdeněk Matějček, Zdeněk Dytrych, and Zdeněk Roth . 2001. "IQ and Personality Traits Assessed in Childhood as Predictors of Drinking and Smoking Behaviour in Middle-Aged Adults: A 24-Year Follow-up Study." *Addiction* 96: 1615–28.

Levasseur-Moreau, Jean, and Shirley Fecteau. 2012. "Translational Application of Neuromodulation of Decision-making." *Brain Stimulation* 5(2): 77–83.

Lusardi, Annamaria, and Olivia S. Mitchell. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature* 52(1): 5–44.

Mata, Rui, Anika K. Josef, and Ralph Hertwig. 2016. "Propensity for Risk Taking across the Life Span and around the Globe." *Psychological Science* 27(2): 231–243.

Mata, Rui, Anika K. Josef, Gregory R. Samanez-Larkin, and Ralph Hertwig. 2011. "Age Differences in Risky Choice: A Meta-Analysis." *Annals of the New York Academy of Sciences* 1235(1): 18–29.

Mather, Mara, Nina Mazar, Marissa A. Gorlick, Nicole R. Lighthall, Jessica Burgeno, Andrej Schoeke, and Dan Ariely. 2012. "Risk Preferences and Aging: The 'Certainty Effect' in Older Adults' Decision Making." *Psychology and Aging* 27(4): 801–816.

Oechssler, Jörg, Andreas Roider, and Patrick W. Schmitz. 2009. "Cognitive Abilities and Behavioral Biases." *Journal Economic Behavior & Organization* 72(1): 147–52.

Pachur, Thorsten, Rui Mata, and Ralph Hertwig. 2017. "Who Dares, Who Errs? Disentangling Cognitive and Motivational Roots of Age Differences in Decisions Under Risk." *Psychological Science* 28(4): 504–518.

Platt, Michael L., and Scott A. Huettel. 2008. "Risky Business: The Neuroeconomics of Decision Making under Uncertainty." *Nature Neuroscience* 11(4): 398–403.

Rabin, Matthew, and Georg Weizsäcker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review* 99(4): 1508–43.

Rushworth, Matthew F. S., and Timothy E. J. Behrens. 2008. "Choice, Uncertainty and Value in Prefrontal and Cingulate Cortex." *Nature Neuroscience* 11(4): 389–97.

Rustichini, Aldo. 2015. "The Role of Intelligence in Economic Decision Making." *Current Opinion in Behavioral Sciences* 5: 32–36.

Rustichini, Aldo, Colin G. DeYoung, Jon E. Anderson, and Stephen V. Burks. 2012. "Toward the Integration of Personality Theory and Decision Theory in the Explanation of Economic and Health Behavior." IZA Discussion Paper 6750.

Rustichini, Aldo, Colin G. DeYoung, Jon E. Anderson, and Stephen V. Burks. 2016. "Toward the Integration of Personality Theory and Decision Theory in Explaining Economic Behavior: An Experimental Investigation." *Journal of Behavioral*

and Experimental Economics 64: 122–37.

Sahm, Claudia R. 2012. "How Much Does Risk Tolerance Change?" *Quarterly Journal of Finance* 2(4): 1–33.

Salthouse, Timothy A. 1996. "The Processing-speed Theory of Adult Age Differences in Cognition." *Psychological Review* 103(3): 403–428.

Schultz, Wolfram, Martin O'Neill, Philippe N. Tobler, and Shunsuke Kobayashi. 2011. "Neural Signals for Reward Risk in Frontal Cortex." *Annals of the New York Academy of Sciences* 1239: 109–117.

Schultz, Wolfram, Kerstin Preuschoff, Colin Camerer, Ming Hsu, Christopher D. Fiorillo, Philippe N. Tobler, and Peter Bossaerts. 2008. "Explicit Neural Signals Reflecting Reward Uncertainty." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363(1511): 3801–11.

Singer, Tania. 2006. "The Neuronal Basis and Ontogeny of Empathy and Mind Reading: Review of Literature and Implications for Future Research." *Neuroscience and Behavioral Reviews* 30(6): 855–63.

Sousa, Sergio. 2010. "Are Smarter People Really Less Risk Averse?" CeDEx Discussion Paper Series 2010-17.

Steinberg, Laurence. 2008. "A Social Neuroscience Perspective on Adolescent Risk-Taking." *Development Review* 28(1): 78–106.

Taylor, Matthew P. 2016. "Are High-Ability Individuals Really More Tolerant of Risk? A Test of the Relationship Between Risk Aversion and Cognitive Ability." *Journal of Behavioral and Experimental Economics* 63: 136–47.

Tobler, Philippe N., and Elke U. Weber. 2014. "Valuation for Risky and Uncertain Choices." Chap. 9 in *Neuroeconomics: Decision Making and the Brain,* 2nd edition, edited by Paul W. Fehr and Ernst Fehr. London: Academic Press.

Tymula, Agnieszka, Lior A. Rosenberg Belmaker, Amy K. Roy, Lital Ruderman, Kirk Manson, Paul W. Glimcher, and Ifat Levy. 2012. "Adolescents' Risk-taking Behavior Is Driven by Tolerance to Ambiguity." *PNAS* 109(42): 17135–40.

Tymula, Agnieszka, Lior A. Rosenberg Belmaker, Lital Ruderman, Paul W. Glimcher, and Ifat Levy. 2013. "Like Cognitive Function, Decision Making across the Life Span Shows Profound Age-Related Changes." *PNAS* 110(42): 17143–48.

Van der Linden, Martial, Serge Brédart, and Annick Beerten. 1994. "Age-Related Differences in Updating Working Memory." *British Journal Psychology* 85(1): 145–52.

van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101(2): 449–72.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 135–154

Are Risk Preferences Stable?

Hannah Schildberg-Hörisch

The dictum "De gustibus non est disputandum" of Stigler and Becker (1977) has shaped economics for decades. They wrote: (p. 76) "[O]ne does not argue about tastes for the same reason that one does not argue over the Rocky Mountains—both are there, will be there next year, too, and are the same to all men." Stigler and Becker (p. 89) argue that changes in individual behavior can and should be explained by changes in prices, incentives, or constraints, with no need for allowing changes in preferences "with the endless degree of freedom they provide."

This essay in part follows the Stigler and Becker approach by maintaining the assumption that individuals maximize utility based on preferences and constraints, and thus neglects more radical abandonments of the standard economic model of individual decision-making such as the use of heuristics (for example, Tversky and Kahneman 1974; Gigerenzer and Gaissmaier 2011). However, in contrast to the conceptual arguments in favor of stability of preferences by Stigler and Becker (1977), we take the view that the extent to which preferences are stable is ultimately an empirical question. In particular, the focus will be on one core dimension of individual preferences: attitude towards risk. In recent years, economists have started to investigate the stability of risk preferences and the evidence has been growing rapidly.

In this paper, we first discuss key methodological prerequisites for an empirical research agenda on the stability of risk preferences: validated measures of

■ *Hannah Schildberg-Hörisch is Professor of Applied Microeconomics, Heinrich-Heine-University Düsseldorf, Germany. Her email address is schildberg-hoerisch@dice.hhu.de.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.135

doi=10.1257/jep.32.2.135

risk preferences and a precise definition of preference stability. The strict definition of preference stability in economics implies that individual risk preferences are constant over time. We then proceed by offering an alternative conceptual framework for preference stability that builds on research regarding the stability of personality traits in psychology. The definition of stability used in psychology implies high levels of rank-order stability across individuals and not that the individual will maintain the same level of a trait over time. Preference parameters are considered as distributions with a mean that is significantly but less than perfectly stable, plus some systematic variance. This framework accommodates evidence on systematic changes in risk preferences over the life cycle, due to exogenous shocks such as economic crises or natural catastrophes, and due to temporary changes in self-control resources, emotions, or stress.

Research on the stability of (risk) preferences is conceptually at the heart of microeconomics. But in addition, systematic changes in risk preferences have vital real-world consequences, because an individual's willingness to take risks predicts aspects of labor market and health outcomes, addictive behaviors, investment, and migration decisions (Barsky, Juster, Kimball, and Shapiro 1997; Hong, Kubik, and Stein 2004; Bonin, Dohmen, Falk, Huffman,and Sunde 2007; Anderson and Mellor 2008; Kimball, Sahm, and Shapiro 2008; Jaeger, Dohmen, Falk, Huffan, Sunde, and Bonin 2010; Dohmen and Falk 2011; Becker, Deckers, Dohmen, Falk, and Kosse 2012; Dawson and Henley 2015; Hsieh, Parker, and van Praag 2017). For example, more risk-averse individuals are less likely to be self-employed and to invest in stocks, and countries with higher aggregate risk aversion have a lower total factor productivity (Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner 2011; Falk, Becker, Dohmen, Enke, Huffman, and Sunde 2017). The empirical evidence implies that individuals become more risk-averse over the life cycle. Thus, aging societies are likely to have lower levels of self-employment and total factor productivity as well as more conservative saving, investment, and voting behavior, with important implications for macroeconomic performance and political outcomes such as labor market protection or the implementation of reforms. Moreover, economic crises and downturns have been shown to increase risk aversion, possibly reducing self-employment and investments in stocks, which in turn can amplify macroeconomic downturns. In finance, this pattern of countercyclical risk aversion (that is, investors are more risk-averse during recessions than booms) offers an explanation for the long-standing puzzle that the equity risk premium seems to be higher during recessions than booms (Shiller 1981; Campbell and Shiller 1989; Cochrane 2011). As a final illustration, evidence that stress, fear, or cognitive load induce temporarily elevated levels of risk aversion has important implications for consumer protection laws, suggesting a case for cooling-off periods in insurance contracts, for example. In sum, understanding systematic changes in individual risk preferences over time seems to be key for policy design, as well as for a better understanding of individual decision-making and macroeconomic outcomes.

Once we accept the possibility of systematic change in risk preferences, an array of fundamental questions arises: How can we evaluate alternative policy options

or perform welfare analyses when individuals' preferences lack complete stability? Can and should policymakers make use of the malleability of risk preferences to promote behavior changes that are deemed desirable (such as giving up smoking or avoiding teenage pregnancies)? Can economists benefit from insights in personality psychology on the degree of stability and malleability of personality traits, a concept somewhat related to economic preferences? We will touch upon these questions before we conclude by pointing out directions for future research.

Prerequisites for a Research Agenda on the Stability of Risk Preferences

Measurement of Risk Preferences

In recent years, economists have started to turn attention to validating their measures of risk preferences—that is, to explicitly documenting (instead of assuming) that commonly used measures of risk preferences fulfill crucial criteria such as internal and external validity (for example, Dohmen et al. 2011; Falk, Becker, Dohmen, Huffman, and Sunde 2016). *Internal validity* (or "convergent validity") implies that different measurement tools of risk preferences (for example, survey-based self-report measures and experimental measures) map into the same underlying construct "risk preferences" and offer a coherent description of the same individual. Measured risk preferences are *externally valid* (or have predictive or behavioral validity) if they have predictive power for actual risky behaviors. For example, Dohmen et al. (2011) document that self-reported risk preferences are a reliable predictor of investment in stocks, self-employment, participation in sports, and smoking, as well as actual risk-taking in an incentivized lottery experiment.

Well-established tools for measuring risk preferences have emerged that allow for comparing distributions of risk preferences across studies and populations. Two approaches prevail: self-reports and incentivized experiments.[1] A prototypical example of a self-reported questionnaire measure is the corresponding question in the German Socio-Economic Panel that is answered using an 11-point Likert scale: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (Wagner, Frick, and Schupp 2007). Some of the most widely used experimental approaches to measure risk preferences encompass Holt and Laury's (2002) price list approach, Gneezy and Potters' (1997) risky investment task, as well as Binswanger's (1980) and Eckel and Grossman's (2002) choice between different gambles. In these experiments, individuals typically choose between different two-outcome lotteries in which higher expected payoffs come at the cost of a higher variance of payoff (that is, more risk).

[1] A third approach infers willingness to take risks from field behavior such as investment decisions. However, such measures do not provide an isolated measure of risk preferences but reflect risk preferences, beliefs about the extent of riskiness of a given behavior, and opportunities to engage in a given behavior, all at the same time.

There are clear-cut tradeoffs between using self-reported survey or incentivized experimental measures of risk preferences (see also Charness, Gneezy, and Imas 2013). Experimental economists consider experiments as the methodological gold standard for measuring preferences since experiments observe real choices with real incentives in well-controlled decision situations that are comparable across individuals. By assigning probabilities to each outcome, experimental measures precisely quantify the risks under consideration, while survey measures might capture risk perception on top of risk preferences. However, experimental measures are costly and time-consuming to implement in large representative samples. Therefore, they often rely on single-item measurement, making them more prone to measurement error. In Gerhardt, Schildberg-Hörisch, and Willrodt (2017), my coauthors and I measure risk preferences by the monetary amount that is needed to make an individual indifferent between two-outcome lotteries with varying expected payoffs and payoff spreads ("risk premium"). We find pairwise Pearson's correlation coefficients on within-subject risk premia in four different choice lists (that is, within the same experimental task format) that range between 0.27 to 0.57. Administering multiple measures of risk preferences to the same individuals, Frey, Pedroni, Mata, Rieskamp, and Hertwig (2017) document partial correlations among seven different experimental measures that are below 0.1 on average compared to around 0.2 among 22 different self-report measures. From this perspective, it's not a surprise that survey measures tend to outperform experimental measures in terms of external validity (for example, Dohmen et al. 2011; see also Mata, Frey, Richter, Schupp, and Hertwig in this issue, who generally take a more skeptical view on experimental measures).

Typically, economists use measures of risk preferences that aim at eliciting an "overall" risk preference, reflecting the common assumption in economics that a single risk attitude governs risk-taking in all risk-related domains such as financial investments, and health- or job-related risks. However, the existence of a single risk preference across all risk-related domains is not undisputed (for example, Weber, Blais, and Betz 2002; Hanoch, Johnson, and Wilke 2006).[2] In a representative sample of the German population, Dohmen et al. (2011) find correlations of about 0.5 across different risk domains such as financial matters, health, career, sports and leisure, and car driving as well as for general risk preference and domain-specific ones. They take this finding as evidence that risk preferences across domains are correlated strongly, but far from perfectly.

In my view, the recent economic research on measurement of risk preferences lays the groundwork for studying stability of risk preferences (for a more skeptical judgment, see Friedman, Issac, James, and Sunder 2014). Nevertheless, there is still a lot of scope for improvement in measurement tools for (risk) preferences. In particular, economists can benefit from adopting the psychometric standards applied to personality traits in psychology for the case of economic preferences (Borghans,

[2]Weber, Blais, and Betz (2002) provide a domain-specific risk-taking (DOSPERT) scale whose 40 items measure risk-taking in the domains of recreational, health/safety, social, and ethical risk; gambling; and investments.

Duckworth, Heckman, and Bas ter Weel 2008). For example, economists should search for the measures with highest test–retest stability in shorter time intervals in panel data in order to reduce measurement error and apply them more broadly. Moreover, if information from multiple survey items, multiple experiments, or both kinds of measurement tools are combined to obtain a single measure of individual risk preference, it may help reduce measurement error.

Definition of Preference Stability in Economics

In economic theory, stability of preferences is defined as stability at the level of the individual (as opposed to stability of the distribution of preferences in a given population). Stability of risk preferences implies that, in the absence of measurement error, one should observe the *same* willingness to take risks when measuring an individual's risk preferences repeatedly over time. Indeed, a standard approach in economics is to attribute any changes in measured risk preferences to measurement error and to consider them as meaningless noise.

In the common economic risk paradigm,[3] a single parameter is sufficient to characterize an individual's risk preferences. The value of this single parameter differs across individuals, spanning the continuum from risk proclivity over risk neutrality to risk aversion, with a large majority of individuals being risk-averse (for example, Dohmen et al. 2011). In the subjective expected utility theory framework (Savage 1954), risk preferences are completely characterized by a parameter that describes the curvature of an individual's utility function. Similarly, in the model-independent concept of risk apportionment (Eeckhoudt and Schlesinger 2006), an individual is classified as risk-averse if that individual prefers a particular lottery to a mean-preserving spread of that lottery. Conversely, an individual who prefers the mean-preserving spread over the original lottery is classified as risk seeking. The intensity of a subject's risk attitude is measured by the monetary amount—the "risk premium"—that is needed to make the subject indifferent between a lottery and a given mean-preserving spread of that lottery. In both approaches, the standard economic definition of stability of risk preferences implies that, in the absence of measurement error, one should obtain the same estimate of the parameter of interest (curvature parameter or risk premium) when measuring an individual's risk preferences repeatedly.

Concepts of Stability of Personality Traits in Psychology

In personality psychology, "traits" are defined as enduring patterns of behavior, thought, and emotion that are relatively stable over time but differ across individuals (Roberts 2009).[4] Personality traits and economic preferences are related in the

[3] In this article, we neglect alternative theories of decision-making under risk or uncertainty such as prospect theory (Kahneman and Tversky 1979) or ambiguity aversion (Ellsberg 1961). However, many arguments put forward in this article apply to preferences under prospect theory or ambiguity as well.

[4] One commonly used taxonomy of personality traits are the Big Five: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (for discussion, see Costa and McCrae 1992).

sense that both are characteristics of an individual that are assumed to be important predictors for individual decision-making. Almlund, Duckworth, Heckman, and Kautz (2011) and Borghans et al. (2008) provide thoughtful introductions to personality psychology for economists and discuss the relation between economic preferences and personality traits in more depth.

Personality traits are considered stable if they meet the criterion of *rank-order stability*. Rank-order stability implies consistency in the rank ordering of individuals according to the intensity of a given trait across repeated measurements and is typically measured by correlations: in this setting, those who are most risk-averse at one time also tend to be most risk-averse at later times. While a high level of rank-order stability is a defining aspect of a trait, mean-level stability is not. *Mean-level stability* refers to consistency in the average level of a trait over time. It is important to stress that the concept of mean-level stability refers to an "average level" or central tendency in repeated measurement, not an exactly constant parameter value. However, personality psychologists acknowledge the existence of systematic changes in the average level of a trait within individual over time. Such changes might occur due to aging, new experiences, or traumatic events, for example.

Opposing the more traditional conception of personality traits being characterized by a mean or central tendency (measured with some error), Fleeson (2001) argues that personality traits should be conceived as density distributions. He offers three lines of evidence that within-person variability in traits is too large and too systematic to be ignored. First, on average, within-person variability in traits over time is as large as variability in traits between individuals. Second, within-person variability is not only measurement error but represents predictable individual differences in reactions to changes in the situation ("conditional traits"). Finally, for a given individual and a given personality trait, both mean and standard deviation of the distribution are parameters that are stable over time. According to this view, "trait concepts are not threatened in their usefulness by the existence of within-person variability, because of the equally large degree of distributional stability" (p. 1019).

Conceptual Framework and Empirical Evidence

Most studies that report results on the stability of adults' risk preferences in panel data document correlations of an individual's risk preferences across time that are significantly positive but typically moderate (Dohmen, Lehmann, and Pignatti 2016; for a detailed and excellent review, see Chuang and Schechter 2015). The reported correlations range from 0.18 to 0.68 for time horizons varying from a few days to five years. About half of the studies use incentivized experiments as opposed to hypothetical experiments or survey questions. In contrast, correlations of risk preferences over time in Chuang and Schechter (2015), Horowitz (1992), Lönnqvist, Verkasalo, Walkowitz, and Wichardt (2015) are largely positive, but not significant. In this symposium, Mata et al. report correlations of risk preferences

of about 0.5 for any yearly time horizon of up to 10 years based on data from the German Socio-Economic Panel (SOEP).

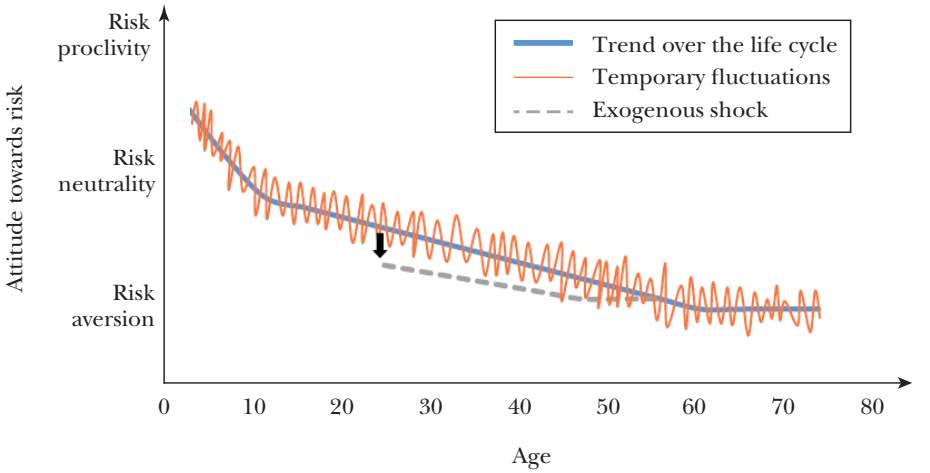
This evidence about the stability of risk preferences can be interpreted as the glass being half-full or half-empty. It is half-full in the sense that the available empirical evidence implies that individual risk preferences do represent a persistent characteristic of an individual that is at least moderately stable over time: correlations over time in panel data are nearly exclusively positive, typically significant, and of medium or large size. It is half-empty because the correlations of risk preferences over time are low enough to cast doubt on the empirical validity of the strict stability definition typically put forward in economics. However, it is important to stress that less-than-perfect correlations in panel data do not provide unequivocal evidence in favor of changing risk preferences. Even if risk preferences were perfectly stable, measurement error in risk preferences could cause low correlations in panel data over time. But given the additional evidence on systematic changes in risk preferences as individuals advance in age, as well as due to exogenous shocks such as economic crises or temporary variations in self-control, emotions, or stress, it seems unlikely that measurement error alone is responsible for the substantial deviations from perfect correlations. Instead, it seems plausible that people's risk aversion changes.

A Conceptual Framework for Preference Stability

In the following, we suggest a framework for studying several possible reasons why an individual's risk aversion may change. The standard economic definition of stability of an individual's risk preferences—absolute stability of a single parameter that is sufficient to characterize an individual's risk preferences—is relaxed in several respects, building on concepts from personality psychology. In particular, we extrapolate Fleeson's (2001) concept of personality traits as distributions to the case of economic preferences and replace the single, constant parameter by a distribution that is characterized by mean and variance (ignoring further moments for the sake of simplicity). Moreover, we do not assume perfect mean-level stability (consistency in the average level of a trait over time). Figure 1 illustrates this framework. For expositional clarity, it focuses on a representative individual and ignores the substantial heterogeneity in risk preferences across individuals. The sketched effect sizes are inspired by empirical findings, and we provide some detailed estimates for each source of variation in risk preferences from selected studies below.

First, the solid line in Figure 1 shows continuous change in the mean-level of risk preferences, reflecting empirical evidence that individuals become more risk-averse over the life cycle. Second, the figure also allows for abrupt mean-level changes in individual risk preferences—as observed in the presence of exogenous shocks like economic crises, natural catastrophes, or violent conflict. This possibility is shown by the dashed line that represents a downward shift of the solid line. Third, the figure displays a distribution around that mean that could be represented by a variance. The variance of the preference distribution allows for temporary variation in risk preferences, which is in line with empirical evidence that temporary

Figure 1
Illustration of the Framework for Studying the Stability of Risk Preferences



Note: This figure illustrates a framework for studying several possible reasons why an individual's risk aversion may change. The solid line shows continuous change in the mean-level of risk preferences, reflecting empirical evidence that individuals become more risk-averse over the life cycle. The dashed line represents a possible downward shift of the solid line for abrupt mean-level changes in individual risk preferences—as observed in the presence of exogenous shocks like economic crises. The jagged line represents temporary variation in risk preferences, in line with empirical evidence that temporary variations in emotions, self-control, or stress cause temporary variation in risk preferences around a baseline or average level.

variations in emotions, self-control, or stress cause temporary variation in risk preferences around a baseline or average level.

Before I discuss the evidence concerning these three reasons for instability of risk preferences, I should stress that this framework does not propose to treat risk preferences as completely stochastic. Age-related changes in adults' risk preferences are modest in size and take place slowly. Exogenous shocks are rare events. Temporary changes in self-control, stress, and emotions induce only temporary, typically small changes in risk preferences. As a result, individual risk preferences are moderately stable over time and sufficiently persistent to be considered an individual trait. However, their degree of stability is too low to be reconciled with the assumption of perfect stability in neoclassical economic theory. Moreover, because change in risk preferences is systematic, it should not be dismissed as meaningless noise.

How Do Risk Preferences Evolve over the Lifetime?

There is a clear pattern of risk preferences over the life cycle: as individuals grow older, they become less willing to take risks. Empirical evidence on risk preferences in childhood and adolescence is largely based on cross-sectional data and

documents systematic changes as children grow. At younger ages, children are more willing to take risks than adults, and a larger share of them behave in a risk-seeking manner (Deckers, Falk, Kosse, and Schildberg-Hörisch 2015; Levin, Hart, Weller, and Harshman 2007; Moreira, Matsushita, and Da Silva 2010; Paulsen, Platt, Huettel, and Brannon 2011). As children grow they become less willing to take risks and in adolescence their risk preferences converge to those of adults (Levin, Hart, Weller, and Harshman 2007; Levin, Weller, Pederson, and Harshman 2007; Paulsen et al. 2011).[5] For example, in Deckers et al. (2015), correlations range from 0.12 to 0.24 for a period of 16 months with an average initial age of 7.8 years, while Levin, Hart, Weller, and Harshman (2007) present a correlation of 0.38 over a three-year period with age ranges of 6–8 and 9–11.

These systematic changes in risk preferences during childhood and adolescence are in line with a standard model of skill formation (Cunha and Heckman 2007). The skills in this model include both cognitive and noncognitive skills—such as risk preferences, patience, self-control, and perseverance. Skill formation is modeled as a dynamic, multistage process. In contrast to a model of stable preferences, children's skills in this model change over time as the result of accumulating investments and the self-reinforcing and cross-fertilizing nature of skills.

From the onset of adulthood to old age, the trend to greater risk aversion continues but is less pronounced (Bucciol and Zarri 2015; Dohmen, Falk, Golsteyn, Huffman, and Wagner 2017; Josef, Richter, Samanez-Larkin, Wagner, Hertwig, and Mata 2016; Sahm 2012; Schurer 2015). Three studies have used panel datasets and disentangle birth cohort and period effects from age effects. Using large representative panel datasets like the DNB Household Survey from the Netherlands and the German Socio-Economic Panel in Germany, which include self-reported measures of the willingness to take risks, Dohmen et al. (2017) find that the willingness to take risks decreases linearly from early adulthood until approximately age 65, after which the slope becomes flatter. In terms of effect size, risk attitudes decrease by about 0.023 standard deviations for each additional year of age. Translating this age effect into an effect on life outcomes, Dohmen et al. (2017) predict that an increase in society's median age of 10 years implies 6 percent less self-employment or 2.5 percent less investment in stocks, ceteris paribus. Building on their work, Schurer (2015) documents that risk tolerance declines strongly for all socioeconomic groups from late adolescence up to age 45. From age 45 onwards, however, the risk tolerance of individuals with high socioeconomic status stabilizes or even increases, while the risk tolerance of individuals with low socioeconomic status continues to drop. Using panel data on hypothetical gambles on lifetime income from the US Health and Retirement Study, Sahm (2012) finds a modest decline in risk tolerance in a sample of older adults (age 45–70).

[5] However, Sutter, Kocher, Glätzle-Rützler, and Trautmann (2013) do not find a significant age trend when studying the risk preferences of 10–18 year-olds.

Because all three of these studies rely on self-reported measures of risk aversion, they do not allow for disentangling whether it is risk preferences, risk perceptions, or constraints (for example, having a higher number of dependents to worry about) that changes with age. Cross-sectional evidence based on hypothetical choices between lotteries (Donkers, Melenberg, and Van Soest 2001) and choices between safe payoffs and lotteries (reviewed in Mata, Josef, Samanez-Larkin, and Hertwig 2011) tends to confirm that risk aversion increases in age, although effect sizes are rather small. This line of research focuses on lotteries with given probabilities and well-defined, rather low payoffs, suggesting that changes in willingness to take risk by age at least partially reflect changes in risk preferences.

Do Exogenous Shocks Affect Risk Preferences Lastingly?

The literature on how exogenous shocks such as natural disasters, violent conflict, or economic crises affect risk preferences is relatively new, but growing quickly.

For negative economic shocks such as the financial crisis in 2008–2009, the evidence rather consistently documents an increase in risk aversion, using a variety of methods. In terms of our framework, such an increase in risk aversion represents an abrupt mean-level change in individual risk preferences. Dohmen, Lehmann, and Pignatti (2016) document this result based on self-reported questionnaire measures of general risk preferences and representative survey data from Germany and Ukraine. Gerrans, Faff, and Hartnett (2015) work with data on financial investors from Australia, the United Kingdom, and the United States who filled out a psychological trait scale that also covers self-assessed financial risk tolerance. Guiso, Sapienza, and Zingales (forthcoming) measure the rise in Italian investors' risk aversion using both a measure of self-assessed financial risk-taking and hypothetical choices between a constant gamble paying €10,000 or zero with equal probability and a sequence of safe payments, similar to the price list approach. In terms of effect size, they find that the risk premium required to accept the risky gamble increased from €1,000 to €2,500 following the crisis. Similarly, the fraction of respondents who say they do not want to take any financial risk rises from 16 to 43 percent. Necker and Ziegelmeyer (2016), working with representative panel data on self-reported financial risk attitudes from Germany, document that households attributing losses to the crisis decreased their risk tolerance.

A potential concern about studies that estimate risk aversion by using the willingness to take financial risks is that it is hard to disentangle whether changes in the willingness to take financial risk reflect changes in risk preferences or beliefs about returns. For example, Weber, Blais, and Betz (2002) document reduced risk-taking of investors due to the financial crisis, which they attribute to changes in subjective expectations of risk and return as opposed to changes in risk preference—which they argue is rather stable in their data. Malmendier and Nagel (2011) provide evidence that is consistent with a beliefs channel, but does not rule out an effect of experiences on risk preferences. All studies suggest that changes in risk preferences due to the financial crisis are not mostly driven by changes in income or wealth

(see also Sahm 2012; Bucciol and Miniaci 2018).[6] Guiso, Sapienza, and Zingales (forthcoming) and Necker and Ziegelmeyer (2016) point at negative emotions as a possible channel for increased risk aversion.

Another strand of the literature studies the relationship between macroeconomic conditions and risk preferences in general, instead of focusing on shocks only. This literature provides evidence for continuous, typically modest, mean-level changes in risk preferences due to continuously changing macroeconomic conditions. For example, Bucciol and Miniaci (2018) use panel data from the representative Dutch Household Survey, self-reported attitudes towards financial risk-taking, and GDP, market returns, and unemployment rates as contemporaneous macroeconomic indicators. Sahm (2012) uses panel data on older adults (age 40–75), hypothetical gambles on lifetime income from the US Health and Retirement Study (HRS) as a measure of risk preferences, and an Index of Consumer Sentiment to proxy the business cycles. Both studies document that individuals are willing to take substantially larger risks during periods of economic growth and are more risk-averse during periods of recession. Malmendier and Nagel (2011) take a more long-term view and examine whether households differ in their willingness to take financial risks depending on the macroeconomic history they experienced over the whole course of their lives. Using repeated cross-section data from the US Survey of Consumer Finances from 1960 to 2007 and controlling for age and year effects, they show that households with higher experienced stock market returns express a higher willingness to take financial risk. More recent experiences receive higher weights, but even returns experienced decades earlier still have some impact. Based on their estimates, the authors extrapolate that for those aged 30 in 2008, the effect of the financial crisis will only have died away after 30 years, pointing at a rather slow fade-out of large shocks.

Research on how natural catastrophes or violent conflict affect risk preferences is inconclusive. The literature review by Chuang and Schechter (2015) finds that natural disasters such as earthquakes, famines, floods, droughts, hurricanes, and tsunamis have been found to either increase risk aversion, or decrease risk aversion, or to have no (consistent) effect on risk preferences. Likewise, the effects of conflict such as civil wars, riots, or political violence show contradictory results, suggesting that conflict may decrease risk aversion (Voors, Nillesen, Verwimp, Bulte, Lensink, and Van Soest 2012) or increase risk aversion (Callen, Isaqzadeh, Long, and Sprenger 2014; Kim and Lee 2014; Moya 2015). Currently, we can only guess about the reasons for these inconsistent results. A large share of the papers that document contradictory effects of violent conflict or natural disasters use experimental data from developing countries, but these tools were typically developed in the context of high-income countries. They may be more likely to produce noisy results in samples that are less educated, partly illiterate, or less used to abstract thinking (for example, Vieider forthcoming; Chuang and Schechter 2015). Moreover, the literature on the effects of natural catastrophes or violent conflict is suffering from a lack

[6] Moreover, major individual-level shocks such as changing labor market status do not seem to affect risk preferences (Sahm 2012; Dohmen, Lehmann, and Pignatti 2016).

of theoretical predictions about the circumstances under which we should expect an increase or decrease in an individual's willingness to take risks. This literature is growing quickly, and in the future, it may become possible to do a meta-analysis that could shed light on the reasons behind the divergent findings.

Are There Systematic but Temporary Variations in Risk Preferences?

A rapidly growing body of research has investigated factors in the decision environment that go beyond prices and constraints, and which might cause systematic but temporary deviations from underlying "baseline risk preferences." This research can be broadly grouped into two areas: temporary variations in an individual's self-control resources, or temporary variation in emotions and stress.

In the area of self-control, a recent class of economic models posits that economic decisions are shaped by the interaction of "dual selves" or "dual systems": a deliberative or long-run and an affective or short-run system (for a review, see Alós-Ferrer and Strack 2014). Several of these models explicitly address decision-making under risk. In particular, lower current levels of self-control resources are assumed to shift the balance of power in favor of the risk-averse short-run self, at the cost of the risk-neutral long-run self. Thus, lower self-control is predicted to induce stronger risk aversion for stakes within a particular range (for example, Fudenberg and Levine 2006, 2011, 2012; Fudenberg, Levine, and Maniadis 2014).[7]

Several laboratory experiments provide causal evidence on the link between self-control and risk preferences by using so-called "ego-depletion tasks" from social psychology (for meta-analyses, see Hagger, Wood, Stiff, and Chatzisarantis 2010; Carter and McCullough 2014; Hagger et al. 2016). Depletion tasks are based on the notion that exerting self-control in one activity consumes self-control resources, thereby increasing self-control costs in subsequent activities (Baumeister, Bratslavsky, Muraven, and Tice 1998). Increased self-control costs in turn will result in lower levels of self-control exertion and thus increase risk aversion according to the models of Fudenberg and Levine (2006, 2011, 2012) and Fudenberg, Levine, and Maniadis (2014). Measuring risk preferences via finely graduated choice lists, in Gerhardt, Schildberg-Hörisch, and Willrodt (2017), my coauthors and I explicitly test hypotheses from the Fudenberg–Levine model. We do not find any evidence for increased risk aversion after self-control depletion, but a small, consistent tendency towards increased willingness to take risks. The same tendency of increased risk-taking under ego-depletion is observed in Friehe and Schildberg-Hörisch (2017), where we measure risk preferences using the risky investment task by Gneezy and Potters (1997). Similarly, Stojić, Anreiter, and Carrasco Martinez (2013) do not find a significant effect of ego-depletion on risk preferences measured via price lists. Benjamin, Brown, and Shapiro (2013, Study 3) and Gerhardt, Biele, Heekeren,

[7]In this literature, the predictions of Loewenstein and O'Donoghue (2005) and Mukherjee (2010) refer to risk-related decision-making under prospect theory as opposed to expected utility theory: for example, Loewenstein and O'Donoghue (2005) predict that lower current levels of self-control induce risk aversion through more pronounced probability weighting.

and Uhlig (2016) use cognitive load manipulations (specifically, memorizing large numbers) that are expected to decrease self-control resources at the time of risky decision-making. Risk preferences are measured by a price list or multiple pairwise lottery choices, respectively. Both studies show that cognitive load induces significantly more risk-averse behavior. In terms of effect sizes, the riskier of two lotteries is chosen in 54 percent of choices in the cognitive load condition and 57 percent in the absence of cognitive load in Gerhardt et al. (2016).

Predictions on how emotions or stress should plausibly influence risk preferences are conflicting. In psychology, the Mood Maintenance Hypothesis (Isen and Patrick 1983) suggests that positive affect induces greater risk aversion, while negative affect leads to a higher willingness to take risks. (The term "affect" refers to conscious, subjectively experienced aspects of an emotion, apart from bodily changes.) The intuition is that individuals in good mood try to protect their good mood by avoiding risks, while individuals in bad mood take risks, trying to improve their mood. The Affect Infusion Model (Forgas 1995) posits the opposite effects. Since risk-related decision-making is a complex process that requires deliberation, this model posits an "affect-priming-mechanism" to be at work—that is, affect may indirectly influence decisions through its influence on selective attention to information or via associative processes. In particular, affect-priming predicts that subjects in bad mood will be more risk-averse than subjects in good mood since they are more attentive to downside risks.

With a few exceptions, empirical evidence on temporary shifts of risk preferences due to changes in emotions or stress tends to be in line with the Affect Infusion Model: that is, negative emotions like fear or stress are typically found to increase risk aversion in studies that establish causal relationships. Most empirical studies on the relation between risk preferences and emotions use a priming approach, so that the resulting shift in emotions can be treated as exogenous. For example, Guiso, Sapienza, and Zingales (forthcoming) induce fear by having students watch a horror movie, then measure risk preferences using a hypothetical choice-list format and find that on average treated students have a 27 percent higher risk premium than untreated ones. Using an incentivized, adapted risky investment task, Cohn, Englemann, Fehr, and Maréchal (2015) prime financial professionals with a financial boom-or-bust scenario and document that subjects are more fearful and substantially more risk-averse in the bust than in the boom condition. In particular, they invest on average 22 percent less into the risky asset in the bust condition than in the boom condition. Moreover, they expose university students to low or high levels of fear by threat of painless or painful electric shocks and show that those with lower levels of fear are willing to take significantly higher risks in ambiguous risky decisions. In Kandasamy et al. (2014), subjects are randomly assigned to taking placebo or hydrocortisone capsules which induce chronic stress over an eight-day period and make incentivized pairwise choices between lotteries. While there is no effect of acute stress (shown by the cortisol response 90 minutes after the first hydrocortisone capsule was taken) on risk preferences, sustained elevation of cortisol leads to greater risk aversion. Cahlíková and Cingl (2017) expose

a randomly chosen subset of participants to a psychosocial stressor in a standard laboratory stress-induction procedure, the Trier Social Stress Test for Groups, and measure risk preferences via multiple price lists. They find that acute psychosocial stress significantly increases risk aversion, a result that is mainly driven by men.[8]

Implications and Conclusion

We started from the premise that it is ultimately an empirical question whether risk preferences are stable over time. The evidence stems from diverse strands of literature, covering the stability of risk preferences in panel data over shorter periods of time, life-cycle dynamics in risk preferences, the possibly long-lasting effects of exogenous shocks on risk preferences as well as temporary variations in risk preferences. Individual risk preferences appear to be persistent and moderately stable over time, but their degree of stability is too low to be reconciled with the assumption of perfect stability in neoclassical economic theory. Inspired by research in personality psychology, we have proposed a framework for preference stability that considers preference parameters as distributions and relies on rank-order stability without imposing mean-level stability. This framework is able to accommodate the empirical evidence on stability and change in risk preferences. Important next steps in research on the stability of risk preferences involve empirical, theoretical, and policy issues.

For empirical research, an important next step would be to seek ways to reduce measurement error in risk preferences. There is a largely unexplored potential for economists to benefit from applying psychometric standards from personality psychology to measures of economic preferences (Borghans et al. 2008; Golsteyn and Schildberg-Hörisch 2017). In particular, it should become standard to measure a single construct like risk preferences with multiple items (experiments and/or questionnaire measures) and to average over those items in order to reduce measurement error. Economists should search for measures of risk preferences with the highest test–retest stability in panel data over shorter periods of time. Moreover, researchers should use measurement tools of risk preferences that have been validated in the context under consideration: for example, a study seeking to measure risk aversion in a largely illiterate sample should use a tool that has been validated in a similar sample.

[8] Among studies that do not find the connection described in the text, Conte, Levati, and Nardi (forthcoming) induce emotions using short film clips and then measure risk preferences by pairwise binary lottery choices and find that fear, sadness, anger, and joviality induce risk-seeking behavior. Also, using the Trier Social Stress Test for Groups, a between-subject design, and incentivized, binary choices between lotteries, Buckert, Schwieren, Kudielka, and Fieback (2014) observe stronger risk proclivity for gains, however only for the small subgroup of participants that show a robust cortisol response to acute stress. As Trautmann (2014) points out, the division of subjects exposed to the stress test into responders and nonresponders possibly induces selection problems and inhibits any causal claims.

As far as theory is concerned, research on the stability of risk preferences might ultimately result in an overarching model of endogenous risk preferences in which risk preferences evolve over time as a function of, among others, aging, exogenous shocks, and changes in the decision environment that encompasses situational factors such as the current level of self-control, stress, or emotions. By acknowledging that individual risk preferences are not carved in stone but may change under specific circumstances, economists would take an important step towards the view commonly held in psychology that the decision environment (beyond incentives and constraints) affects individual decisions.

For many research questions and circumstances, it will still be fine to adopt the simpler textbook model of stable risk preferences as an as-if approach and to investigate how individuals with given preferences react to changes in incentives and constraints. However, systematic changes in risk preferences over time severely complicate policy advice and welfare analysis. After all, policy advice is typically based on predicting (or in case of evaluation, retrospectively documenting) behavioral reactions to institutional changes under the assumption that preferences are stable. Preference changes are a major threat to that approach. To give just one example, consider policymakers who react to a financial crisis with a new regulation. If the financial crisis has induced the population to become more risk-averse, but public policy assumes constant risk preferences before and after the crisis, those proposing and analyzing the new policy will fail to predict behavioral responses adequately.

On the other side, acknowledging these systematic instances of changes in risk preferences opens up new possibilities for public policy. If exogenous shocks affect risk preferences lastingly, policymakers could expose individuals to positive "shocks" in their social environment. For example, such interventions might aim at reducing the high levels of risk proclivity that are associated with drug addiction, criminal activities, or teenage pregnancies. Gutman and Schoon (2013) and Kautz, Heckman, Diris, ter Wel, and Borghans (2014) survey evidence on the effects of interventions such as mentoring programs, social and emotional learning programs, center-based care, residential-based education programs, or programs aimed at improving parenting practices. Some of the surveyed interventions have been found to decrease overly risky *behaviors*, but no evaluation directly measures risk *preferences*. An important prerequisite for the successful timing of interventions in childhood and adolescence is identifying critical periods in which risk preferences are especially malleable (Cunha and Heckman 2007), an open challenge for future research.

Finally, a recognition that risk aversion may shift seems to imply that people have multiple preferences, leaving open the question of which kind of preferences public policy and welfare analysis should rely on. While this article has focused on the stability of risk preferences, similar lines of reasoning apply to other dimensions of economic preferences—in particular time preferences and social preferences—and empirical evidence concerning their stability and systematic instances of changes in these areas is starting to accumulate as well. Evidence of systematic preference changes over time deprives economics of a clean analytical foundation

for assessing the welfare impact of policies and behaviors, and there is no consensus for how to deal with the issue. For example, O'Donoghue and Rabin (2003), Thaler and Sunstein (2003), Glaeser (2006), O'Donoghue and Rabin (2006), Bernheim (2009), Bernheim and Rangel (2009), and Chetty (2015) propose highly divergent approaches for how to do welfare economics if choices do not unambiguously reveal preferences. Others have proposed turning away from welfare criteria that are based on preference satisfaction, suggesting an opportunity-based approach to welfare (as in Sugden 2004) or approaches rooted in happiness economics (for example, O'Donnell, Deaton, Durand, Halpern, and Layard 2014). Addressing the issue of how to conduct research and policy while acknowledging the reality of changing preferences seems certain to present difficult challenges.

■ *I am grateful to Bart Golsteyn, Gordon Hanson, Enrico Moretti, and Timothy Taylor for helpful comments.*

References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." IZA Discussion Papers 5500.

Alós-Ferrer, Carlos, and Fritz Strack. 2014. "From Dual Processes to Multiple Selves: Implications for Economic Behavior." *Journal of Economic Psychology* 41(April): 1–11.

Anderson, Lisa R., and Jennifer M. Mellor. 2008. "Predicting Health Behaviors with an Experimental Measure of Risk Preference." *Journal of Health Economics* 27(5): 1260–74.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112(2): 537–79.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74(5): 1252–65.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4: 453–78.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro. 2013. "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11(6): 1231–55.

Bernheim, B. Douglas. 2009. "Behavioral Welfare Economics." *Journal of the European Economic Association* 7(2–3): 267–319.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124(1): 51–104.

Binswanger, Hans P. 1980. "Attitudes toward Risk: Experimental Measurement in Rural India." *American Journal of Agricultural Econmics* 62(3): 395–407.

Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. 2007. "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics* 14(6): 926–37.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel. 2008. "The Economics

and Psychology of Personality Traits." *Journal of Human Resources* 43(4): 972–1059.

Bucciol, Alessandro, and Raffaele Miniaci. 2018. "Financial Risk Propensity, Business Cycles and Perceived Risk Exposure." *Oxford Bulletin of Economics and Statistics* 80(1): 160–83.

Bucciol, Alessandro, and Luca Zarri. 2015. "The Shadow of the Past: Financial Risk Taking and Negative Life Events." *Journal of Economic Psychology* 48(June): 1–16.

Buckert, Magdalena, Christiane Schwieren, Brigitte M. Kudielka, and Christian J. Fiebach. 2014. "Acute Stress Affects Risk Taking But Not Ambiguity Aversion." *Frontiers in Neuroscience* vol. 8. https://www.frontiersin.org/article/10.3389/fnins.2014.00082.

Cahlíková, Jana, and Lubomír Cingl. 2017. "Risk Preferences under Acute Stress." *Experimental Economics* 20(1): 209–236.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger. 2014. "Violence and Risk Preference: Experimental Evidence from Afghanistan." *American Economic Review* 104(1): 123–48.

Campbell, John Y., and Robert J. Shiller. 1989. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1(3): 195–228.

Carter, Evan C., and Michael E. McCullough. 2014. "Publication Bias and the Limited Strength Model of Self-Control: Has the Evidence for Ego Depletion Been Overestimated?" *Frontiers in Psychology* 5: 823.

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. "Experimental Methods: Eliciting Risk Preferences." *Journal of Economic Behavior and Organization* 87: 43–51.

Chetty, Raj. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105(5): 1–33.

Chuang, Yating, and Laura Schechter. 2015. "Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results." *Journal of Development Economics* 117: 151–70.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance* 66(4): 1047–1108.

Cohn, Alain, Jan Engelmann, Ernst Fehr, and Michel André Maréchal. 2015. "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals." *American Economic Review* 105(2): 860–85.

Conte, Anna, M. Vittoria Levati, and Chiara Nardi. Forthcoming. "Risk Preferences and the Role of Emotions." *Economica*.

Costa, Paul T., and Robert R. McCrae. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97(2): 31–47.

Dawson, Chris, and Andrew Henley. 2015. "Gender, Risk, and Venture Creation Intentions." *Journal of Small Business Management* 53(2): 501–15.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Hannah Schildberg-Hörisch. 2015. "How Does Socio-Economic Status Shape a Child's Personality?" IZA Discussion Papers 8977.

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101(2): 556–90.

Dohmen, Thomas, Armin Falk, Bart Golsteyn, David Huffman, and Uwe Sunde. 2017. "Risk Attitudes across the Life Course." *Economic Journal* 127(605): F95–F116.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522–50.

Dohmen, Thomas, Hartmut Lehmann, and Norberto Pignatti. 2016. "Time-Varying Individual Risk Attitudes over the Great Recession: A Comparison of Germany and Ukraine." *Journal of Comparative Economics* 44(1): 182–200.

Donkers, Bas, Bertrand Melenberg, and Arthur Van Soest. 2001. "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *Journal of Risk & Uncertainty* 22(2): 165–95.

Eckel, Catherine C., and Philip J. Grossman. 2002. "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Fisk." *Evolution and Human Behavior* 23(4): 281–95.

Eeckhoudt, Louis, and Harris Schlesinger. 2006. "Putting Risk in Its Proper Place." *American Economic Review* 96(1): 280–89.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75(4): 643–69.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David B. Huffman, and Uwe Sunde. 2017. "Global Evidence on Economic Preferences." NBER Working Paper 23943.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." IZA Discussion Paper 9674.

Fleeson, William. 2001. "Toward a Structure- and

Process-Integrated View of Personality: Traits as Density Distributions of States." *Journal of Personality and Social Psychology* 80(6): 1011–27.

Forgas, Joseph P. 1995. "Mood and Judgment: The Affect Infusion Model (AIM)." *Psychological Bulletin* 117(1): 39–66.

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3(10): e1701381.

Friedman, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder. 2014. *Risky Curves: On the Empirical Failure of Expected Utility.* London: Routledge.

Friehe, Tim, and Hannah Schildberg-Hörisch. 2017. "Self-Control and Crime Revisited: Disentangling the Effect of Self-Control on Risk Taking and Antisocial Behavior." *International Review of Law & Economics* 49: 23–32.

Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review* 96(5): 1449–76.

Fudenberg, Drew, and David K. Levine. 2011. "Risk, Delay, and Convex Self-Control Costs." *American Economic Journal: Microeconomics* 3(3): 34–68.

Fudenberg, Drew, and David K. Levine. 2012. "Timing and Self-Control." *Econometrica* 80(1): 1–42.

Fudenberg, Drew, David K. Levine, and Zacharias Maniadis. 2014. "An Approximate Dual-Self Model and Paradoxes of Choice under Risk." *Journal of Economic Psychology* 41: 55–67.

Gerhardt, Holger, Guido P. Biele, Hauke R. Heekeren, and Harald R. Uhlig. 2016. "Cognitive Load Increases Risk Aversion." 2016–011. *SFB 649 Discussion Paper 2016-011.* http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2016-011.pdf.

Gerhardt, Holger, Hannah Schildberg-Hörisch, and Jana Willrodt. 2017. "Does Self-Control Depletion Affect Risk Attitudes?" *European Economic Review* 100: 463–87.

Gerrans, Paul, Robert Faff, and Neil Hartnett. 2015. "Individual Financial Risk Tolerance and the Global Financial Crisis." *Accounting & Finance* 55(1): 165–85.

Gigerenzer, Gerd, and Wolfgang Gaissmaier. 2011. "Heuristic Decision Making." *Annual Review of Psychology* 62: 451–82.

Glaeser, Edward L. 2006. "Paternalism and Psychology." *University of Chicago Law Review* 73(1): 133–56.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics* 112(2): 631–45.

Golsteyn, Bart H. H., and Hannah Schildberg-Hörisch. 2017. "Challenges in Research on Preferences and Personality Traits: Measurement, Stability, and Inference." *Journal of Economic Psychology* 60: 1–126.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. Forthcoming. "Time Varying Risk Aversion." *Journal of Financial Economics.*

Gutman, Leslie Morrison, and Ingrid Schoon. 2013. "The Impact of Non-Cognitive Skills on Outcomes for Young People: Literature Review." November, 21. Education Endowment Foundation. Available at: http://www.lsce-mena.org/institute-of-education-university-of-london-2013.

Hagger, Martin S., Chantelle Wood, Chris Stiff, and Nikos L. D. Chatzisarantis. 2010. "Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis." *Psychological Bulletin* 136(4): 495–525.

Hagger, Martin S. et al. 2016. "A Multilab Preregistered Replication of the Ego-Depletion Effect." *Perspectives on Psychological Science* 11(4): 546–573.

Hanoch, Yaniv, Joseph G. Johnson, and Andreas Wilke. 2006. "Domain Specificity in Experimental Measures and Participant Recruitment: An Application to Risk-Taking Behavior." *Psychological Science* 17(4): 300–304.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–55.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. 2004. "Social Interaction and Stock-Market Participation." *Journal of Finance* 59(1): 137–63.

Horowitz, John K. 1992. "A Test of Intertemporal Consistency." *Journal of Economic Behavior and Organization* 17(1): 171–82.

Hsieh, Chihmao, Simon C. Parker, and C. Mirjam van Praag. 2017. "Risk, Balanced Skills and Entrepreneurship." *Small Business Economics* 48(2): 287–302.

Isen, Alice M., and Robert Patrick. 1983. "The Effect of Positive Feelings on Risk Taking: When the Chips Are Down." *Organizational Behavior and Human Performance* 31(2): 194–202.

Jaeger, David A., Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, and Holger Bonin. 2010. "Direct Evidence on Risk Attitudes and Migration." *Review of Economics and Statistics* 92(3): 684–89.

Josef, Anika K., David Richter, Gregory R. Samanez-Larkin, Gert G. Wagner, Ralph Hertwig, and Rui Mata. 2016. "Stability and Change in Risk-Taking Propensity across the Adult Life Span." *Journal of Personality and Social Psychology* 111(3):

430–50.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–92.

Kandasamy, Narayanan, Ben Hardy, Lionel Page, Markus Schaffner, Johann Graggaber, Andrew S. Powlson, Paul C. Fletcher, Mark Gurnell, and John Coates. 2014. "Cortisol Shifts Financial Risk Preferences." *PNAS* 111(9): 3608–13.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success." *IZA Discussion Paper 8696*.

Kim, Young Il, and Jungmin Lee. 2014. "The Long-Run Impact of a Traumatic Experience on Risk Aversion." *Journal of Economic Behavior and Organization* 108: 174–86.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2008. "Imputing Risk Tolerance From Survey Responses." *Journal of the American Statistical Association* 103(483): 1028–38.

Levin, Irwin P., Stephanie S. Hart, Joshua A. Weller, and Lyndsay A. Harshman. 2007. "Stability of Choices in a Risky Decision-Making Task: A 3-Year Longitudinal Study with Children and Adults." *Journal of Behavioral Decision Making* 20(3): 241–52.

Levin, Irwin P., Joshua A. Weller, Ashley A. Pederson, and Lyndsay A. Harshman. 2007. "Age-Related Differences in Adaptive Decision Making: Sensitivity to Expected Value in Risky Choice." *Judgment and Decision Making* 2(4): 225–33.

Loewenstein, George, and Ted O'Donoghue. 2005. "Animal Spirits: Affective and Deliberative Processes in Economic Behavior." May, https://odonoghue.economics.cornell.edu/files/2015/10/will5_05-227fjlg.pdf.

Lönnqvist, Jan-Erik, Markku Verkasalo, Gari Walkowitz, and Philipp C. Wichardt. 2015. "Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison." *Journal of Economic Behavior and Organization* 119: 254–266.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics* 126(1): 373–416.

Mata, Rui, Anika K. Josef, Gregory R. Samanez-Larkin, and Ralph Hertwig. 2011. "Age Differences in Risky Choice: A Meta-Analysis." *Annals of the New York Academy of Sciences* 1235(1): 18–29.

Moreira, Bruno, Raul Matsushita, and Sergio Da Silva. 2010. "Risk Seeking Behavior of Preschool Children in a Gambling Task." *Journal of Economic Psychology* 31(5): 794–801.

Moya, Andrés. 2015. "Violence, Psychological Trauma, and Induced Changes in Risk Attitudes in Colombia." http://andresmoya.weebly.com/uploads/1/3/8/3/13836830/violence_riskaversion_andresmoya.pdf.

Mukherjee, Kanchan. 2010. "A Dual System Model of Preferences Under Risk." *Psychological Review* 117(1): 243–55.

Necker, Sarah, and Michael Ziegelmeyer. 2016. "Household Risk Taking after the Financial Crisis." *Quarterly Review of Economics and Finance* 59: 141–60.

O'Donnell, Gus, Angus Deaton, Martine Durand, David Halpern, and Richard Layard. 2014. "Wellbeing and Policy." Report commissioned by the Legatum Institute, London. March. Available at: https://www.li.com/programmes/the-commission-on-wellbeing-and-policy.

O'Donoghue, Ted, and Matthew Rabin. 2003. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review* 93(2): 186–91.

O'Donoghue, Ted, and Matthew Rabin. 2006. "Optimal Sin Taxes." *Journal of Public Economics* 90(10–11): 1825–49.

Paulsen, David J., Michael L. Platt, Scott A. Huettel, and Elizabeth M. Brannon. 2011. "Decision-Making under Risk in Children, Adolescents, and Young Adults." *Frontiers in Psychology* 2. https://www.frontiersin.org/article/10.3389/fpsyg.2011.00072.

Roberts, Brent W. 2009. "Back to the Future: Personality and Assessment and Personality Development." *Journal of Research in Personality* 43(2): 137–45.

Sahm, Claudia R. 2012. "How Much Does Risk Tolerance Change?" *Quarterly Journal of Finance* 2(4): 1250020.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley.

Schurer, Stefanie. 2015. "Lifecycle Patterns in the Socioeconomic Gradient of Risk Preferences." *Journal of Economic Behavior & Organization* 119: 482–95.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71(3): 421–36.

Stigler, George J., and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum." *American Economic Review* 67(2): 76–90.

Stojić, Hrvoje, Michael R. Anreiter, and José A. Carrasco Martinez. 2013. "An Experimental Test of the Dual Self Model." Unpublished paper, Barcelona, Spain.

Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences." *American*

Economic Review 94(4): 1014–33.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103(1): 510–31.

Thaler, Richard H., and Cass R. Sunstein. 2003. "Libertarian Paternalism." *American Economic Review* 93(2): 175–79.

Trautmann, Stefan T. 2014. "Risk Taking under Stress: The Role(s) of Self-Selection. A Comment on Buckert et al. (2014)." *Frontiers in Neuroscience,* vol. 8. https://www.frontiersin.org/article/10.3389/fnins.2014.00197.

Tversky, Amos, and Daniel A. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124–31.

Vieider, Ferdinand M. Forthcoming. "Certainty Preference, Random Choice, and Loss Aversion: A Comment on 'Violence and Risk Preference: Experimental Evidence from Afghanistan.'" *American Economic Review.*

Voors, Maarten J., Eleonora E. M Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review* 102(2): 941–64.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP)—Scope, Evolution and Enhancements." *SSRN Electronic Journal* 127(1): 139–69.

Weber, Elke U., Ann-Renée Blais, and Nancy E. Betz. 2002. "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perception and Risk Behaviors." *Journal of Behavioral Decision Making* 15(4): 263–90.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 155–172

Risk Preference: A View from Psychology

Rui Mata, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig

Investing in financial markets, engaging in criminal activity, or consuming recreational and possibly illicit drugs are examples of behaviors that involve trading-off potential costs and benefits associated with some degree of risk and uncertainty. Many psychologists aim to uncover the extent to which stable personality characteristics—psychological traits—account for why individuals differ in their appetite for risk and in their decision to engage in such behaviors. The endeavors of psychologists not only reflect an effort to understand human behavior per se, but also aim to better diagnose and prevent undesirable levels of risk taking, with the ultimate goal of improving the physical or mental health and the financial well-being of individuals and populations. In what follows, we use the term "risk preference" to refer to such a psychological trait (or collection of traits) and explore the extent to which both psychologists and economists can use it to explain individual differences in people's appetite for risk.

Debates surrounding the nature of risk preference and its measurement have a long history in psychology and economics, and the number of discussion points

■ *Rui Mata is Professor for Cognitive and Decision Sciences and Renato Frey is a Researcher in Cognitive and Decision Sciences, both at the University of Basel, Basel, Switzerland. Mata and Frey are also Adjunct Researchers, Max Planck Institute for Human Development, Berlin, Germany. David Richter is Research Associate and Jürgen Schupp is the Director of the German Socio-Economic Panel Study, both at the German Institute for Economic Research, Berlin, Germany. Ralph Hertwig is Director of the Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. Their email addresses are rui.mata@unibas.ch, renato.frey@unibas.ch, drichter@diw.de, jschupp@diw.de, and hertwig@mpib-berlin.mpg.de.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.155

doi=10.1257/jep.32.2.155

is large (Bernoulli 1738 [1954]; Edwards 1954; Slovic 1964; Schonberg, Fox, and Poldrack 2011; Friedman, Isaac, James, and Sunder 2014). In psychology, risk preference is commonly defined as the propensity to engage in behaviors or activities that are rewarding yet involve some potential for loss, including substance use, or criminal activities that may be associated with considerable physical and mental harm to individuals (Steinberg 2013). In economics, risk preference more often refers to the tendency to engage in behaviors or activities that involve higher variance in returns, regardless of whether these represent gains or losses, and is often studied in the context of monetary payoffs involving lotteries (Harrison and Rutström 2008). Beyond such differences in definition and scope between fields, which we will not fully address, there are shared and unresolved conceptual and measurement issues overdue for consideration by both fields. We argue that psychology offers conceptual and analytic tools that can help advance the discussion on the nature of risk preference and its measurement in the behavioral sciences. We also provide an overview of strengths and weaknesses of two different measurement traditions of risk preferences that have coexisted in psychology and, to some extent, in economics: the revealed and stated preference traditions (Beshears, Choi, Laibson, and Madrian 2008; Appelt, Milch, Handgraaf, and Weber 2011; Charness, Gneezy, and Imas 2013). Lurking beneath these measurement aspects are broader conceptual issues. Let us briefly preview three before discussing them in more detail in the remainder of the article: temporal stability, convergent validity, and predictive validity.

Psychological traits, by definition, show some degree of *temporal stability*. Consequently, any theorizing about risk preference as a psychological trait must ask whether it shows a degree of stability over time that approximates what has been established for other major traits, such as intelligence, or, alternatively, is more similar to the stability of transitory psychological states, such as emotional states. Of course, no psychological trait is perfectly stable, and it may be subject to systematic variation as a function of specific contextual influences (Caspi, Roberts, and Shiner 2005). Such a view is compatible with our proposal that risk preference can both be seen as a stable psychological trait and yet show systematic and sizable changes as a function of specific life stages or momentary shocks (see also Schildberg-Hörisch, in this issue).

Convergent validity refers to the degree to which different measures of a psychological construct capture a common underlying characteristic or trait. Do measures of risk preference all capture a unitary psychological trait that is indicative of risky behavior across various domains, or do they capture various traits that independently contribute to risky behavior in specific areas of life, such as financial, health, and recreational domains (Weber, Blais, and Betz 2002; Highhouse, Nye, Zhang, and Rada 2017)? This need not be an either–or choice. For example, research on the trait of intelligence suggests that a single general factor can account for the largest share of variance (approximately 50 percent) in performance across many different tasks, with the rest of the variance being accounted for by more specific factors such as visual-spatial or logical-mathematical intelligence (Deary 2001). Similar results have been obtained for psychopathology: About 50 percent of variance in symptomatology is captured by a general factor, which is in line with the fact that about half of

individuals who meet diagnostic criteria for one disorder also meet diagnostic criteria for a second one (Caspi et al. 2014; Castellanos-Ryan et al. 2016). Critically, recent work on risk preference suggests that it may share the psychometric structure of such major psychological traits, by which over 50 percent of the systematic variance in measures of risk preference are accounted for by a general factor, with the remaining variance being shared among several additional specific factors (Frey, Pedroni, Mata, Rieskamp, and Hertwig 2017). Consequently, it may be important to consider the explanatory power of a general trait of risk preference in addition to more specific ones when accounting for individual differences in the appetite for risk.

Predictive validity refers to the extent to which a psychological trait has power in forecasting behavior. For example, intelligence and major personality traits, such as some of the Big Five traits (openness, conscientiousness, extraversion, agreeableness, neuroticism), have been shown to predict important life outcomes, such as academic and professional achievement (Schmidt and Hunter 2004; Richardson, Abraham, and Bond 2012). Such work suggests that it is important to examine the short- and long-term outcomes of risk preference—something that is still largely lacking in current psychological (and economic) research.

In what follows, we discuss the current empirical knowledge on risk preferences in light of these three arguments. However, first, we provide an overview of the revealed and stated preference measurement traditions, which have coexisted in both psychology and economics in the study of risk preferences. Without acknowledging their existence and understanding their somewhat difficult relation, it is hard to make concerted progress in research on risk preference.

Two Measurement Traditions

In his presidential address to the American Psychological Association, Lee Cronbach (1957), a towering figure of 20th century psychology, distinguished between two research streams that run through the history of the still young—and back then even younger—discipline of scientific psychology. One stream, he argued, is *experimental* psychology (see also Hertwig and Ortmann 2001). Its emphasis is on well-controlled experimental designs and on the goal of rigorously testing the influence of selected situational variables on behavior, cognition, and emotion, often using objective measures—such as overt choices and associated reaction times—as outcomes of interest. The other stream, *correlational* psychology, relies on observational and correlational designs to understand cross-situational and intra-individual consistency of the same behavior, cognition, and emotion, often with the aid of self-reports in response to standardized survey measures. Whereas experimenters' interest lies primarily in the impact of the variations they caused, the concern of correlators is with the (co)variation of individuals' behavior across naturally occurring situations.

Six decades later, the partition of psychological research into these two streams is still noticeable (Tracy, Robins, and Sherman 2009)—and perhaps nowhere more so than in research on the construct(s) of risk preference (Appelt et al. 2011; Frey et al.

2017). This distinction is also reflected in two major measurement approaches: one that mostly employs behavioral paradigms, and another that predominantly uses self-reports. These two broad approaches can also be identified, alongside others, in the economics literature (Beshears et al. 2008; Charness, Gneezy, and Imas 2013).[1]

The behavioral stream in psychology focuses on understanding the cognitive or neural correlates of risk preference. This work often emphasizes the structural properties of tasks and environments that are associated with sometimes surprisingly different and even seemingly inconsistent behaviors (Kahneman and Tversky 1979; Mata, Josef, Samanez-Larkin, and Hertwig 2011). For example, a long tradition in both economics and psychology uses choices between lotteries to understand how individuals deal with gains and losses or specific types of incentive structures (Kahneman and Tversky 1979; Holt and Laury 2002). This type of research is alive and well in its somewhat splendid isolation—an issue to which we return shortly. For example, recent experimental efforts have tried to understand the *description–experience gap* that arises from differences in the presentation format of risk information (Hertwig and Erev 2009). For example, the numerical description of risks ("stated probabilities") in canonical lottery tasks gives rise to choices indicative of overweighting of small probabilities, but sequential experience of risk first-hand through sampling of outcomes is associated with choices as if people underweight small probabilities (Wulff, Mergenthaler-Canseco, and Hertwig 2018). A large swath of research now aims to identify the neural basis of choice in such experience-based and description-based paradigms using functional neuroimaging and other neuroscientific methods (Glimcher and Fehr 2014; Knutson and Huettel 2015). Researchers from this approach often focus on uncovering the psychological processes underlying choices in a specific behavioral paradigm, but often with little or no investigation of how such processes generalize across paradigms and time.

Studies using self-report measures seek to elicit *stated preferences* in response to hypothetical or real-world behaviors. For example, respondents may be asked to rate themselves on a rating scale with opposite poles being "not at all willing to take risks" and "very willing to take risks," or express the likelihood of engaging in some risky behavior—"How likely would you be to go white-water rafting at high water in the spring?" A growing body of work on risk preference builds primarily on findings from either single-item (Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner 2011) or multiple-item self-report measures of risk preference (Blais and Weber 2006). For example, this type of data has been used to study stable individual characteristics, such

[1]For the sake of completeness, it should be noted that there are other approaches to studying and measuring risk preference in both psychology and economics. Frey et al. (2017) distinguished between behavioral measures (assessing revealed preferences), self-reported propensity measures (assessing stated preferences), and self-reported frequency measures (tracking specific and observable behaviors). Other approaches include the use of epidemiological data from population statistics, such as crime or cause-specific mortality (Steinberg 2013), actual behavior as captured from administrative or survey data (Moffitt et al. 2011), or observer reports from relatives or acquaintances (Roberts, Lejuez, Krueger, Richards, and Hill 2014). However, the bulk of work on risk preference rests on behavioral and self-report measures, so we focus on those here.

as the genetic basis of risk preference (see also Benjamin et al. 2012; Beauchamp, Cesarini, and Johannesson 2017) as well as to uncover cohort (Malmendier and Nagel 2011; Dohmen, Falk, Golsteyn, Huffman, and Sunde 2017), life span (Josef, Richter, Samanez-Larkin, Wagner, Hertwig, and Mata 2016; Dohmen et al. 2017), and momentary (Browne, Jaeger, Richter, and Steinorth 2016) changes in risk preference. Importantly, such self-report preference measures are now included in a number of panel assessments, such as the German Socio-Economic Panel (Wagner, Frick, and Schupp 2007), the US Health and Retirement Survey (Fisher, Gideon, Hsu, and McFall 2011), the British Household Panel (Galizzi, Machado, and Miniaci 2016), the Swedish Screening Across the Lifespan Twin survey (Beauchamp, Cesarini, and Johannesson 2017), the Swiss Household Panel (Mamerow, Frey, and Mata 2016), and the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Clark and Lisowski 2017). These panel studies are important data troves for revealing more about the associates and determinants of risk preference.

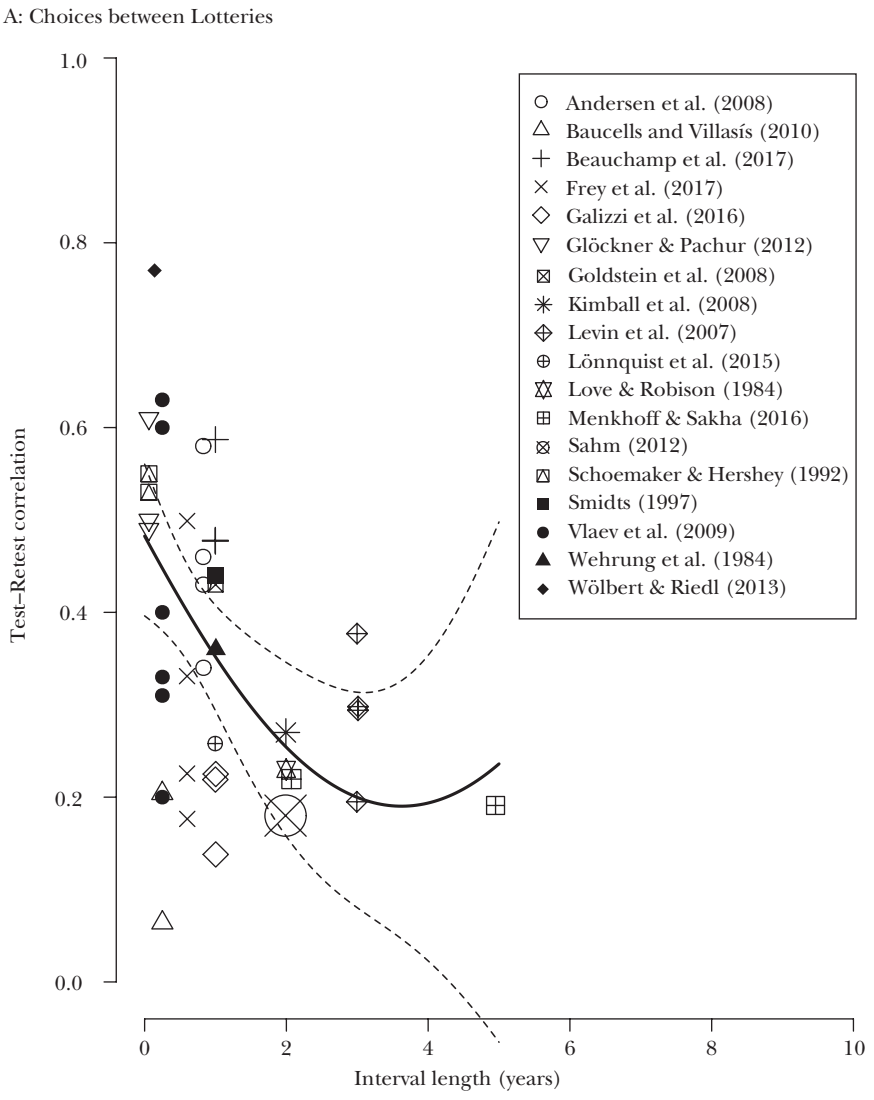
Both behavioral and self-report measures of risk preference have been subject to criticism. For example, some have voiced concern about the lack of generalizability across behavioral elicitation methods (Friedman et al. 2014). There is also a fair amount of skepticism in both economics (Beshears et al. 2008) and psychology (Haeffel and Howard 2010) about self-reports representing little more than "cheap talk." In our view, the relative strengths and weaknesses of the two measurement approaches as well as their possible links should be determined empirically. Unfortunately, and echoing Cronbach's (1957) diagnosis of psychologists' firm commitment to either one or the other methodology and associated theoretical constructs, the behavioral and the self-report approaches to measuring risk preference usually exist side-by-side with little or no empirical or theoretical integration.

Next, we turn to some of the work produced by the two approaches to studying risk preference, emphasize the strengths and weaknesses of both, and address implications for a more general theory of risk preference. In particular, we provide some evidence that self-report measures represent stable indicators of risk preference whereas widely used behavioral measures do not—and possibly as a consequence, there is often little agreement between the two. This realization has at least one important implication for psychologists and economists studying risk preference: It suggests that measures of risk preference cannot be used interchangeably when predicting outcomes of interest.

Temporal Stability

Do revealed (behavioral) and stated (self-report) risk preferences show similar levels of temporal stability? For an admittedly preliminary answer, we took advantage of narrative reviews of past work (Chuang and Schechter 2015) and drew on our knowledge of the literature to identify published findings and datasets that allowed us to compute test–retest reliability of revealed and stated risk preferences. Specifically, we identified studies reporting test–retest reliability of choices between

Figure 1
Meta-analysis of Test–Retest Stability of Risk Preferences

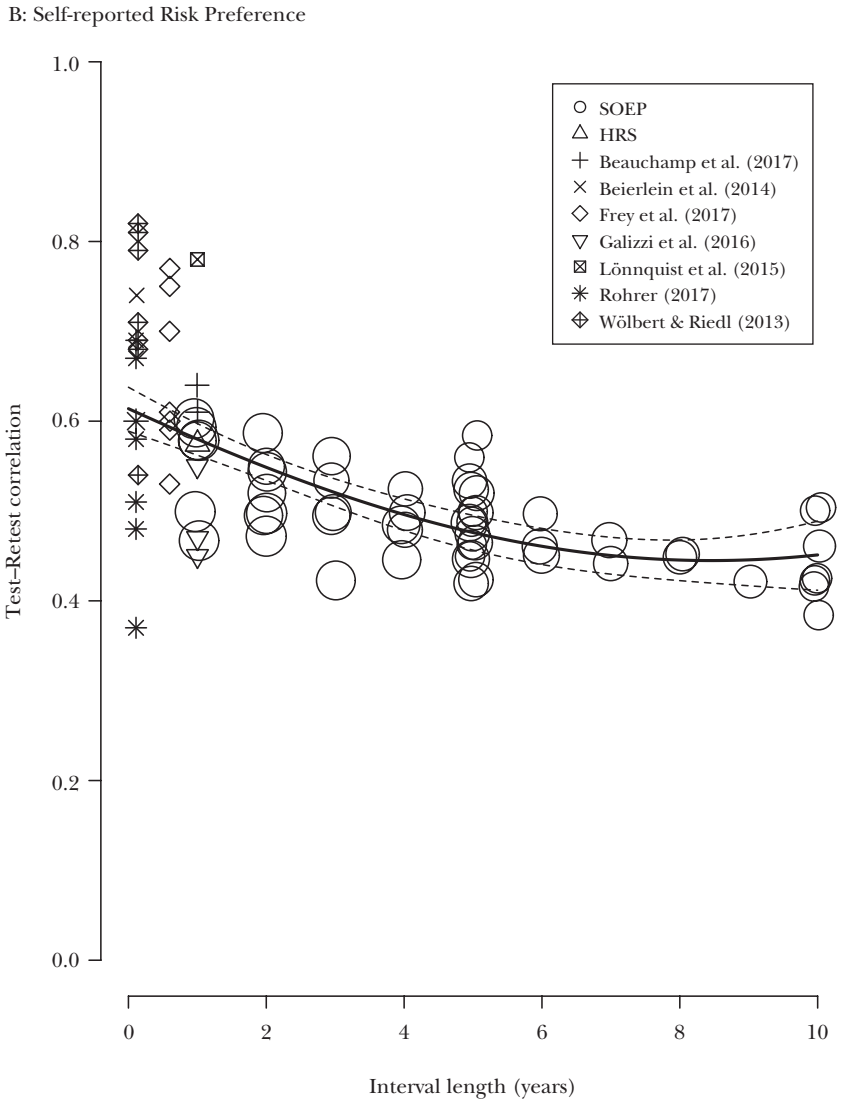


(Continued on next page)

monetary lotteries (for example, Harrison and Ruttström 2008), as well as studies and datasets reporting test–retest reliability of self-report items, with those items probing the propensity to take risks either in general or in specific domains of life, such as financial, health, and social domains (for example, Dohmen et al. 2011).

Figure 1 depicts the meta-analytic scatterplots of test–retest correlations for choices between lotteries (Figure 1A) and for self-reported risk preference (Figure 1B). The test–retest correlations help assess the extent to which the same

Figure 1 (Continued)
Meta-analysis of Test–Retest Stability of Risk Preferences



Note: Figure 1 presents the meta-analytic scatterplots of test–retest correlations for choices between lotteries (Panel A) and self-reported risk preference (Panel B). Symbols represent correlations between two measurement occasions obtained from published literature (references provided in the figure legends) and our own calculations (German Socio-Economic Panel, SOEP; American Health and Retirement Survey, HRS). Note that a small amount of jitter was added to each point to better distinguish points at the same interval length. The size of each point is proportional to the inverse variance (larger symbols = more precision). The solid line represents the weighted regression line including a linear and a quadratic term for interval length from a random effects meta-analysis (dashed lines correspond to 95 percent confidence intervals). We conducted the analyses using the package metafor for R (Viechtbauer 2010). Data and code are provided online with the article at the journal website, at https://www.aeaweb.org/journals/jep.

rank-ordering of individuals is preserved across two waves. To our knowledge, no data are available about the temporal stability for choices between lotteries with retest intervals longer than five years. Data on temporal stability of up to 10 years are available for self-report measures, albeit stemming mostly from one source, the German Socio-Economic Panel (Wagner, Frick, and Schupp 2007). Our analysis suggests that after five years, the measures taken from choices between lotteries show test–retest correlations of around .2 (although there is considerable uncertainty around that estimate). In contrast, the corresponding correlations for self-report are around .5 and these values do not seem to decline much across a 10-year period. Indeed, the level of stability found for self-report measures of risk preference is only slightly below the 10-year stability estimates for major personality traits, such as the Big Five, which are estimated at about .6, and shows greater stability than measures of life satisfaction, self-esteem, and affect, which are estimated to range between .35 and .4 for a 10-year period (Anusic and Schimmack 2016).

A potential criticism of the meta-analysis for measures of revealed preference is that it relies on choices between lotteries, and such choices may be perceived as artificial, therefore failing to engage participants. However, we have examined test–retest reliability of other prominent behavioral risk preference measures, including measures designed to be more engaging, such as the Balloon Analogue Risk Task (Lejuez et al. 2002) or the Columbia Card Task (Figner, Mackinlay, Wilkening, and Weber 2009). Such measures show low levels of test–retest reliability similar to those found using choices between lotteries across a delay of six months (Frey et al. 2017).

Does high temporal stability of risk preferences for individuals, at least for stated preferences, mean that there are no systematic changes within individuals over shorter or longer time scales? No. Research on personality suggests that high temporal stability in differences between individuals across long intervals is compatible with population mean-level changes in psychological traits (Roberts and DelVecchio 2000). Stability and change are compatible because mean-level changes—say, changes across the lifespan—represent average patterns affecting many or all individuals in the population, whereas test–retest reliability captures preservation of the relative *rank-ordering* of individuals, regardless of mean-level differences. This point may be easier to appreciate with an example. Intelligence is one of the most stable constructs known to psychology because of evidence of preserved rank-order stability (within a cohort) across decades (Deary 2001). However, intelligence can show dramatic and systematic changes as a function of momentary shocks, such as sleep deprivation (Lim and Dinges 2010) as well as long-term changes across the life span, including considerable decline in fluid components, such as reasoning and memory, with aging (Baltes, Staudinger, and Lindenberger 1999; Lindenberger 2014). Consistent with the concurrent presence of stability and change, we and others have found high test–retest reliability (Josef et al. 2016) *as well as* systematic mean-level reductions in risk-taking propensity with age in longitudinal examinations of self-reported risk-taking propensity (Josef et al. 2016; Dohmen et al. 2017).

One outstanding issue concerning individual and age-related changes in stated preferences is the extent to which they are indicative of "real" changes as opposed to mere changes in individuals' use of reference points across time. A similar issue has been raised in the domain of subjective well-being, where some have argued that age differences in self-reported measures may represent different benchmarks or reference classes (Weimann, Knabe, and Schöb 2015). Presently, we cannot offer a satisfactory response to this possible objection. Ideally, one would tackle this issue by using measures that are robust to this criticism, such as self-report measures that provide a relatively stable referential context (for example, "how risk taking are you relative to those of your age?") or, of course, behavioral measures in which reference points can be firmly and transparently established and systematically varied.

To summarize, risk preference measured from stated preferences emerges as a construct with considerable temporal stability, although revealed preference measures do not show such stability. Moderate rank-order stability in stated risk preferences is accompanied by sizable mean-level differences across the life span as well as significant variation within individuals. Consequently, the evidence suggests that present and future theories of risk preference need to account for both stable differences between individuals as well as systematic variation within individuals.

Convergent Validity

A key question in psychological research on risk preference has been whether it can be thought of as a domain-general tendency (similar to a general factor of intelligence, *g*, affecting behaviors implicating intelligence across many diverse contexts), or whether it should be construed as a multidimensional or domain-specific construct, with specific tendencies regarding wealth, health, or social exchange, to name just a few (for example, Slovic 1964; Weber, Blais, and Betz 2002). One way to approach this question empirically is to ask whether different measures of risk preference such as behavioral and self-report measures speak with one voice and converge in what they suggest about the individual.

Several studies on issues unrelated to risk have found that differences in experimental design can make a very large difference in behavioral patterns: for example, Berg, Dickhaut, and McCabe (2005) found large variations in behavioral patterns in laboratory experiments using different economic institutions, and Hertwig and Erev (2009) have found systematic differences and even preference reversals depending on whether risk information was described or experienced through repeated sampling. Further, the reported correlations between measures of risk preference are typically low (Dohmen et al. 2011; Galizzi, Machado, and Miniaci 2016). Such results cast doubt on the convergent validity of established risk preference measures. In what follows, we detail our recent efforts to assess the convergent validity of risk preference measures. We find a serious gap between different methods of eliciting risk preferences; in particular, we find a divide between stated (self-report) and

revealed (behavioral) preference measures, as well as among different behavioral measures.

First, in a study with 1,507 participants who completed a comprehensive battery of 39 risk preference measures—including a range of stated and revealed preference measures—we found that correlations between measures from the revealed and stated preference traditions were weak ($r = 0.06$; Frey et al. 2017). Moreover, the correlations among the nine different behavioral measures were substantially weaker ($r = 0.08$) than those among the 29 self-report measures ($r = 0.20$), even though the latter intentionally capture risk preference in diverse domains, such as financial, health, recreational, and social. The correlations between behavioral measures were not increased when parameters from specific decision models, such as expected utility theory or cumulative prospect theory, were used to describe individual's choices (Pedroni, Frey, Bruhin, Dutilh, Hertwig, and Rieskamp 2017). We also conducted a psychometric analysis using a bifactor model that directly accounts for shared variance across all measures with a single factor, leaving any residual variance to be captured by yet other specific, orthogonal factors. The bifactor analysis suggested that a general risk preference factor accounts for over 60 percent of the explained variance across measures, with the remaining variance captured by more domain-specific factors. Crucially, though this general factor explained substantial variance across self-report measures, it did not generalize to the behavioral measures. Overall, our psychometric analysis suggests that there is a large shared component that can be thought of as a general factor of risk preference bridging different domains of life that is captured from self-report (albeit not behavioral) measures. The idea of a general risk preference is in line with the robust observation that major psychological traits account for large portions of variance in subjective reports or behavior (Deary 2001; Caspi et al. 2014).

Second, we recently conducted a study on the gap between risk preference measures and its implications for understanding individual, sex, and age differences in risk preference, using the Innovation Sample of the German Socio-Economic Panel (Richter and Schupp 2015). In this study, we used different elicitation methods to survey a relatively large, age-heterogeneous, representative sample of the population, which ensures considerable variance in the outcomes of interest. Specifically, 951 individuals between 18 and 80 years of age were asked to complete different measures, including self-report measures of risk-taking propensity as well as incentivized behavioral measures of risk taking, involving decisions based on either described or experienced risk (Frey, Richter, Schupp, Hertwig, and Mata 2018). We were thus able to analyze the convergent validity of the three different measure types. Our findings are similar to past work on the description–experience gap, which suggests a gap in choice behavior between the measures involving the same lottery choices but presented in description mode or in experienced mode (Hertwig and Erev 2009; Wulff et al. 2018). Furthermore, we observed a gap between behavioral and self-report measures in their intercorrelations and their covariates. More precisely, the self-report, but not the behavioral measures, show the common patterns of sex and age differences identified in previous work, whereby males show higher

levels of risk-taking propensity relative to females, and younger adults show higher levels of risk-taking propensity relative to older adults (Josef et al. 2016; Mata, Josef, and Hertwig 2016). These data suggest not only a separation between self-reported and revealed preference measurements, but also systematic differences in how they relate to some demographic covariates.

Third, we have conducted several other studies that show that different behavioral measures also do not coalesce in providing a similar picture of age differences, which is potentially a result of the differential cognitive demands they impose (Mata et al. 2011; Frey, Mata, and Hertwig 2015; Mamerow et al. 2016). In a meta-analysis, we found that those behavioral measures of risk preference that involve considerable learning and memory demands are more likely to indicate large age differences in risk preferences (Mata et al. 2011). Specifically, whether older adults tend be more risk-seeking relative to younger adults, or vice versa, is likely to depend on the architecture of the choice task. For instance, older adults appear as if they seek more risk, relative to younger adults, whenever learning is necessary to overcome a task-specific anchor to choose a seemingly attractive but ultimately disadvantageous risky option. These results suggest one cause for the gap *between* revealed and stated preferences and even *within* revealed preferences. Revealed preferences are derived from measures that enlist processes that are also subject to cognitive or learning abilities and thus to inter- and intra-individual (during a life span) variations on those processes (for additional discussion of the role of cognitive abilities see the article by Dohmen, Falk, Huffman, and Sunde in this symposium).

To summarize, at present, there appears to be little hope for establishing a clear link between self-report and behavioral measures of risk preference, not only because measures from the two traditions do not correlate with each other, but also because revealed preference measures, that is, behavioral measures, fail to converge. Nevertheless, extant work suggests that stated preferences partly derive from a general risk preference component that accounts for a large portion of variance across life domains. As discussed in the next section, whether stated or revealed preference measures provide a better account of individuals' propensity for risk should be judged in light of prospective studies involving predictive validity of real-world behavior.

Predictive Validity

Real-world financial institutions such as banks and insurance companies have shown little use for revealed risk preference measures when recommending their products—perhaps because of the surprisingly limited predictive validity of utility and risk constructs obtained from revealed preference measures for real-world choices (Friedman et al. 2014). Unfortunately, there are few studies in the literature involving the measurement of risk preference to predict *objective* measures of real-world outcomes. Those few studies suggest, first, that self-reports and informant reports, assessing risk preference or related constructs, do have considerable

predictive validity for real-world outcomes such as teenage pregnancy, drug use, or financial security, even when controlling for other factors such as intelligence or socioeconomic status (Moffitt et al. 2011; Caspi et al. 2016; Beauchamp, Cesarini, and Johannesson 2017). Second, self- and informant reports are potentially more powerful than behavioral measures in this regard (White, Moffitt, Avshalom, Jeglum, Needles, and Stouthamer-Loeber 1994). In addition, interventions that have targeted specific at-risk groups identified through self-report measures of related constructs show promising results (Conrod et al. 2013), whereas the complementary evidence for the power of behavioral measures is still lacking.

To summarize, the current scant evidence suggests no advantage of revealed (behavioral) over stated preference measures in predicting real-world outcomes. While there are some promising results concerning the predictive validity of stated risk preference, data concerning the predictive validity of behavioral measures and comparisons between self-report and behavioral approaches are sorely needed. Clearly, more prospective longitudinal designs are required for both measurement paradigms. Such studies are difficult, time-consuming, and expensive to conduct. Unfortunately, long-standing panels, such as the German Socio-Economic Panel (SOEP), are not, at this time, equipped with psychometrically sound behavioral measures (for example, measures with satisfactory test–retest stability, batteries exhibiting convergent validity), nor objectively measured criterion variables (for example, credit reports, drug tests from biological samples) to permit fast progress in this regard. However, there is some work that links risk preference data from existing surveys to administrative data, such as education or income (for example, Beauchamp et al. 2017), and we hope and expect that more will follow.

A Look Ahead

Risk preference, when measured through stated, self-reported preferences, displays trait-like characteristics, such as high temporal stability across years and high convergent validity between different measurement instruments spanning different life domains. Furthermore, stated preferences seem to show significant predictive validity for a number of economic and health outcomes, dispelling the notion of self-assessments as simply "cheap talk." However, the picture emerging from studies using revealed (behavioral) preference measures is less promising, with problems of poorer temporal stability, confounds related to high demands on learning, memory, or numeracy skills, and a relative lack of evidence concerning their predictive validity.

Many important phenomena in research on risk preference are still insufficiently understood. What explains the lack of convergence between stated and revealed preference measures? Why do revealed preference measures display so little convergent validity among themselves? What is the relative predictive validity of stated relative to revealed preference measures? In light of the fundamental nature of such questions, we hope that psychologists and economists team up to

conduct the necessary research to address them. We should emphasize that the debate on bridging the divide between different measures, such as the self-report and behavioral measures of risk preference, is not unlike that taking place in economics concerning the link between subjective and objective measures of well-being (Deaton and Stone 2013). To harvest the potential of these constructs and their value for actual policy-making, they need to still be better operationalized, measured, and understood.

Looking ahead, we identify two main avenues for future work on the study of risk preference. First, we hope to have helped to convince researchers interested in risk preference to undertake the painstaking task of examining the temporal stability, convergent validity, and predictive validity of their favorite measures. A time-honored tradition, such as relying on choices between monetary gambles, cannot substitute for this foundational work. Our own goal for future work is to develop and study a toolbox of measures to assess their strengths: perhaps some measures may be better at gauging a trait-like and domain-general component of risk preference, whereas others may be better suited to gauge domain-specific components. For example, it is possible that some behavioral measures may be better in simulating the specific incentive structure and choice architecture of a real-world context for which behavior is to be predicted. One interesting avenue toward a toolbox and taxonomy of risk preference measures is theory-driven task construction and decomposition using computational or neural methods that can disentangle risk preference from cognitive demands or other individual characteristics (Wallsten, Pleskac, and Lejuez 2005; Helfinstein et al. 2014). However, we suspect that computational and neural methods offer no panacea for the lack of temporal stability and convergent validity of the currently available behavioral measures (Frey et al. 2017; Pedroni et al. 2017).

Second, we need to make conceptual progress by addressing the psychological primitives or traits underlying individual differences in the appetite for risk. There is some agreement in the psychological literature about the existence of a few major psychological traits, such as a general factor of intelligence, *g*, and a few basic dimensions of personality (as one example, extraversion). However, there are still ongoing debates about distinctions within such constructs. In intelligence research, some lines of research focus on a general factor (Deary 2001) whereas others investigate specific facets such as the distinction between crystallized versus fluid intelligence (Baltes et al. 2007). Similarly, in the field of personality there are ongoing debates about whether to distinguish one, two, five, or yet more dimensions of personality (for example, Block 2010). The place for risk preference in this uncertain "periodic table" of psychological elements is yet unclear. Psychology has a tradition of introducing new constructs without full concern for their conceptual or empirical distinction. In this context, risk preference, sensation-seeking, impulsivity, self-control, grit, will-power, self-regulation, or conscientiousness are only some of the monikers that psychologists have introduced to explain individual differences in the appetite for risky activities, for example drug use, crime, and financial investment (Cross, Copping, and Campbell 2011; Roberts, Lejuez, Krueger, Richards, and

Hill 2014; Sharma, Markon, and Clark 2014). In line with the notion that psychological traits are general, we suspect that such labels characterize largely the same trait, and our empirical work suggests considerable overlap between such constructs (Frey et al. 2017). In practice, empirical studies investigating the temporal stability, and convergent and predictive validity of such different constructs and their respective measures will be fundamental in making conceptual progress. Psychology is already moving in that direction by initiating studies directly aimed at uncovering the amount of variance shared by measures originally proposed in the context of different traits (Eisenberg et al. 2018; Frey et al. 2017). The results of this work promise to be of immediate practical relevance to all behavioral scientists interested in determining how many and what kind of risk preference measures to include in their studies and models.

To conclude, risk preference, at least when measured through stated (self-reported) preferences, may be thought of as a moderately stable, general psychological trait, and, thus, an important variable to consider in psychological and economic theories and policy-making. Nevertheless, the measurement of risk preference needs more attention, and the usefulness of behavioral measures to uncover a stable psychological risk preference trait seems, at this time, surprisingly limited. Future research on risk preference needs to develop and deploy both stated (self-report) and revealed (behavioral) risk preference measures in prospective longitudinal studies in order to uncover their convergent and differential predictive validity for important economic and other life outcomes. In the meantime, behavioral scientists should be aware that not all measures of risk preference are created equal.

■ *We are grateful to the editors Gordon Henson, Enrico Moretti, and Timothy Taylor for their feedback; to Ulrike Malmendier for helpful discussions on the topic; to Anne Albrecht for help with coding studies for the meta-analyses; and to Laura Wiles and Ann Norman for editing the manuscript. This work was partly funded by a grant to the first author from the Swiss National Science Foundation (156172; http://p3.snf.ch/project-156172).*

References

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. "Lost in State Space: Are Preferences Stable?" *International Economic Review.* 49(3): 1091-112.

Anusic, Ivana, and Ulrich Schimmack. 2016. "Stability and Change of Personality Traits, Self-Esteem, and Well-Being: Introducing the Meta-Analytic Stability and Change Model of

Retest Correlations." *Journal of Personality and Social Psychology* 110(5): 766–81.

Appelt, Kirstin C., Kerry F. Milch, Michel J. J. Handgraaf, and Elke U. Weber. 2011. "The Decision Making Individual Differences Inventory and Guidelines for the Study of Individual Differences in Judgment and Decision-Making Research." *Judgment and Decision Making* 6(3): 252–62.

Baltes, Paul B., Ursula M. Staudinger, and Ulman Lindenberger. 1999. "Lifespan Psychology: Theory and Application to Intellectual Functioning." *Annual Review of Psychology* 50: 471–507.

Baucells, Manel, and Antonio Villasís. 2010. "Stability of Risk Preferences and the Reflection Effect of Prospect Theory." *Theory and Decision* 68(1–2): 193–211.

Beauchamp, Jonathan P., David Cesarini, and Magnus Johannesson. 2017. "The Psychometric and Empirical Properties of Measures of Risk Preferences." *Journal of Risk and Uncertainty* 54(3): 203–37.

Beierlein, Constanze, Anastassiya Kovaleva, Christoph J. Kemper, and Beatrice Rammstedt. 2014. "Eine Single-Item-Skala zur Erfassung von Risikobereitschaft: Die Kurzskala Risikobereitschaft-1 (R-1)" [A Single Item Scale for the Measurement of Proneness to Risk: The Short Scale Risk Proneness-1 (R-1)]. (GESIS Working Papers 2014|34) Cologne, Germany: GESIS.

Benjamin, Daniel J., David Cesarini, Matthijs J. H. M. van der Loos, Christopher T. Dawes, Philipp D. Koellinger, Patrik K. E. Magnusson, Christopher F. Chabris, Dalton Conley, David Laibson, Magnus Johannesson, and Peter M. Visscher. 2012. "The Genetic Architecture of Economic and Political Preferences." *PNAS* 109(21): 8026–31.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 2005. "Risk Preference Instability across Institutions: A Dilemma." *PNAS* 102(11): 4209–14.

Bernoulli, Daniel. 1738 [1954]. "Exposition of a New Theory on the Measurement of Risk." *Econometrica* 22(1): 23–36.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "How Are Preferences Revealed?" *Journal of Public Economics* 92(8–9): 1787–94.

Blais, Ann-Renée, and Elke U. Weber. 2006. "A Domain-Specific Risk-Taking (DOSPERT) Scale for Adult Populations." *Judgment and Decision Making* 1(1): 33–47.

Block, Jack. 2010. "The Five-Factor Framing of Personality and Beyond: Some Ruminations." *Psychological Inquiry* 21(1): 2–25.

Browne, Mark J., Verena Jaeger, Andreas Richter, and Petra Steinorth. 2016. "Family Transitions and Risk Attitude." MRIC Working Paper 32, Munich Risk and Insurance Center.

Caspi, Avshalom, Renate M. Houts, Daniel W. Belsky, Sidra J. Goldman-Mellor, HonaLee Harrington, Salomon Isreal, Madeline H. Meier, Sandhya Ramrakha, Idan Shalev, Richie Poulton, Terrie E. Moffitt. 2014. "The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders?" *Clinical Psychological Science* 2(2): 119–37.

Caspi, Avshalom, Renate M. Houts, Daniel W. Belsky, Hona Lee Harrington, Sean Hogan, Sandhya Ramrakha, Richie Poulton, and Terrie E. Moffitt. 2016. "Childhood Forecasting of a Small Segment of the Population with Large Economic Burden." *Nature: Human Behaviour* 1, Article 0005.

Caspi, Avshalom, Brent W. Roberts, and Rebecca L. Shiner. 2005. "Personality Development: Stability and Change." *Annual Review of Psychology* 56: 453–84.

Castellanos-Ryan, Natalie, et al. 2016. "The Structure of Psychopathology in Adolescence and Its Common Personality and Cognitive Correlates." *Journal of Abnormal Psychology* 125(8): 1039–52.

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. "Experimental Methods: Eliciting Risk Preferences." *Journal of Economic Behavior and Organization* 87: 43–51.

Chuang, Yating, and Laura Schechter. 2015. "Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results." *Journal of Development Economics* 117: 151–70.

Clark, William A. V., and William Lisowski. 2017. "Prospect Theory and the Decision to Move or Stay." *PNAS* 114(36): E7432–E77440.

Conrod, Patricia J., Maeve O'Leary-Barrett, Nicola C. Newton, Lauren Topper, Natalie Castellanos-Ryan, Clare Mackie, and Alain Girard. 2013. "Effectiveness of a Selective, Personality-Targeted Prevention Program for Adolescent Alcohol Use and Misuse: A Cluster Randomized Controlled Trial." *JAMA Psychiatry* 70(3): 334–9.

Cronbach, Lee J. 1957. "The Two Disciplines of Scientific Psychology." *American Psychologist* 12(11): 671–84.

Cross, Catharine P., Lee T. Copping, and Anne Campbell. 2011. "Sex Differences in Impulsivity: A Meta-Analysis." *Psychological Bulletin* 137(1): 97–101.

Deary, Ian J. 2001. "Human Intelligence Differences: A Recent History." *Trends in Cognitive Sciences* 5(3): 127–30.

Deaton, Angus, and Arthur A. Stone. 2013. "Two Happiness Puzzles." *American Economic Review* 103(3): 591–7.

Dohmen, Thomas, Armin Falk, Bart H. H. Golsteyn, David Huffman, and Uwe Sunde. 2017. "Risk Attitudes Across the Life Course." *Economic*

Journal 127(605): F95–F116.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522–50.

Edwards, Ward. 1954. "The Theory of Decision Making." *Psychological Bulletin* 51(4): 380–417.

Eisenberg, Ian W., et al. 2018. "Applying Novel Technologies and Methods to Inform the Ontology of Self-Regulation." *Behaviour Research and Therapy*, 101: 46-57.

Figner, Bernd, Rachael J. Mackinlay, Friedrich Wilkening, and Elke U. Weber. 2009. "Affective and Deliberative Processes in Risky Choice: Age Differences in Risk Taking in the Columbia Card Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(3): 709–30.

Fisher, Gwenith, Mike Gideon, Joanne W. Hsu, and Brooke Helppie McFall. 2011. "Cognitive Economics Study: Development and Methodology." November. Cognitive Economics Project, University of Michigan, Ann Arbor, USA. http://ebp-projects.isr.umich.edu/CogEcon/docs/dev+meth.pdf.

Frey, Renato, Rui Mata, and Ralph Hertwig. 2015. "The Role of Cognitive Abilities in Decisions from Experience: Age Differences Emerge as a Function of Choice Set Size." *Cognition* 142: 60–80.

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3(10): e1701381.

Frey, Renato, David Richter, Jürgen Schupp, Ralph Hertwig, and Rui Mata. 2018. "Identifying Correlates of Risk Preference: A Systematic Approach Using Specification Curve Analysis" Unpublished paper.

Friedman, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder. 2014. *Risky Curves: On the Empirical Failure of Expected Utility.* New York: Routledge.

Galizzi, Matteo M., Sara R. Machado, and Raffaele Miniaci. 2016. "Temporal Stability, Cross-Validity, and External Validity of Risk Preferences Measures: Experimental Evidence from a UK Representative Sample." Available at SSRN: http://dx.doi.org/10.2139/ssrn.2822613.

Glimcher, Paul W., and Ernst Fehr. 2014. *Neuroeconomics: Decision Making and the Brain*, 2nd edition. San Diego, USA: Academic Press, Elsevier.

Glöckner, Andreas, and Thorsten Pachur. 2012. "Cognitive Models of Risky Choice: Parameter Stability and Predictive Accuracy of Prospect Theory." *Cognition* 123(1): 21–32.

Goldstein, Daniel G., Eric J. Johnson, and William F. Sharpe. 2008. "Choosing Outcomes versus Choosing Products: Consumer-Focused Retirement Investment Advice." *Journal of Consumer Research* 35(3): 440–56.

Haeffel, Gerald J., and George S. Howard. 2010. "Self-Report: Psychology's Four-Letter Word." *American Journal of Psychology* 123(2): 181–88.

Harrison, Glenn W., and E. Elisabet Rutström. 2008. "Risk Aversion in the Laboratory." *Research in Experimental Economics*, Vol. 12: Risk *Aversion in Experiments,* edited by James C. Cox and Glenn W. Harrison, pp. 41–196. Bingley, UK: Emerald.

Helfinstein, Sara M., Tom Schonberg, Eliza Congdon, Katherine H. Karlsgodt, Jeanette A. Mumford, Fred W. Sabb, Tyrone D. Cannon, Edythe D. London, Robert M. Bilder, and Russell A. Poldrack. 2014. "Predicting Risky Choices from Brain Activity Patterns." *PNAS* 111(7): 2470–75.

Hertwig, Ralph, and Ido Erev. 2009. "The Description–Experience Gap in Risky Choice." *Trends in Cognitive Sciences* 13(12): 517–23.

Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24(3): 383–403.

Highhouse, Scott, Christopher D. Nye, Don C. Zhang, and Thaddeus B. Rada. 2017. "Structure of the DOSPERT: Is There Evidence for a General Risk Factor?" *Journal of Behavioral Decision Making* 30(2): 400–06.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–55.

Josef, Anika K., David Richter, Gregory R. Samanez-Larkin, Gert G. Wagner, Ralph Hertwig, and Rui Mata. 2016. "Stability and Change in Risk-Taking Propensity across the Adult Life Span." *Journal of Personality and Social Psychology* 111(3): 430–50.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory—Analysis of Decision under Risk." *Econometrica* 47(2): 263–92.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2008. "Imputing Risk Tolerance from Survey Responses." *Journal of the American Statistical Association* 103(483): 1028–38.

Knutson, Brian, and Scott A. Huettel. 2015. "The Risk Matrix." *Current Opinion in Behavioral Sciences* 5(October): 141–46.

Lejuez, Carl W., Jennifer P. Read, Christopher W. Kahler, Jerry B. Richards, Susan E. Ramsey, Gregory L. Stuart, David R. Strong, and Richard A. Brown. 2002. "Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task

(BART)." *Journal of Experimental Psychology: Applied* 8(2): 75–84.

Levin, Irwin P., Stephanie S. Hart, Joshua A. Weller, and Lyndsay A. Harshman. 2007. "Stability of Choices in a Risky Decision-Making Task: A 3-Year Longitudinal Study with Children and Adults." *Journal of Behavioral Decision Making* 20(3): 241–52.

Lim, Julian, and David F. Dinges. 2010. "A Meta-Analysis of the Impact of Short-Term Sleep Deprivation on Cognitive Variables." *Psychological Bulletin* 136(3): 375–89.

Lindenberger, Ulman. 2014. "Human Cognitive Aging: Corriger la fortune?" *Science* 346(6209): 572–78.

Lönnqvist, Jan-Erik, Markku Verkasalo, Gari Walkowitz, and Philipp C. Wichardt. 2015. "Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison." *Journal of Economic Behavior and Organization* 119: 254–66.

Love, Ross O., and Lindon J. Robison. 1984. "An Empirical Analysis of the Intertemporal Stability of Risk Preference." *Southern Journal of Agricultural Economics* 16(1): 159–66.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics* 126(1): 373–416.

Mamerow, Loreen, Renato Frey, and Rui Mata. 2016. "Risk Taking across the Life Span: A Comparison of Self-Report and Behavioral Measures of Risk Taking." *Psychology and Aging* 31(7): 711–23.

Mata, Rui, Anika K. Josef, and Ralph Hertwig. 2016. "Propensity for Risk Taking across the Life Span and around the Globe." *Psychological Science* 27(2): 231–43.

Mata, Rui, Anika K. Josef, Gregory R. Samanez-Larkin, and Ralph Hertwig. 2011. "Age Differences in Risky Choice: A Meta-Analysis." *Annals of the New York Academy of Sciences* 1235(1): 18–29.

Menkhoff, Lukas, and Sahra Sakha. 2016. "Determinants of Risk Aversion over Time: Experimental Evidence from Rural Thailand." Available at SSRN: http://doi.org/10.2139/ssrn.2785467.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W. Roberts, Stephen Ross, Malcolm R. Sears, W. Murray Thomson, and Avshalom Caspi. 2011. "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety." *PNAS* 108(7): 2693–8.

Pedroni, Andreas, Renato Frey, Adrian Bruhin, Gilles Dutilh, Ralph Hertwig, and Jörg Rieskamp. 2017. "The Risk Elicitation Puzzle." *Nature: Human Behaviour* 1: 803–9.

Richardson, Michelle, Charles Abraham, and Rod Bond. 2012. "Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis." *Psychological Bulletin* 138(2): 353–87.

Richter, David, and Jürgen Schupp. 2015. "The SOEP Innovation Sample (SOEP IS)." *Schmollers Jahrbuch* 135(3): 389–400.

Roberts, Brent W., and Wendy F. DelVecchio. 2000. "The Rank-Order Consistency of Personality Traits from Childhood to Old Age: A Quantitative Review of Longitudinal Studies." *Psychological Bulletin* 126(1): 3–25.

Roberts, Brent W., Carl Lejuez, Robert F. Krueger, Jessica M. Richards, and Patrick L. Hill. 2014. "What Is Conscientiousness and How Can It Be Assessed?" *Developmental Psychology* 50(5): 1315–30.

Rohrer, Julia M. 2017. "Test–Retest Reliabilities of Scales Included in the Socio-Economic Panel Study." OSF, Open Science Framework, February 7, https://osf.io/3ncbt.

Sahm, Claudia R. 2012. "How Much Does Risk Tolerance Change?" *Quarterly Journal of Finance* 2(4): 1250020.

Schmidt, Frank L., and John Hunter. 2004. "General Mental Ability in the World of Work: Occupational Attainment and Job Performance." *Journal of Personality and Social Psychology* 86(1): 162–73.

Schoemaker, Paul J. H., and John C. Hershey. 1992. "Utility Measurement: Signal, Noise, and Bias." *Organizational Behavior and Human Decision Processes* 52(3): 397–424.

Schonberg, Tom, Craig R. Fox, and Russell A. Poldrack. 2011. "Mind the Gap: Bridging Economic and Naturalistic Risk-Taking with Cognitive Neuroscience." *Trends in Cognitive Sciences* 15(1): 11–19.

Sharma, Leigh, Kristian E. Markon, and Lee Anna Clark. 2014. "Toward a Theory of Distinct Types of 'Impulsive' Behaviors: A Meta-Analysis of Self-Report and Behavioral Measures." *Psychological Bulletin* 140(2): 374–408.

Slovic, Paul. 1964. "Assessment of Risk Taking Behavior." *Psychological Bulletin* 61(3): 220–33.

Smidts, Ale. 1997. "The Relationship Between Risk Attitude and Strength of Preference: A Test of Intrinsic Risk Attitude." *Management Science* 43(3): 357–70.

Steinberg, Laurence. 2013. "The Influence of Neuroscience on US Supreme Court Decisions about Adolescents' Criminal Culpability." *Nature Reviews: Neuroscience* 14(7): 513–8.

Tracy, Jessica L., Richard W. Robins, and Jeffrey W. Sherman. 2009. "The Practice of Psychological Science: Searching for Cronbach's Two Streams in Social–Personality Psychology." *Journal of*

Personality and Social Psychology 96(6): 1206–25.

Viechtbauer, Wolfgang. 2010. "Conducting Meta-Analyses in R with the Metafor Package." *Journal of Statistical Software* 36(3): 1–48.

Vlaev, Ivo, Nick Chater, and Neil Stewart. 2009. "Dimensionality of Risk Perception: Factors Affecting Consumer Understanding and Evaluation of Financial Risk." *Journal of Behavioral Finance* 10(3): 158–81.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP)—Scope, Evolution and Enhancements." *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 127(1): 139–69

Wallsten, Thomas S., Timothy J. Pleskac, and Carl W. Lejuez. 2005. "Modeling Behavior in a Clinically Diagnostic Sequential Risk-Taking Task." *Psychological Review* 112(4): 862–80.

Weber, Elke U., Ann-Renée Blais, and Nancy E. Betz. 2002. "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors." *Journal of Behavioral Decision Making* 15(4): 263–90.

Wehrung, Donald A., Kenneth R. MacCrimmon, and Kent M. Brothers. 1984. "Utility Assessment: Domains, Stability, and Equivalence Procedures." *INFOR: Information Systems and Operational Research* 22(2): 98–115.

Weimann, Joachim, Andreas Knabe, and Ronnie Schöb. 2015. *Measuring Happiness: The Economics of Well-Being*. Cambridge, MA: MIT Press.

White, Jennifer L., Terrie E. Moffitt, Caspi Avshalom, Dawn Jeglum, Douglas J. Needles, Magda Stouthamer-Loeber. 1994. "Measuring Impulsivity and Examining its Relationship to Delinquency." *Journal of Abnormal Psychology* 103(2): 192–205.

Wölbert, Eva, and Arno Riedl. 2013. "Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity." Netspar Discussion Paper no. 07/2013-044. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2354907.

Wulff, Dirk U., Max Mergenthaler-Canseco, and Ralph Hertwig. 2017. "A Meta-Analytic Review of Two Modes of Learning and the Description–Experience Gap." *Psychological Bulletin*, 144(2): 140-76

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 173–192

Space, the Final Economic Frontier

Matthew Weinzierl

The Soviet Union launched its Sputnik satellite in 1957. A year later, the National Advisory Committee for Aeronautics, a little-known agency that had played a limited role in pursuing basic research in aeronautics since 1915, was transformed into the National Aeronautics and Space Administration. The surge of US government spending on human spaceflight through the Apollo program in the 1960s cemented a public-sector centralized model of the US space sector, putting NASA at its hub for the next 50 years. NASA set the strategy for exploration and use of space, and it also coordinated the market's structure, which largely involved government purchases from prominent aerospace firms. As NASA historian Joan Lisa Bromberg (1999) wrote of those early years: "[NASA Administrator James L.] Webb believed that national space policy should not be turned over to private firms. It was government acting in the public interest that had to determine what should be done, when it should be done, and for how much money."

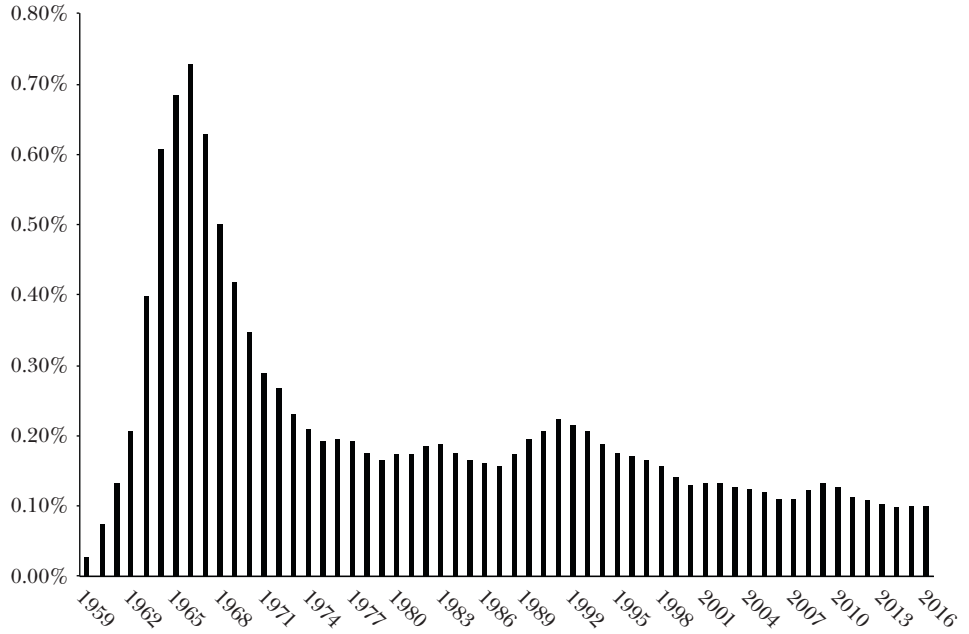
After decades of centralized control of economic activity in space, NASA and US policymakers have begun to cede the direction of human activities in space to commercial companies. Figure 1 shows that NASA garnered more than 0.7 percent of GDP in the mid-1960s, but that level fell precipitously in the late 1960s and then gradually but persistently over the next 40 years to around 0.1 percent of GDP today. Meanwhile, space has become big business, with $300 billion in annual revenue. Recent valuations of innovative space firms like SpaceX ($21 billion), Orbital ATK ($7.8 billion), and dozens of small startups (receiving $2.8 billion in funding

■ *Matthew Weinzierl is Professor of Business Administration, Harvard Business School, Boston, Massachusetts. His email address is mweinzierl@hbs.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.173

doi=10.1257/jep.32.2.173

Figure 1
NASA Budget as a Share of GDP



in 2016) suggest the market is optimistic about what's next. Recent high-profile successes, most recently the launch and return of SpaceX's Falcon Heavy rocket, are generating a new surge of public interest and enthusiasm.

The shift from public to private priorities in space is especially significant because a widely shared goal among commercial space's leaders is the achievement of a large-scale, largely self-sufficient, developed space economy. Jeff Bezos, whose fortune from Amazon has funded the innovative space startup Blue Origin, has long stated that the mission of his firm is "millions of people living and working in space." Elon Musk (2017), who founded SpaceX, has laid out plans to build a city of a million people on Mars within the next century. Both Neil deGrasse Tyson and Peter Diamandis have been given credit for stating that Earth's first trillionaire will be an asteroid-miner (as reported in Kaufman 2015). Such visions are clearly not going to become reality in the near future. But detailed roadmaps to them are being produced (National Space Society 2012), and recent progress in the required technologies has been dramatic (Metzger, Muscatello, Meuller, and Mantovani 2013). If such space-economy visions are even partially realized, the implications for society—and economists—will be enormous. After all, it will be our best chance in human history to create and study economic societies from a (nearly) blank slate. Though economists should treat the prospect of a developed space economy with healthy skepticism, it would be irresponsible to treat it as science fiction.

In this article, I provide an analytical framework—based on classic economic analysis of the role of government in market economies—for understanding and managing the development of the space economy. That framework has three components: 1) establishing the market through decentralization of decision making and financing for human space activities; 2) refining the market through policies that address market failures and ensure a healthy market structure; and 3) tempering the market through regulation in pursuit of social objectives. The next three sections will focus on these issues. Some of the topics are familiar from Earth, while others are unique to space, but most of these questions—despite the pioneering work of space-focused economists such as Macauley and Toman (1991, 2004, 2005), Hertzfeld (1992, 2007), and MacDonald (2014, 2017)—remain largely unaddressed. I will focus on the US space sector, but the framework applies equally well to the efforts of any spacefaring nation.

Establishing Markets in Space: Decentralization

The Slow Decline of Centralization

Since the start of the Space Age, private-sector leaders have been issuing warnings that a centralized model would undermine progress on public and, especially, commercial priorities in space. For example, Ralph Cordiner (1961), the one-time chairman and CEO of General Electric, foresaw much of the development of the government-directed space sector over the subsequent several decades while forcefully arguing that, eventually, space's "development shall be under our traditional competitive enterprise system."

The economic logic for the centralized model was clear, and for several decades it achieved its (remarkable) goals. Public goods such as national security, national pride, and basic science are typically underprovided if left to the market, and NASA was founded to provide them during the Cold War. Its command-and-control structure grew naturally from that objective, as the merits of decentralization took a back seat to the imperative of directed action. Under this model, the United States has been the leading space power and NASA has occupied the technological frontier. Most prominently, the success of the Apollo missions (including the 1969 moon landing) inspired grand visions for what would come next. In the early 1970s, studies of space colonization and diversified space-based economies proliferated, even at the highest levels of the space program (O'Neill 1976).

But after the last of the Apollo missions in 1972, NASA—and thus the US space sector—struggled to find a second act. Part of the reason was that the tight connection between the Apollo program and competition with the Soviet Union made NASA's budget vulnerable to the sense that the mission had already been accomplished (Logsdon 2015). Apollo astronaut Buzz Aldrin said: "After the Apollo lunar missions, America lost its love of space—there was no concentrated follow-up and we didn't have any clear objectives" (as quoted in Sunyer 2014).

When NASA decided that its next emphasis would be on the Space Transportation System, better known as the Shuttle, it applied largely the same centralized approach it had used in the 1960s, but with more mixed results. The first flight of the Columbia space shuttle was in 1981. Successive shuttle flights enabled two decades of achievements by NASA, including the construction of the International Space Station (ISS) and Hubble Space Telescope, and they demonstrated American technological prowess. But the Shuttle's costs were higher than hoped (roughly two-thirds of NASA's human spaceflight budget and around $220 billion in 2017 dollars) and its performance weaker (it missed more half of its planned annual flights). Moreover, public goods were prioritized over commercial priorities, handicapping the growth of the commercial space sector. Logsdon (2011), a prominent space expert, has written: "[I]t was probably a mistake to develop this particular space shuttle design, and then to build the future U.S. space program around it."

After two tragic accidents, with the Challenger shuttle in 1986 and the Columbia shuttle in 2003, momentum turned away from the Shuttle and the centralized model of space it represented. The President's Commission on Implementation of United States Space Exploration Policy (2004) came to a striking conclusion: "NASA's role must be limited to only those areas where there is irrefutable demonstration that only government can perform the proposed activity." The shuttle program was cancelled in 2011, leaving the United States in the embarrassing position of not being able to launch humans from domestic soil.

The vulnerabilities of centralized control will be familiar to any economist: weak incentives for the efficient allocation of resources, poor aggregation of dispersed information, and resistance to innovation due to reduced competition. In addition to these concerns, NASA's funding and priorities were subject to frequent, at times dramatic, revision by policymakers, making it hard for the space sector to achieve even the objectives set at the center (Handberg 1995; Logsdon 2011).

Anticipating these vulnerabilities, reform advocates had made previous pushes for at least partial decentralization and a greater role for the private sector in space. Near the dawn of the Shuttle era, President Ronald Reagan signed the Commercial Space Launch Act of 1984, saying: "One of the important objectives of my administration has been, and will continue to be, the encouragement of the private sector in commercial space endeavors." That same year saw the creation of the Office of Commercial Programs at NASA and the Office of Commercial Space Transportation in the Department of Transportation (NASA 2014). However, these early seeds would have to wait until the end of the Shuttle program to bear fruit.

An instructive contrast is provided by the approach the US government took to the development of the commercial satellite market. In 1962, Congress created COMSAT, a for-profit, private corporation owned by common shareholders and a group of telecommunications companies (though three of the company's 15 board seats were to be appointed by the US President). NASA was officially charged with providing technical advice to COMSAT, and the agency was given responsibility for COMSAT's launches. The idea behind this public–private partnership was to leverage

the expertise of NASA to jump-start a private communications satellite industry. It was "industrial policy with a vengeance" in the words of NASA historian Bromberg (1999), and it led to the rapid deployment and use—for both public and private purposes—of the vast array of satellites that dominate the space economy today.

The Rise of New Space

When the shuttle program itself ended in 2011, commercialization-minded reformers in both the public and private space sectors seized their opportunity. In the words of Bretton Alexander, an executive at Blue Origin and former White House space official: "The failure of NASA to find a replacement for the shuttle for 30 years shattered the idea of NASA being in charge … When the shuttle was retired, it created this void that allowed NASA to look to the commercial sector" (quoted in Weinzierl and Acocella 2016).

The decentralized set of space companies that emerged is generally known as "New Space." Table 1 offers a (necessarily incomplete) overview of some of the main companies currently active in commercialization of space. The "space access" companies focus on launching people and payload into space. The "remote sensing" companies provide images of Earth and are closely related to the "satellite data and analytics" companies, which also serve a range of other customers. The "habitats and space stations" companies plan to provide secure facilities for manufacturing, research, and even tourism in so-called "low Earth orbit" (the space between 160 km and 2,000 km of altitude). The "beyond low Earth orbit" companies have goals ranging from space manufacturing to asteroid mining to colonization of the Moon and Mars. Not listed in the table are research and investment firms, whose increased involvement in space suggests a maturing of the sector as a wider range of investors seek information and access. Leading examples of these include Bryce Space and Technology and an array of investment firms ranging from those focused on space (for example, Space Angels) to those devoting a small share of their large resources to space (for example, Bessemer and Draper Fisher Jurvetson).

Funding for New Space companies comes from a variety of sources. A set of high-profile entrepreneurs—Elon Musk, Jeff Bezos, Richard Branson, Paul Allen, and others—have used their wealth to overcome high fixed-cost barriers to entry, launching companies based on new approaches to the technology and management of space access. According to leading space industry analyst Bryce Space and Technology (2017), outside investment in start-up New Space firms has risen from less than $500 million per year from 2001 to 2008 to roughly $2.5 billion per year in 2015 and 2016.[1]

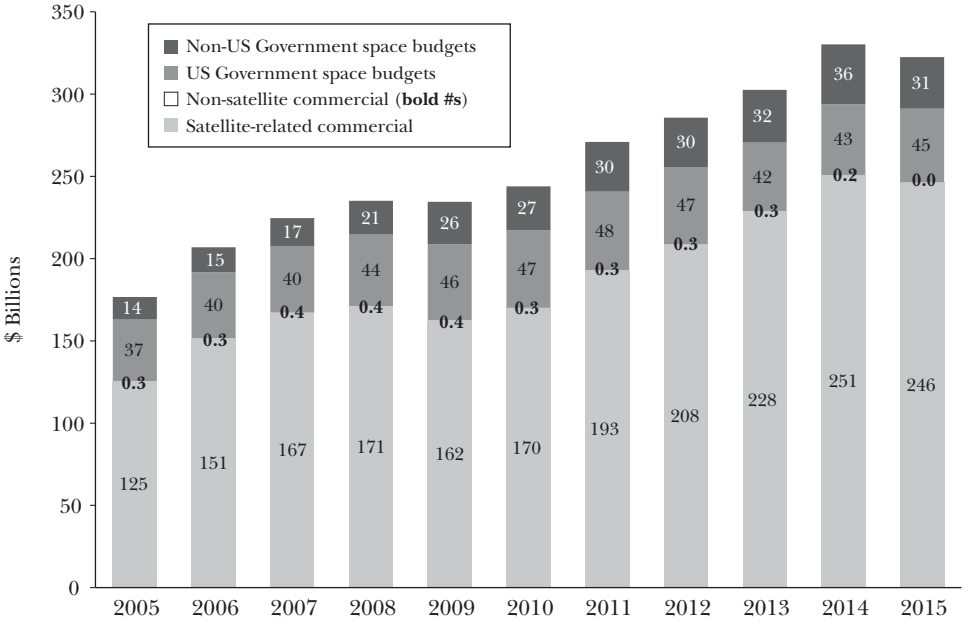
[1] In 2006, levels were higher, as there were large debt offerings (by the satellite provider Protostar and broadband provider WildBlue—now ViaSat). Investment flows grew to roughly $2 billion per year from 2009 to 2011, thanks mainly to interest from private equity firms and substantial debt offerings by Ligado Networks (broadband), Digital Globe (Earth imaging—recently merged into Maxar), and O3b (a satellite constellation provider). The years 2013 and 2014 saw some large acquisitions in this sector, including Monsanto acquiring the Climate Corporation ($930 million), Google acquiring TerraBella ($478 million, later sold to Planet), and SES acquiring O3b ($730 million). Levels in 2015 and 2016 included inflows of venture capital that were larger than $1.5 billion each year (Bryce Space and Technology 2017).

Table 1
A Sample of Companies Involved in Commercial Space Activities

Sector	*Company (alphabetical by sector)*	*Year founded*	*Full-time equivalent workers (2016)*[a]	*Products/Services*
Space access	Astrobotic	2008	11–50[b]	Transportation to the Moon
	Blue Origin	2000	875	Launch vehicles and engines, space tourism
	Boeing Aerospace	1978	2,800	Crewed LEO transportation
	Masten Space Systems	2004	11–50[b]	Suborbital launches of small payloads
	Orbital ATK	1982	12,700	Orbital launches of satellites and ISS cargo
	Sierra Nevada Corp.	1963	3,094	Cargo and crewed LEO transportation
	Space Adventures	1998	17	Crewed LEO, lunar transport, and tourism
	SpaceX	2002	5,420	Reusable launch vehicles, colonization
	Stratolaunch Systems	2011	501–1000[b]	Air-launched orbital launch services
	World View Enterprises	2012	11–50[b]	High-altitude private spaceflight balloons
	United Launch Alliance	2006	4,000	Orbital launch services
	Virgin Galactic	2004	200	Space tourism; rapid commercial flight
	XCOR Aerospace	1999	23	Suborbital launches, human spaceflight
Remote sensing	Iceye	2012	11–50[b]	Synthetic aperture radar remote sensing
	Planet (including Terra Bella)	2010	251–500[b]	Earth imaging and video, data provision
	Spire Global Inc.	2006	101–250[b]	Data gathering; Earth observation network
Satellite data access and analytics	Analytical Space	2016	10	Optical LEO comms network, full service
	Astroscale	2013	11–50	Space Debris Removal
	Bridgesat	2015	3	Optimal comms network, hardware
	Kepler Communications	2015	5	Internet communications to crafts in orbit
	Maxar	n/a	5,000+	Diversified: satellites, imaging, robotics
	OneWeb	2012	101–250[b]	Large-scale satellite constellation
	Oxford Space Systems	2013	11–50[b]	Deployable satellite structures
	Qwaltec	2001	58	Satellite and network operations
	Skywatch	2014	11–50[b]	Satellite data integration Earth observation
	Vector Space Systems	2016	11–50[b]	Micro satellite space vehicle
Habitats and space stations	Axiom	2015	11–50[b]	Commercial space station building off ISS
	Bigelow Aerospace	1999	135	Inflatable space habitats
	Ixion Initiative Team	2016	n/a	Commercial use of rocket upper stages
	Made In Space	2010	50	Additive manufacturing in space
	Nanoracks	2009	40	Payload transport, deployment hardware
	Space Tango	2014	5–10	Microgravity research platforms
Beyond low Earth orbit	Deep Space Industries	2012	11–50[b]	Asteroid mining
	Golden Spike	2010	11–50[b]	Human lunar expeditions
	Mars One	2011	11–50[b]	Mars colonization
	Moon Express	2010	51–100[b]	Moon exploration and mining
	Planetary Resources, Inc.	2010	11–50[b]	Asteroid mining

Source: List and descriptions of companies compiled from the Commercial Spaceflight Federation website and author research.
Note: LEO is "low Earth orbit." ISS is the International Space Station.
[a] Employee data is from private communications with companies or Capital IQ, US Department of Labor, unless otherwise noted:
[b] Data from Crunchbase;
[c] Capital IQ, third-party data.

Figure 2
Space Sector Revenue



Source: The Space Report (Space Foundation 2018).
Note: Classification adjusted by the authors to separate satellite-related from other commercial revenue. Non-U.S. governments include (in descending order of amount of revenue) ESA, China, Russia, Japan, France, along with several others (which recorded less than $1 billion in 2015).

Figure 2 shows estimates from *The Space Report* (Space Foundation 2018) that revenues in the space sector have climbed from less than $200 billion in 2005 to more than $300 billion in recent years, with the vast majority of that activity related to satellite technology for telecommunications and other services. The rest is the space budgets of governments—US and others—and commercial revenues from nonsatellite space services). This dominance of the satellite business in space revenue is likely to hold for the foreseeable future, especially given projections of substantial growth in small satellite constellations for Earth observation, where published forecasts (Henry 2016) see revenue of $22 billion over the next decade.

Credible estimates of the ultimate economic potential of space in the long term are elusive, as many of its most ambitious plans have very uncertain prospects. As one example, a 2014 report by the Boston Consulting Group put global spending on luxury travel at $460 billion and the overall luxury "experiences" market at $1.8 trillion (Abtan et al. 2014). Some New Space companies such as Blue Origin are working to claim a slice of this vast market for space, but there is substantial skepticism toward space tourism among many in the industry. Revenues from space manufacturing or asteroid-mining will be negligible in the near term and perhaps also in the medium term, though active commercial research toward both is being

funded in the marketplace. In the end, whether lower-cost access and infrastructure for working in space will generate an economic reason to be in space—as current investors hope and expect—remains unclear.

At this point, the terminology of "New Space" has come to represent not just a new generation of companies (after all, well-established firms like Boeing and Orbital Sciences are also important players) or a steady growth in space-sector revenues, but rather a new approach. In the centralized model, private firms working with NASA were largely insured against the enormous risks of investments in space through cost-plus contracts, but they had little ability to participate in the potential gains from a commercialized space market. In the "New Space" approach, private firms share in the enormous risks and (potential) returns of investments in space (Achenbach 2013; see also Weinzierl and Acocella 2016).

A Channel for Decentralization: Commercial Orbital Transportation Services

As the Shuttle program wound down, the primary channel by which NASA and the rising New Space sector came together to solve the space access problem—and thereby provide an example of how decentralization can work—was a set of public–private partnerships called Commercial Orbital Transportation Services (COTS). In 2005, Congress funded COTS with $500 million (less than 1 percent) of NASA's five-year budget, with the goal of "challenging private industry to establish capabilities and services that can open new space markets while meeting the logistics transportation needs of the International Space Station" (NASA 2014). As Lambright (2016) writes in a history of the program, "[NASA Administrator Michael Griffin's] vision was to build a new commercial space industry." In particular, it was hoped that COTS would lower cargo—and eventually crew—transportation costs and thus help to open up a set of untapped opportunities in low Earth orbit.

The key innovation in the Commercial Orbital Transportation Services program was to make NASA a customer and partner, not a supervisor, of its private contractors. In particular, COTS contracts replaced conventional cost-plus procurement for customized products with fixed-price payments for the generic capabilities of delivering and disposing or returning cargo and transporting crew to low Earth orbit (in other settings, COTS is an acronym for "commercial off the shelf").[2] This change shifted risk from NASA to private firms, reducing the need for NASA to use a combination of intensive monitoring and cost-plus contracts to control costs and encourage innovation.

New Space companies welcomed the new approach: their investors were comfortable taking on risk; innovation and efficiency were (they argued) their key advantages over established players; and they found intensive monitoring to be costly and invasive. Firms were given the freedom and responsibility to design and produce their products as they saw best, with NASA providing insight rather than

[2] More specifically, COTS agreements were structured using so-called Other Transaction Authority under the rubric of Space Act Agreements, replacing Federal Acquisition Regulation (FAR) rules that had governed the vast majority of NASA contracts prior to COTS.

oversight. Moreover, firms would retain the ownership of the intellectual property created for the COTS, whereas under previous contracts, the government was the default holder of intellectual property because the work was done at its behest, not for the broad marketplace (NASA 2014).

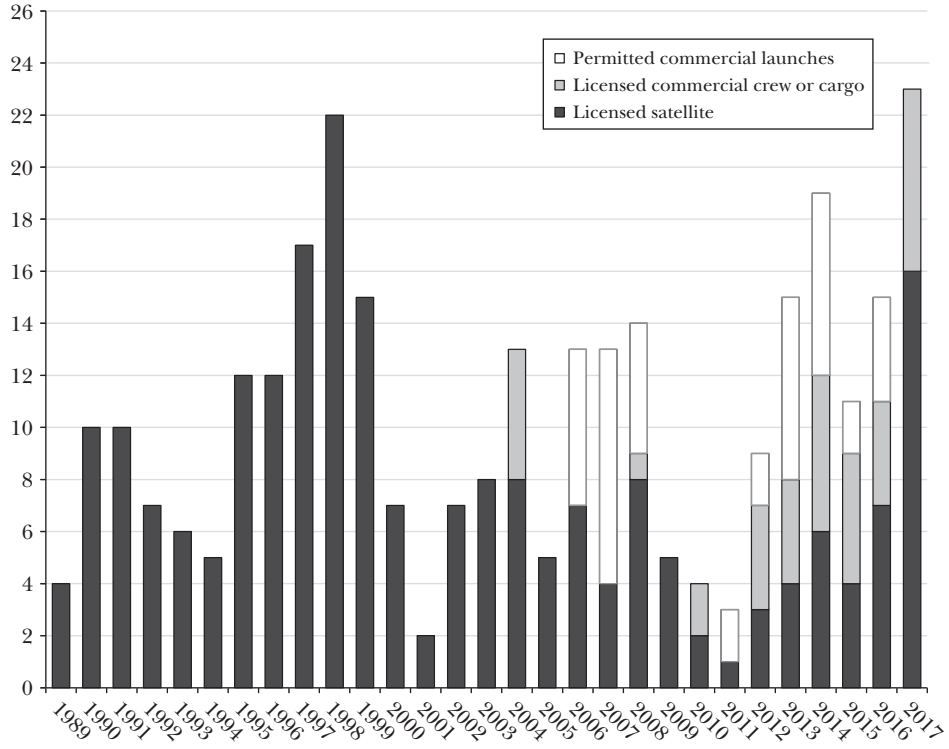
The Commercial Orbital Transportation Services program offered several advantages for NASA. First, the agency could leverage private capital to acquire its required services more cheaply: NASA (2014) reported that COTS provided "U.S.-based cargo transportation services at a significantly lower cost than previous Space Shuttle flights." In particular, NASA (Zapata 2017) provided a detailed breakdown of the cost savings from COTS, concluding that the all-in cost to deliver a kilogram of cargo to the International Space Station was approximately $89,000 through SpaceX and $135,000 through Orbital Sciences, one-third and one-half the $272,000 estimated cost per kilogram that would have been possible with the Space Shuttle. Second, and related, COTS would allow NASA to redirect its time and budget to projects like basic science and exploratory research. As NASA Administrator Charlie Bolden noted: "These agreements are significant milestones in NASA's plans to take advantage of American ingenuity to get to low Earth orbit, so we can concentrate our resources on deep space exploration" (as cited in Morring 2011; see also NASA 2014; Launius 2014).

Despite its appeal, the Commercial Orbital Transportation Services program was initially viewed by some within the established space sector as, at best, a backup plan for the more conventional approach. NASA already had in place a multifaceted exploration and space access program called Constellation, and part of that program (Ares 1/Orion) was focused on low Earth orbit. But the Constellation program ran over budget and behind schedule. When it was eventually cancelled by President Obama, COTS became far more than a backup.

In fact, the Commercial Orbital Transportation Services program has been making core contributions to achieving NASA's missions. By 2008, two companies had convinced the agency of their ability to provide full resupply services to the International Space Station, and NASA awarded fixed-price contracts for 20 flights valued at $3.5 billion to SpaceX and Orbital Sciences under a successor program, Commercial Resupply Services (CRS). These flights are now a main way in which the space station is resupplied. Even the program's missteps were seen as making progress: when NASA cancelled one of the initial contracts after the partner company, Rocketplane Kistler, failed to meet benchmarks, the agency proved that it took its role as a "customer" seriously (Lambright 2016). The successes of the cargo programs led to the Commercial Crew Development program, a multiphase project that has culminated in scheduled crew transportation to the space station by SpaceX and Boeing before 2020. In just over a decade, the relationship between the US space program and commercial providers had shifted, in NASA's (2014) words, "From Contingency to Dependency."

Moreover, these public–private partnership programs spurred activity and innovation within the space sector that presage a broadening of the space economy. To take one particularly important example, they fed a new surge of

Figure 3
FAA-Licensed and Permitted Commercial Launches by Objective



Source: Federal Aviation Administration (FAA) 2018.
Note: This figure displays the number of commercial launches that were officially licensed by the FAA (for satellite delivery or for missions related to resupplying the International Space Station with crew or cargo) or that were permitted by the FAA (permits for experimental launches can be granted in less time and with fewer requirements than a full license, pursuant to the 2004 Commercial Space Launch Amendments Act).

private nonsatellite-related commercial launch activity, as shown in Figure 3, that included a drive toward "reusability"—that is, the capacity to employ components of launch vehicles and spacecraft multiple times. Many in the space sector have expressed sentiments in agreement with SpaceX's Elon Musk, who has said: "If one can figure out how to effectively reuse rockets just like airplanes, the cost of access to space will be reduced by as much as a factor of a hundred. A fully reusable vehicle has never been done before. That really is the fundamental breakthrough needed to revolutionize access to space" (as quoted in SpaceX 2015). SpaceX's successful demonstrations of reusability for its launch vehicle (in 2016), its cargo capsule (in 2017), and most recently its heavy-launch vehicle (in 2018) were therefore seen as watershed moments in both aerospace technology and the commercialization of space. Musk has made clear the importance to his company's success

of its participation in the public–private partnerships: "SpaceX could not do this without NASA. Can't express enough appreciation," he tweeted in February 2017.

The Broader Commercialization of Low Earth Orbit

In March 2017, the US space sector took a further step toward decentralization with the signing of the NASA Transition and Authorization Act, a comprehensive and bipartisan reauthorization bill. In essence, policymakers decided to go beyond asking commercial providers to carry out what would previously have been NASA missions, such as carrying people and payload to the International Space Station, and to cede the direction of activities in low Earth orbit to commercial space providers. If this transition succeeds, NASA will adopt a more targeted role focused on space exploration and basic science, the public goods that have long been its core competencies, leaving the economic development of space largely to the private sector. Historians such as Launius (2014) suggest there is a historical analogue to this relationship in the commercial aviation industry, where the US government played a critical role in basic research in the mid 20th century while leaving the operation of the aviation sector in private hands.

Despite the success of public–private partnerships in resupplying missions to the International Space Station, commercialization comes with risks, and the case for broader commercialization in low Earth orbit is hotly debated. Critics often argue that New Space companies are piggybacking in various ways: for example, off NASA technology that took decades to develop, and through marginal-cost pricing for the use of NASA facilities (NASA 2014) and indemnification from risk. A related critique is that public–private partnerships channeling resources away from established space contractors risk undermining the institutional knowledge and economies of scale that have been built up over decades. Finally, it is unclear whether NASA will stay hands-off as the scope of commercial space activities expands both in low Earth orbit and beyond (for discussion, see Martin 2011). In fact, current debates over the path to Mars provide a clear example of these tensions, and their resolution will tell us a great deal about the future of the space sector.

Clearly, a number of questions remain to be addressed on the way to a decentralized space economy. Will the public–private partnership approach be an effective model for encouraging further commercialization, or would a clearer separation of public and private sectors be more effective? How should the industrial structure of commercial space be influenced by the public sector, including NASA? Will decentralization of economic activity in space focused on private goods undermine or bolster support for NASA and the public goods it produces?

Refining the Market: Addressing Market Failures

The original justifications for NASA included its ability to provide public goods like basic science, national pride (Logsdon 2004; Launius 2006a), and support of

national security (although NASA is a civilian agency). In other words, NASA was a response to classic market failures. As the evolving economics of space push toward a greater role for market forces, risks of other market failure arise. Two examples are already complicating the sector's development: the problem of complementarities and coordination (which in turn is related to a risk of insufficient competition), and the problem of externalities like those caused by space debris.

Complementarities and Coordination

Many New Space companies have business models that make sense only when other, complementary models are already in place. Consider some technologies widely believed to be essential for the commercialization of space: low-cost, frequent launch capabilities; in-space manufacturing; scalable habitats; in-space resource extraction and energy collection; and reliable radiation shielding and debris mitigation. Individually, each of these technologies has only a limited payoff. Low-cost launches are still expensive if there is nothing to do and nowhere to go in space. Building habitats for manufacturing or tourism is of no use if they cannot be secured from the dangers of space. And so on. If these technologies were realized together, however, they would form a self-sustaining system with potentially enormous profit potential. In the economics of human space activities, the whole may be much greater than the sum of the parts.

One can imagine a self-reinforcing virtuous cycle of development that would support the space economy. For example, cheaper and more frequent rocket launches might facilitate short-term tourism, along with industrial and scientific experimentation on suborbital and orbiting spacecraft. If these activities become routine, demand might rise for commercial habitats to support longer flights. In turn, these habitats could generate demand for resources in space, increasing the opportunities for workers and residents.

But one can also reasonably doubt that such an ideal path will be realized easily or without some nudges along the way. Limits on or asymmetries of information, the high level of risk inherent in space, and the challenges of capturing surplus from such complementarities will make it difficult to move forward on the most efficient path—or even to move forward at all.[3]

Even if the market "succeeds" in capturing these complementarities, the economics of the sector suggest that the outcome would feature a high degree of concentration. After all, complementarities mean large profits for actors that integrate the pieces of the whole, and entrepreneurs at the forefront of New Space (Jeff Bezos, Elon Musk, Richard Branson, and others) are masters of such a strategy on Earth. Economies of scale and scope have, in fact, always characterized commercial

[3] Consider, for instance, a classic stag hunt game in which an inferior but less-risky equilibrium is selected rather than the more efficient coordinated equilibrium. In this game, two individuals go hunting. Each must choose whether to hunt for a high-value stag or low-value hare. However, choosing a hare is guaranteed to succeed, while choosing a stag only succeeds if the other person also chooses "stag." See Brynjolfsson and Milgrom (2013) for a relevant review of complementarities in economics.

space: NASA historian Bromberg (1999) points out that one of the agency's earliest goals was to retain competition among its contractors and avoid monopolization.

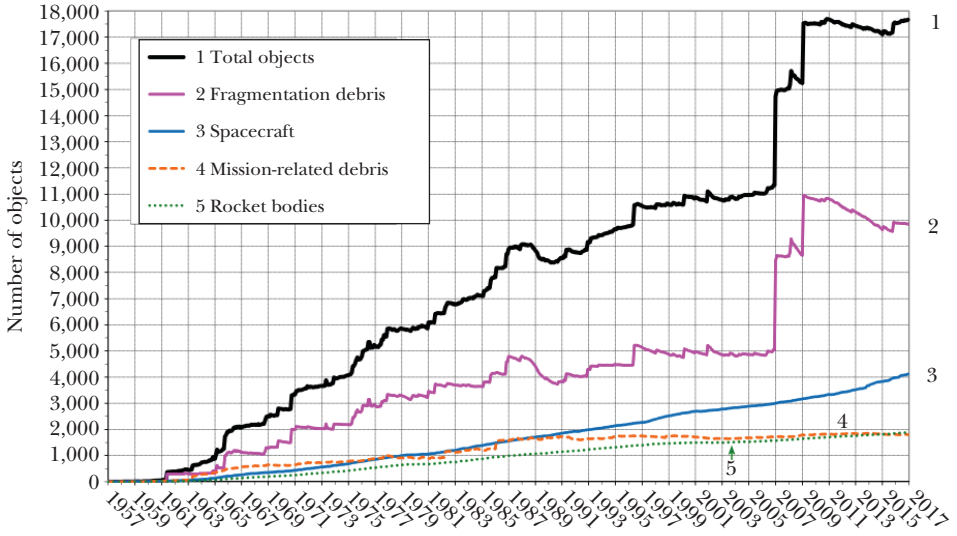
Carefully designed public-sector coordination can help: indeed, Hertzfeld (1992) made similar arguments at an earlier stage of the US space sector's development, when commercialization was far less advanced. NASA's recent efforts at coordinating the commercialization of space have scored some successes.

For example, Commercial Orbital Transportation Services and related programs not only subsidized commercial launch vehicles, they also maintained a competitive market structure through a diversified set of award contracts. The Commercial Crew Development program awarded contracts to six companies in its first round, four companies (plus three more without funding) in its second round, three in its third round, and two in its final round (NASA 2014). NASA has tried to play a similar role in encouraging habitat technologies. Most prominently, Bigelow Aerospace has been allowed to dock its inflatable expandable activity module on the International Space Station to prepare for its use in modular commercial stations. But NASA has also actively partnered with five other companies to develop deep-space habitat technology through its NextSTEP and NextSTEP-2 public–private partnerships (for details, see https://www.nasa.gov/nextstep).

Historical analogies suggest lessons for how the public sector can play this facilitative role. Launius (2014) provides an in-depth analysis of six relevant historical episodes. The construction of the US transcontinental railroad in the late 19th century is commonly cited in the space community as an example of how government support—massive in that case—can facilitate development of a new frontier. (Donaldson and Hornbeck 2016 find that growth in the American West was moderately higher as a result.) The story of the railroads suggests the range of forms such support might take: direct transfers, lower taxes, guaranteed contracts, and even grants of property. The story of the railroads also reveals risks of such efforts, however, as early government support led to a concentration of economic (and political) power. The differences between space and such an analogy are instructive, as well. Unlike with the railroads and the West, rockets are the only means of accessing space and no national government has authority over property rights in space. Also, while the railroads linked communities of eager customers, demand for easy access to space is still nascent and will depend on the development of complementary technologies. Launius's other five case studies are a diverse group—fostering the aerospace industry; creating the telephone industry; supporting research in Antarctica; advancing public works; and making accessible conservation zones (scenic and cultural)—each of which provides additional lessons.

The complementarities at the heart of developing a commercial space sector raise a number of policy questions. What role should the government play in coordinating and subsidizing these interdependent technologies? Which forms of subsidy—cost-sharing, revenue guarantees, prizes—would be most effective? If the provision of these linchpin technologies turns out to have the features of natural monopoly, how should policymakers respond? How will the surplus from such an interdependent set of inputs be shared among its participants?

Figure 4
Space Debris
(monthly number of objects in Earth orbit)



Source: From NASA (2017) with only minor stylistic changes.

Crowding and the Space Debris

The development of space is already generating concerns about overuse and crowding in the most useful regions of low Earth orbit. In time, such concerns seem likely to spread to the richest asteroids and orbital space in general. In fact, Earth's orbital space is already being described as "congested, contested, and competitive" (Duff-Brown 2015). To illustrate this problem in more detail, consider the case of space debris.

Space debris—including defunct satellites, spacecraft parts, and the pieces created by collisions between them—is accumulating, as shown in Figure 4. Even small debris can inflict major damage: a piece of metal the size of a cherry carries the explosive power of a grenade when in orbit. Current estimates are that 23,000 objects larger than 10 centimeters in diameter, 500,000 particles between 1 and 10 centimeters, and over 100 million particles smaller than 1 centimeter are flying through low Earth orbit. Most of these objects have been created in just the past ten years, as shown in Figure 4, in part due to two major events. As explained in Weinzierl and Acocella (2016b), "On Feb 10, 2009, an active US communications satellite (Iridium 33) exploded on impact with a defunct Russian satellite (Kosmos 2251), spewing 2,200 trackable objects and hundreds of thousands of smaller, undetectable fragments into Earth's orbit. ... In 2007, a Chinese weather satellite (Fengyun-1C) was destroyed by a kinetic kill vehicle traveling at nearly 18,000 mph as part of China's anti-satellite ballistic missile test, creating over 2,000 pieces of

trackable objects—those larger than 10 centimeters in diameter—and an estimated 150,000 smaller fragments." While the current threats from debris are generally considered manageable through shielding and avoidance technology, the long-term problem is daunting, especially when considering the enormous increase in the size and number of orbiting objects required for a developed space economy. Warnings of an uncontrollable chain reaction of debris-generating collisions—in which debris creates collisions that lead to more debris—came as early as the 1970s from NASA scientist Donald Kessler, and the issue is only becoming more pressing with time.

The space debris problem is a classic example of negative externalities but in a setting in which the conventional remedies suggested by economic analysis and applied on Earth have limited traction. For example, Hanson (2016) suggests a standard Pigouvian price on debris, but also notes that a main obstacle is the lack of any space taxing authority. A Coasian (1960) solution in which affected parties negotiate to internalize externalities will be difficult in the case of space debris because this approach requires clearly delineated property rights, and no such rights exist in space. A polycentric governance solution as in Ostrom (2009), in which public and private actors would collectively manage orbital debris in a way similar to how a range of actors manage large-scale irrigation projects and water rights in some emerging economies, may be possible but faces an uphill battle. After all, the conditions under which Ostrom found this kind of cooperation most promising—including the ability to monitor and discipline actions—are missing in space (Weinzierl, Acocella, and Yamazaki 2016). In short, without some centralized action, space debris could generate an outcome similar to the tragedy of the commons.[4]

International agreements have made some progress on the issue of space debris by requiring that objects put into space in the future have automatic de-orbiting capabilities, but the main provision of international treaties relevant to debris—the assignment of responsibility for debris to the party or country from which it was first launched—has fallen far short. In fairness, identifying the origin of pieces of debris is difficult, assigning responsibility for an object having become debris (say, due to a collision with another object) is often impossible, and enforcing countries' obligations threatens their national security and economic interests in other assets. The analogy to global climate change, where a decades-long effort to generate international coordination has gradually confronted these obstacles, is both useful and daunting. A more encouraging analogy is to international efforts to reverse the depletion of the ozone layer, where over the several decades multiple rounds of agreements have turned the tide. Advocates of action on space debris often point to the need for public awareness of the problem, a factor often credited with encouraging swift action on the ozone layer.

[4]Some industry consortia have recently proposed self-regulation to address space debris (as reported in Foust 2017). Hertzfeld, Weeden, and Johnson (2016) suggest that these efforts will be more effective if they focus on how the debris problem differs from the textbook "tragedy of the commons" scenario.

With this challenging landscape, economists have the tools to pose and address some key questions. Are private interests, like those of satellite providers or space tourists, likely to create sufficient demand for debris removal and a more systematic stewardship of space? If not, what policies can governments adopt, or what markets can governments create, to price or regulate these externalities? How can these negative externalities be internalized without working against the subsidization merited by the positive externalities discussed above? Can unilateral actions succeed, or is cooperation across countries imperative? How can historical (or current) examples inform our answers to these questions?

Tempering the Market: Pursuing Social Objectives

Even an established, efficient space marketplace offers no guarantee that the pursuit of private priorities in space will serve the public or respect the public's ethical judgments. Some questions lie outside the natural scope of economists (for example, with regard to our moral responsibility to preserve outer space as we find it). But if we fail to exert oversight over the space economy, its legitimacy—and thus its success—will be undermined.

As a tangible example of the challenges in protecting the public interest without handicapping the private space economy, consider the case of asteroid mining. A number of private companies are interested in mining asteroids for precious metals, in-space manufacturing inputs, habitat materials, and (perhaps most likely) water. The technological challenges to asteroid mining are formidable, but the regulatory landscape is also a risk. The heart of the economic issue is who has the right to mine and profit from the resources to be found in asteroids. As Krolikowski and Elvis (2017) caution, if commercial interest in asteroids conflicts with the public's interest in them for scientific exploration or space settlement—for example, because mining destroys material of interest to scientists while extracting material that is useful to settlers—how are such conflicts to be sorted out?

Similar legal and ethical challenges apply to the management of two terrestrial frontiers: Antarctica and the oceans. In Antarctica, international treaties have kept development to a minimum, at least for the next several decades. As discussed by Ehrenfreund, Race, and Labdon (2013), the Antarctic Treaty System commits signatories to a range of limitations intended to leave undisturbed the Antarctic ecosystem, the most important of which are the prohibitions on military and mineral resource extraction activities. Scientific research and exploration, including tourism, are allowed but carefully managed by international bodies. Similar goals animate the treaties governing the management of the oceans—the UN Convention on the Law of the Seas—but centuries of military and commercial activities (and claims) complicate the picture. For example, the United States has not formally ratified the Convention and has, at times, expressed concern over its proposals on mining rights and fees applied to the international seabed beyond the defined economic zones of coastal countries. In the oceans, the tension between

economic and environmental priorities is therefore more apparent than in Antarctica, perhaps because there is more economically at stake.

Existing international space treaties neither endorse nor prohibit the private use of resources in space. The 1967 Outer Space Treaty, which continues to be the main framework for international cooperation, strikes an ambiguous middle ground on the development and use of resources in space. It encourages—but does not require—cooperation on responsible use. An attempt by some nations to put in place a more restrictive agreement, the 1979 Moon Treaty, has not been signed by any spacefaring nation. The resulting ambivalence over property rights in space has had no real effects for decades. But with the rise of commercial space, choosing a regulatory approach to property rights has taken on new urgency.

The United States upset the regulatory status quo—and facilitated the growth of asteroid mining companies—by passing the Commercial Space Launch Competitiveness Act in 2015, a law that grants property rights to the resources on a planetary body (though not to the body itself) to whoever "gets there first." The law's treatment of property rights reflects the principle that the first actor to utilize a resource earns the right, as the law says, "to possess, own, transport, use, and sell." The fundamental tradeoff rooted within this approach is that a property right granted in this way may be utilized in a way that conflicts with society's interests, but without that right the resource may be left undeveloped altogether. A resolution to this tradeoff offered by Locke (1689) and made famous by Robert Nozick (1974) is the so-called "Lockean proviso," in which appropriation of a resource is justifiable if each individual is left at least as well off as in a world where all resources had remained unowned. This justification was at the heart of supporters' case for the 2015 act.

While some other countries were critical of the bold creation of property rights in space by the Commercial Space Launch Competitiveness Act, arguing that space resources should be common property, others rushed to follow suit. For example, small but high-income Luxembourg has played a key role in commercial space as the headquarters of SES, a major satellite owner and operator. In the context of space resources, Luxembourg's key advantage is its regulatory responsiveness to firms. In fact, both of the leading asteroid mining companies—Planetary Resources and Deep Space Industries—have opened offices in Luxembourg and praised the country's business-friendly setting. In other words, Luxembourg is positioning itself to be for asteroid-mining companies what Delaware has been in recent decades for major American firms.

It appears that the right of private companies to mine and profit from asteroids is quickly being formalized. An open question is whether, if asteroid miners ever turn their visions into reality, these legal commitments will hold. The distributional questions arising from the development of space will be contentious. Complicating matters further, some of the greatest disparities in the returns from space may be across countries or generations—or even across on-Earth and off-Earth societies—rather than within traditional boundaries.

The uncoordinated structure of space regulation raises a number of questions that economists might help to pose and answer. As the space economy is developed,

how will the value it creates be shared among the countries, and people, on Earth and off, now and in future generations? Does competition across nations pose a risk of a race to the regulatory "bottom" in the context of asteroid mining? What is the first-best structure of property rights in space, and what is the (politically) constrained second-best option?

Concluding Thoughts

The successful economic development of space tests the limits of imagination. However, it might plausibly share some of the features of postwar American suburbanization. In each case, the locations from which emigration occurred (urban cores; Earth) were becoming polluted, crowded, and fractious. Innovations in transportation were making migration feasible for workers (mass transit and automobiles; low-cost launch). Innovations in residential technology were making housing workers in the new locations possible (mass-produced housing units; space habitats). Complementarities were leading a proliferation of supportive activities to develop (shopping malls and office parks; resource extraction and in-space manufacturing).

One can even imagine that "supraurban" societies in space would compete to attract settlers and workers, extending Tiebout (1956) competition—with its benefits and costs—in a new direction. For economists, the possibility of extraterrestrial experimentation with alternative institutional and policy arrangements will bring to mind issues that have arisen with the so-called "seasteading" movement to found autonomous floating city-states (challenges to which are discussed in Friedman and Taylor 2012) and Romer's (2010) proposed "charter cities," which are jurisdictions within existing countries whose institutions are designed on a "clean sheet" basis (although political resistance has handicapped their development).

The achievement of such visions will take time, perhaps a very long time. Many of the key questions for the economic development of space will be technological. But there will also be considerable room for scholars of economic development, industrial organization, public finance, economic history, and other specialties, to begin the work of understanding, improving, and even shaping the development of the space economy.

■ *Thanks to Henry Hertzfeld, Roger Launius, Benjamin B. Lockwood, John Logsdon, Alexander MacDonald, Brent Neiman, and Danny Yagan and participants in the Working Group on the Business and Economics of Space at Harvard Business School for helpful discussions and to Enrico Moretti and Timothy Taylor for valuable editorial advice. Angela Acocella provided outstanding suggestions and research assistance.*

References

Abtan, Olivier, Antonio Achille, Jean-Marc Bellaiche, Youlee Kim, Vincent Lui, Amitabh Mall, Antonella Mei-Pochtler, and Sarah Willersdorf. 2014. "Shock of the New Chic: Dealing with New Complexity in the Business of Luxury." Boston Consulting Group Report, January 30.

Achenbach, Joel. 2013. "Which Way to Space: Flights of Fancy May Launch the Industry's Future." *Washington Post*, November 23.

Bromberg, Joan Lisa. 1999. *NASA and the Space Industry*. Johns Hopkins University.

Bryce Space and Technology. 2017. "Start-Up Space: Update on Investment in Commercial Space Ventures." (The Space Start-up Report.) Bryce.

Brynjolfsson, Erik, and Paul Milgrom. 2013. "Complementarity in Organizations." Chap. 1 *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, 11–55. Princeton University Press.

Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.

Cordiner, Ralph J. 1961. "Competitive Private Enterprise in Space." Chap. 10 in *Peacetime Uses of Outer Space*, edited by Simon Ramo. McGraw-Hill.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics* 131(2): 799–858.

Duff-Brown, Beth. 2015. "The Final Frontier has Become Congested and Contested." Stanford Center for International Security and Cooperation, March 4. http://cisac.fsi.stanford.edu/news/security-space-0.

Ehrenfreund, Pascale, Margaret Race, and David Labdon. 2013. "Responsible Space Exploration and Use: Balancing Stakeholder Interests." *New Space* 1(2): 60–72.

Henry, Caleb. 2016. "Euroconsult Study Says SmallSat Market Could Be Worth $22 Billion over Next 10 Years." *Via Satellite*, July 7.

Hertzfeld, Henry R. 2007. "Globalization, Commercial Space and Spacepower in the USA." *Space Policy* 23(4): 210–20.

FAA. 2018. U.S. Federal Aviation Administration online data, accessed March 2018: https://www.faa.gov/data_research/commercial_space_data/launches/?type=license and http://www.faa.gov/data_research/commercial_space_data/launches/?type=permitted.

Foust, Jeff. 2017. "Smallsat Developers Propose Self-Regulation to Address Orbital Debris Concerns." *SpaceNews*, August 15. http://spacenews.com/smallsat-developers-propose-self-regulation-to-address-orbital-debris-concerns/.

Friedman, Patri, and Brad Taylor. 2012. "Seasteading: Competitive Governments on the Ocean." *Kyklos* 65(2): 218–35.

Handberg, Roger. 1995. *The Future of the Space Industry: Private Enterprise and Public Policy*. Quorum.

Hanson, Ward. 2016. "Pricing Space Debris." *New Space* 2(3): 143–44.

Hertzfeld, Henry R. 1992. "Economic Issues Facing the United States in International Space Activities." In *Space Economics*, edited by Joel S. Greenberg and Henry R. Hertzfeld, pp. 417–35. AIAA.

Hertzfeld, Henry R., Brian Weeden, and Christopher D. Johnson. 2016. "Outer Space: Ungoverned or Lacking Effective Governance?: New Approaches to Managing Human Activities in Space." *SAIS Review of International Affairs* 36(2): 15–28.

Kaufman, Ellie. 2015. "More Students Are Going to Space School—And It May Be a Better Idea Than You Think." *Mic*, June 24.

Krolikowski, Alanna, and Martin Elvis. 2017. "In Pursuit of Science, Settlement, Security, or Sales? Making Policy for Asteroid Activities." Working paper.

Lambright, W. Henry. 2016. "Leadership, Change, and Public–Private Partnerships: A Case Study of NASA and the Transition from Space Shuttle to Commercial Space Flight." Improving Performance Series. The IBM Center for The Business of Government.

Launius, Roger D. 2006a. "Compelling Rationales for Spaceflight? History and the Search for Relevance." Chap. 2 in *Critical Issues in the History of Spaceflight*, edited by Steven J. Dick and Roger D. Launius. Washington, DC: NASA.

Launius, Roger D. 2006b. "Assessing the Legacy of the Space Shuttle." *Space Policy* 22(4): 226–34.

Launius, Roger D. 2014. "Historical Analogs for the Stimulation of Space Commerce." Monographs in Aerospace History, no. 54, July 1. NASA.

Launius, Roger D. 2017. Personal communication.

Launius, Roger D., and Howard E. McCurdy. 1997. *Spaceflight and the Myth of Presidential Leadership*. University of Illinois Press.

Locke, John. 1689. "Of Property." Chap. 5 of *Second Treatise of Government*.

Logsdon, John M. 2004. "A Sustainable Rationale for Human Spaceflight." *Issues in Science and Technology* 20(2): 31–34.

Logsdon, John M. 2011. "A New US Approach to Human Spaceflight?" *Space Policy* 27(1): 15–19.

Logsdon, John M. 2015. *After Apollo: Richard Nixon and the American Space Program*. Palgrave Macmilian.

Macauley, Molly K. 2004. "Economics of Space." In *Space Politics and Policy,* pp. 181–200. Springer.

Macauley, Molly K. 2005. "Advantages and Disadvantages of Prizes in a Portfolio of Financial Incentives for Space Activities." *Space Policy* 21(2): 121–28.

Macauley, Molly K., and Michael A. Toman. 1991. "Providing Earth Observation Data from Space: Economics and Institutions." *American Economic Review* 81(2): 38–41.

MacDonald, Alexander. 2014. "Emerging Space: The Evolving Landscape of 21st Century American Spaceflight." NASA.

MacDonald, Alexander. 2017. *The Long Space Age: The Economic Origins of Space Exploration from Colonial America to the Cold War*. Yale University Press.

Martin, Paul K. 2011. "NASA's Commercial Crew Development Program." Prepared statement to the US House of Representatives Committee on Science, Space, and Technology, October 26, https://oig.nasa.gov/IG_Statement_NASAs_CCDev_Program_10_26_2011.pdf.

Metzger, Philip. T., Anthony Muscatello, Robert P. Mueller, and James Mantovani. 2013. "Affordable, Rapid Bootstrapping of the Space Industry and Solar System Civilization." *Journal of Aerospace Engineering* 26(1): 18–29.

Morring, Frank Jr. 2011. "The New Space Race." *Aviation Week*, April 25. http://archive.aviationweek.com/issue/20110425.

Musk, Elon. 2017. "Making Humans a Multiplanetary Species." New Space 5(2).

NASA. 2014. *Commercial Orbital Transportation Services: A New Era in Spaceflight*. https://www.nasa.gov/sites/default/files/files/SP-2014-617.pdf.

NASA. 2017. "Monthly Number of Objects in Earth Orbit by Object Type." A figure in *Orbital Debris Quarterly News* 21(1): 12. https://orbitaldebris.jsc.nasa.gov/quarterly-news/pdfs/odqnv21i1.pdf.

National Space Society. 2012. *Milestones to Space Settlement: An NSS Roadmap*. http://space.nss.org/milestones-to-space-settlement-an-nss-roadmap/.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Basic Books

O'Neill, Gerard K. 1976. *The High Frontier: Human Colonies in Space*. Bantam.

Ostrom, Elinor. 2009. "Beyond Markets and States: Polycentric Governance of Complex Economic Systems." Nobel Prize Lecture, delivered in Stockholm, Sweden, December 8.

President's Commission on Implementation of United States Space Exploration Policy. 2004. *A Journey to Inspire, Innovate, and Discover*. June. http://www.nasa.gov/pdf/60736main_M2M_report_small.pdf.

Space Foundation. 2018. *The Space Report*. Available at: https://www.thespacereport.org/.

SpaceX. 2015. "Reusability: The Key to Making Human Life Multi-Planetary." June 10. http://www.spacex.com/news/2013/03/31/reusability-key-making-human-life-multi-planetary.

Sunyer, John. 2014. "The New Market Space: Billionaire Investors Look beyond Earth." *Financial Times*, February 28.

Romer, Paul. 2010. "Technologies, Rules, and Progress: The Case for Charter Cities." Center for Global Development Essay, March 2010. Accessed at: http://www.cgdev.org/content/publications/detail/1423916.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(5): 416–24.

Weinzierl, Matthew, and Angela Acocella. 2016. "Blue Origin, NASA, and New Space." Harvard Business School Case 716-012.

Weinzierl, Matthew, Angela Acocella, and Mayuka Yamazaki. 2016. "Astroscale, Space Debris, and Earth's Orbital Commons." Harvard Business School Case 716-037.

Zapata, Edgar. 2017. "An Assessment of Cost Improvements in the NASA COTS/CRS Program and Implications for Future NASA Missions." NASA. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170008895.pdf.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 193–208

Dave Donaldson: Winner of the 2017 Clark Medal

Daron Acemoglu

The 2017 John Bates Clark Medal of the American Economic Association was awarded to Dave Donaldson for his path-breaking contributions in international trade. Donaldson's work sheds light on some of the central questions of international economics, ranging from the economic and welfare implications of market integration within a country to testing the core empirical predictions of models of international trade based on comparative advantage. In these areas, empirical work faces the challenge of taking into account the broader equilibrium implications of changes in policies or economic conditions—that is, the possibility that bilateral relations between two regions or countries will affect others via trade diversion or their effects on equilibrium prices. Donaldson's work has managed to address these challenges by combining careful theory with detailed and creative empirical work. Indeed, this research strategy has turned Dave into a leader in the revival of empirical work in international trade.

Dave Donaldson, a native of Canada, grew up in Toronto. He graduated from high school in 1997 and moved to Trinity College, Oxford, with a scholarship to study physics. The debates and protests about globalization in the late 1990s piqued Dave's interest in economics and carried him to the London School of Economics. Economics stuck with Dave, and vice versa, and after completing his master's degree in 2003, he continued on to the PhD program.

The questions that had motivated Dave to delve into economics combined with his childhood interests in geography and maps (his favorite sport was orienteering,

■ *Daron Acemoglu is the Elizabeth and James Killian Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. His email address is Daron@mit.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.193

doi=10.1257/jep.32.2.193

Dave Donaldson

which combines cross-country running with off-trail navigation using detailed topographical maps) and led to his chosen specialization: empirical international trade that takes geography seriously. He graduated from the London School of Economics in 2009, with a dissertation focusing on the effects of railways and economic integration on cross-district inequality of economic and social outcomes in India. He joined the MIT faculty that year and spent the next five years there, while also visiting Harvard and Stanford during that period. In 2014, he moved to Stanford University, where he was a faculty member in April 2017 when he was awarded the Clark Medal. He left Stanford and returned to MIT in July 2017.

In this essay, I will attempt to put Dave's research in the broader context of work over the last several decades on market integration and various topics in international trade. I will focus on the key papers listed by number in Table 1.

Background

Many of the classic works in economics such as Adam Smith's (1776) *Wealth of Nations* or David Ricardo's (1817) *Principles of Political Economy and Taxation* centered on international trade. Despite this illustrious background and some of the best minds in economics having devoted their careers to this topic, there is still much we do not know about the extent of gains from trade—and even about whether countries trade in the way that economic theories predict. Given that uncertainty, it is perhaps unsurprising that international trade is often such a controversial topic.

Table 1
Selected Papers by Dave Donaldson

1. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." Forthcoming. *American Economic Review.*
2. "Railroads and American Economic Growth: A 'Market Access' Approach," (with Richard Hornbeck). 2016. *Quarterly Journal of Economics* 131(2): 799–858.
3. "How Large Are the Gains from Economic Integration? Theory and Evidence from US Agriculture, 1880–1997," (joint with Arnaud Costinot). NBER Working Paper 22946.
4. "Who's Getting Globalized? The Size and Implications of Intranational Trade Costs," (joint with David Atkin). 2014. Unpublished.
5. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," (with Arnaud Costinot and Ivana Komunjer). 2012. *Review of Economic Studies* 79(2): 581–608.
6. "Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence," (with Arnaud Costinot). 2012. *American Economic Review* 102(3): 453–58.
7. "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," (with Arnaud Costinot and Cory Smith). 2016. *Journal of Political Economy* 124(1): 205–248.
8. "The Elusive Pro-Competitive Effects of Trade," (with Costas Arkolakis, Arnaud Costinot, and Andrés Rodríguez-Clare). Forthcoming. *Review of Economic Studies.*
9. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," (with Rodrigo Adao and Arnaud Costinot). 2017. *American Economic Review* 107(3): 633–89.
10. "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect," (with Arnaud Costinot, Margaret Kyle, and Heidi Williams). 2016. NBER Working Paper 22538.
11. "Comparative Advantage and Optimal Trade Policy," (with Arnaud Costinot, Jonathan Vogel, and Iván Werning). 2015. *Quarterly Journal of Economics* 130(2): 659–702.

The readily visible costs from trade with other countries include firm shutdowns and jobs lost at those firms or in those industries, as well as reduced diversity of national production. The benefits of trade often seem less apparent.

Trade questions are some of the "big questions" of economics, and have become, if anything, more salient in recent years, not just because of the evident discontent of a large share of the public concerning international trade, but also because some prominent recent research has highlighted some of the costs of trade and the uncertainties about benefits. For example, Autor, Dorn, and Hanson (2013) study the employment losses in US local labor markets following the rise in imports from China, and find costs that are higher than many economists had presumed. Rodrik (1997) and a few others have offered an influential contrarian voice on whether the gains from globalization are as large as what many economists have argued.

Overall, measurement of gains from an expansion of trade remains difficult because of several interrelated challenges.

The first and fundamental challenge is one of counterfactual analysis: that is, a researcher does not observe what regions or countries would produce and how productively it would be produced if expanded trade had not come into existence in the first place. For example, what would the patterns of production, economic organization, and transportation have looked like across the 19th-century United States, or across 19th-century India, if the building of a vast rail network had not opened up new possibilities for market integration?

A second challenge is that any decision about linking any two local markets, whether via railways or trade agreements, involves choices made by economic and political agents, and such choices are likely to be correlated with current and future economic prospects. For example, we cannot extrapolate the estimates of increased trade from a rail link between an area with rich agricultural or mineral resources and a port city to argue that railroads connecting two isolated, resource-poor areas would have the same effects.

Third, equilibrium effects also complicate inference—linking two markets will not only change economic outcomes in these two markets, but potentially in many others. The possible changes may include diversion of imports or exports to the newly linked markets, and changes in the prices of goods and factors resulting from increased trade and specialization. Put in the language of microeconometrics, when investigating within- (or, for that matter, between-) country trade, the "stable unit treatment value assumption"—the bedrock of simple empirical strategies—is violated almost by design.

Addressing these issues in a systematic way required fundamental advances within the profession to generate credible empirical designs and develop models of trade and economic geography. The former would help us to exploit sources of variation that come closer to identifying exogenous changes that arise from market integration, and the latter would discipline how we can move from local effects to an inference regarding (general) equilibrium outcomes. A major part of Dave Donaldson's work is at the forefront of these challenges.

Market Integration

The classic theory of international economics has been on trade *between* countries. Yet in most countries, the integration of markets has been a slow and still incomplete historical process. The study of gains from *within-country* integration illustrates the challenges facing the measurement of gains from an expansion of trade.

Robert Fogel's (1964) seminal work revived interest in the consequences of within-country market integration and has shaped much of the literature in the subsequent five decades. The hallmark of Fogel's work was his focus on the contribution of railroads to economic growth in 19th-century America, possibly the most iconic conduit of market integration for most countries (or at least for those countries not fortunate enough to be crisscrossed by rivers and canals). Fogel sought to

spell out an alternative history of how transportation networks like canals and roads might have evolved in the absence of the railroads, which led to his often-cited conclusion that "the railroad did not make an overwhelming contribution to the production potential of the economy" (p. 235).

Not surprisingly, Fogel's conclusions were controversial, as captured pithily by Paul David's (1969) memorable title "Professor Fogel on and off the Rails." But setting aside the specifics of that debate, there was a more enduring methodological contribution in Fogel's work: the emphasis that economists need to worry about counterfactuals. Fogel's approach to building a counterfactual for 19th-century US railroads was ingenious, but it is fair to say that it did not fully tackle the counterfactual questions involved.

Dave Donaldson confronts this set of questions in the historical context of India in a paper [numbered 1 in Table 1] which grew out of his PhD dissertation. In several ways, this choice of context for this study is ideal. India has been an exemplar of a nonintegrated internal market for most of its history (and arguably even today). However, the railway investments of the British Raj were a huge step in bringing somewhat closer integration of Indian districts; between 1853 and 1930, British authorities laid over 67,000 kilometers of rail tracks in colonial India.

The first step in Dave's approach is to develop a credible empirical strategy to estimate the *local* effects of railway access. At a conceptual level, the goal is to answer the question: If one Indian district is randomly allocated access to the railway network, how much does that district gain relative to another district that is randomly denied access to this network? Obviously, we don't have this random allocation in practice, and comparing a district that does get access to the railway network, to one that doesn't, won't do, because there is quite a bit of planning on the part of a relatively sophisticated bureaucracy on where the railway network should be located.

Dave's strategy here is to exploit the archival sources in several ways. He builds a new dataset of district-level real incomes in India and also obtains detailed information about the building of its railway network. He also collected information on railroad lines proposed to be built that, for some reason, did not get built or got built only with considerable delay. Using these data sources, Dave compares districts that got access to the railway network, not to all of those that didn't, but only to those that seemed to be desirable locations for rail stations and, in fact, had a rail station planned for them, but in the end, didn't get it. This strategy enables Dave to establish that railroads reduced interregional price differentials and, consistent with theory, reduced the responsiveness of prices to local productivity shocks. It also provides an estimate of the effects of railways on local (agricultural) incomes: districts that got further integrated with other parts of India gained about 16 percent more agricultural income relative to those that did not.

Of course, we cannot extrapolate from this 16 percent estimate of gains to specific districts to conclude that Indian incomes overall grew by 16 percent (or grew by 16 percent times the fraction of GDP that was in districts that got access to the rail network). It is possible that districts not connected to the rail network

experienced a decrease in income because trade got further diluted away from them, or going in the opposite direction, these districts may have benefited from changes in equilibrium prices. These indirect effects would need to be taken into account in any overall calculation.

One can partially deal with these concerns by looking for effects on districts neighboring those that received a rail station, where we expect indirect effects might most likely be felt. Indeed, Dave shows in the working paper version of [1] that a neighbor's access to railways reduced a district's real income level significantly. But a reduced-form empirical analysis focusing on neighbors is not sufficient; in full general equilibrium, *all* districts might be affected.

This is where [1] turns to the advances in the theory of international trade. In particular, Dave turns to the "Ricardian" mode of trade developed by Eaton and Kortum (2002; described in this journal in Eaton and Kortum 2012), which provides a tractable setup in which patterns of trade and their welfare implications can be studied in the presence of trade costs and productivity differences across locations. In Eaton and Kortum's approach, each region (or country) has different productivities in the production of different goods (which is the Ricardian element) and shipping goods from one district to another is costly. The resulting equilibrium determines factor prices within each region, which in turn pins down the cheapest producer of each good for each region after taking shipping costs into account. The model also verifies that reducing transport costs between two regions will affect equilibrium prices and thus can affect all regions. What makes the model tractable is that although there are price differences across regions even for goods with the same origin, because of trade costs, there is a straightforward pattern of flows between any two regions. These flows are related to the classic "gravity equation," which links bilateral trade between two countries to their "economic masses" (GDPs) and the geographic and other types of distance between the two.

The structure implied by Eaton and Kortum's (2002) model also links welfare gains from trade to a sufficient statistic, the "trade share" of a region's expenditure on its own goods (in autarky this trade share is of course one). This result enables Dave Donaldson to verify that this theoretical approach is a good approximation to the effects of railways on real agricultural incomes throughout India by confirming whether all of the effects of railways work through this trade share. Once the model's implications are thus validated, Dave can use its structure to estimate the overall welfare consequences of the improved transportation network in colonial India, and he infers that any spillovers on other districts are not quantitatively large. (This conclusion also explains why the estimated quantitative effects of market integration are, even if not identical, similar to those of Fogel's approach, which did not consider these indirect effects.)

This work is an excellent specimen of the style of work that has made Dave a leader in the study of empirical effects of trade across regions and countries. The project is motivated by a challenging question concerning the overall (general) equilibrium effects of a change in transport infrastructure. This question is answered by combining new data, careful reduced-form empirical work, and theory

and inference about structural parameters that are informative about equilibrium effects that go beyond local impacts.

Dave also investigated the effect of railways in joint work with Richard Hornbeck in [2] by revisiting the effects of the so-called "iron horse" in the 19th-century United States. The challenging questions again concern equilibrium effects of the massive expansion of the railway network. To tackle these questions, Dave and Richard make three methodological advances relative to previous work. First, they again use a Ricardian model of trade that builds on Eaton and Kortum (2002), where the effect of reduced transport costs on the economy of a region (here county) is linked to the "market access" of that region. The measure of market access takes into account not just the density of the railway links, but how getting connected to a hub such as Chicago provides Midwestern counties further access to other markets. This provides an alternative implementation of the same gravity-equation structure as that in [1]. Second, they focus on the value of agricultural land, which under the assumption of competitive markets should capture the current and future improvements from improved access. Third, they build a detailed county-level dataset of the railway network and canals exploiting the geographic information system (GIS) network database.

Their estimates show a strong correlation between changes in the measures of market access, driven by the rollout of the railway network, and long-run changes in the value of agricultural land. A 10 percent increase in market access is associated with a 5 percent increase in the value of agricultural land. But as in [1], this reduced-form relationship may reflect endogenous choices of where railroads were built. As a partial method to deal with this problem, the authors use the source of variation due to water market access in 1870: specifically, higher water market access in 1870 implies a lower change in overall market access due to railroads between 1870 and 1890, because high water market access counties make railways less useful at improving access at the margin. This source of variation leads to even larger estimates, now implying that a 10 percent increase in market access leads to an over 11 percent increase in the value of agricultural land.

With these estimates at hand, the paper proceeds to perform the same counterfactual as Fogel's (1964) classic work, investigating what the consequences would be of removing all the railroads in 1890. They find that such a step would have reduced the total value of US agricultural land in 1890 by approximately 60 percent. This number is fairly large. But if we view land values as corresponding to the present discounted returns of land, assume that there was no anticipation of the expansion of the railway network before 1870 and no anticipation of further productivity-enhancing investments in affected counties, and take an interest rate of 5 percent, then the effects of railways on land values they estimate are equivalent to an increase of about 3 percent of total national income annually (which is only modestly larger than Fogel's estimate).

While [1] and [2] focus on historical studies of railways, in [3] Dave Donaldson and his frequent collaborator and colleague Arnaud Costinot turn to the effects of overall US economic integration in the agricultural sector between 1880 and 1997,

which has dramatically reduced the role of distance and enabled much greater specialization in agriculture. This paper takes a more holistic approach to the question of gains from trade and adopts a more ambitious approach than previous work by recognizing that the productivity of different plots of land vary greatly depending on what crops are grown there. To make matters more challenging, even though we do observe the allocation of different plots of land to different crops today, we do not know what this would have been and how much productivity would have been lower had it not been for this major process of integration over the last century and a half.

To overcome this problem, the authors adapt Costinot's (2009) earlier theoretical framework to model the allocation of heterogeneous land parcels to different crops across 2,636 US counties. To implement this approach, as in any Ricardian model, one needs to have estimates of the productivity of a given land parcel for every crop—since in the absence of integration, they may have chosen to produce some of the crops that they are not currently producing. Their ingenious idea is to use the modern production function of crops to infer the comparative advantage patterns. They do this using data from the agronomic (GAEZ) project from the Food and Agriculture Organization of the United Nations. This database uses agronomic models and high-resolution data on geographic characteristics and climatic conditions to predict crop yields at the level of relatively small grid cells (roughly 10 km by 10 km at the equator) covering the entire globe. Under the assumption that within-county comparative advantage patterns of the late 19th century are not reversed today, Dave and Arnaud show how to use data on total farm sales in a county, total output per crop, and total land allocated to each crop to infer the unique vector of crop prices and crop-and-county productivity shocks that is consistent with profit-maximization and factor-market clearing in the county. They assume that production functions are linear, so they can solve for equilibrium using computationally straightforward linear programming analysis.

One way of validating this approach is to compare the price implications of the model to data. Although county-level crop prices are not observable, state-level prices are. The authors show that the implications of the model that follow from profit maximization and factor market clearing are highly correlated with observed state-prices. The same data also confirm the decline of spatial price dispersion over time.

The ultimate objective is to estimate the contribution of greater integration of agricultural markets to economic growth. Using this powerful framework, the paper estimates that a significant fraction of economic growth of agricultural output (perhaps as much as 80 percent of it) may be due to economic integration.

In Donaldson's joint work with David Atkin in [4], they take a very different approach to studying the implications of within-country trade barriers, directly measuring how prices vary within a country as a function of distance. To achieve this, they use barcode-level price data from Ethiopia, Nigeria, and the United States (included as a comparison), and collect new data on the origin of products in order to determine which are the location pairs that are trading and hence have price gaps that are directly informative of trade barriers. They then develop a model of

pass-through to discipline the empirical work and also allow for markups that vary over space, which is essential to understanding whether price differences simply reflect within-country trade costs, or also reflect differential pricing strategies and markups set by intermediaries.

Focusing on a sample of goods that are identified at the barcode level enables the authors to minimize any unobserved quality differences, and reveals sizable cross-country differences in the effects of distance on prices. In particular, the effects of distance on trade costs appear to be four or five times larger in Ethiopia and Nigeria than in the United States. But this is only part of the story. Atkin and Donaldson show that markets are less competitive in remote locations, and consequently, the gains from globalization, which reduce trade costs, are greater for consumers in locations closer to where goods originate and lower for those in distant locations. These findings imply that the gains from reduced trade costs following from globalization may not benefit some consumers as much, and in fact such gains are likely to be unequally distributed for reasons beyond those already emphasized in the literature—in particular, because of differential pass-through and markups.

Empirics of Comparative Advantage

David Ricardo's (1817) classic analysis of international trade, which links trade patterns and specialization to an economy's comparative advantage, is of course one of the mainstays of the economic canon. All the same, systematic empirical investigation of the predictions of the approach have been few and far between. This is both because tractable models of Ricardian trade specifying bilateral trade flows when countries may specialize only in a subset of the available goods were not developed until recently, and also because detailed data for the empirical analysis were not widely available. To be sure, there have been plenty of empirical papers linking exports to various measures of productivity to get to one of the key implications of Ricardian comparative advantage—that countries should export more in sectors in which they are more productive. Yet often these empirical exercises can seem ad hoc because they were not explicitly linked to the predictions of a fully specified Ricardian model of trade.

However, Dave Donaldson's work offers an in-depth empirical investigation of the predictions and implications of the patterns of comparative advantage in Ricardian trade models. It is fitting that the American Economic Association awarded Dave Donaldson the Clark Medal almost on the 200th anniversary of the publication of David Ricardo's *Principles of Political Economy and Taxation* (which was published on April 19, 1817). Though they have not fully tested the implications of the Ricardian model, Dave's study [1] of the effects of railways in India and his joint work with Costinot [4] on welfare gains from economic integration in US agriculture have been important precursors of this type of research. Dave's explicit work in this area consists of several papers coauthored with with Arnaud Costinot, and in some instances with other coauthors as well.

In [5], Arnaud and Dave team up with Ivana Komunjer to overcome the challenges confronting empirical studies of the general Ricardian framework. They first develop a rich structural Ricardian model of trade. This framework goes beyond Eaton and Kortum's (2002) model to specify trade patterns as a function of productivity differences in a setting with multiple countries, several goods, and multiple varieties of each good. It explicitly allows countries to specialize in a subset of the goods depending on their comparative advantage and factor prices. The framework provides a particularly straightforward form of the basic Ricardian mechanism—whereby countries export relatively more in sectors in which they are relatively more productive—and provides closed-form expressions for bilateral flows as a function of observed productivity patterns. Crucially, the model takes into account the fact that a country will not produce all varieties of every good but rather those varieties in which it is relatively more efficient. This analytical structure implies that differences in observed productivity tend to be smaller than true differences in productivity because of a "trade selection effect"—countries tend to produce the varieties in which they are more productive.

Combining the empirical equation that emerges from the theory with data on trade flows and producer prices, which should reflect productivity, the authors establish that countries do export goods where their relative productivity is higher, as one might have expected. For example, their core estimates imply that a 1 percent increase in relative productivity is associated with a 6.5 percent increase in relative exports of a country. Using their estimate of the key structural parameter of the model—the dispersion of productivity across varieties within a sector—the authors proceed to quantify the welfare impact of this Ricardian channel across sectors. They find that cross-industry differences in productivity generate only a small part of the gains from trade, and instead it appears to be comparative advantage differences at the within-industry level that account for most of the gains from trade.

This paper [5] is an important one, because it makes significant advances relative to the previous literature on teasing out the predictions of a canonical Ricardian model of international trade and confronting them with data. Though the data used is not very fine-grained (sectoral data from the Groningen Productivity Database), the theoretical predictions are borne out in the empirical analysis, and do suggest that Ricardian comparative advantage plays an important role in the observed trade patterns and in the welfare gains from international trade.

In [6], Arnaud and Dave turn to the much more detailed agricultural data from the GAEZ database (mentioned above), which allow them to use more fine-grained variation than in [5] while tackling the fundamental problem of Ricardian trade empirics—specialization means that we do not observe the productivity of the country in the goods that it imports. They show how the parcel-level information from this database can be used to make the Ricardian model more operational. As in [5], but this time using GAEZ-predicted agricultural productivity rather than measured manufacturing productivity, the authors document a positive correlation between cross-country comparative advantage and cross-country patterns of specialization in agriculture.

In [7], Arnaud and Dave team up with Cory Smith to investigate how Ricardian trade and specialization patterns might mitigate the global implications of climate change in agricultural markets. This paper also uses the rich data from GAEZ, this time to investigate how changes in production patterns within countries might mitigate the adverse consequences of climate change. To do this, they develop a detailed micro-founded model of allocation of land to crops and trade patterns, related to [3], but now applied to over 9,000,000 grid cells from 187 countries from the GAEZ data for 10 distinct crops. Crucially, the GAEZ dataset is available both under contemporary growing conditions and under the climate change scenarios used by the UN's Intergovernmental Panel on Climate Change (IPCC), which is the core input into the paper's counterfactual exercises.

By feeding these micro-level productivity shocks into their general equilibrium trade model, Arnaud, Dave, and Cory estimate that, absent a reallocation of land parcels to crops, unmitigated climate change will translate into large negative productivity shocks for many countries around the world, decreasing world welfare substantially. However, there is enough heterogeneity in these shocks over space that the reallocation of production according to comparative advantage across crops within each field reduces the welfare impact of climate change by an order of magnitude. Furthermore, there is so much productivity heterogeneity across fields within countries that allowing countries to adjust their patterns of trade internationally appears to have very small effects on the welfare consequences of climate change. The key to reducing the very negative effects of climate change, therefore, lies in changing production patterns within countries, and not so much in international trade.

In [8], Arnaud and Dave join forces with Costas Arkolakis and Andrés Rodríguez-Clare to investigate the gains from trade when trade affects markups. Earlier work by Arkolakis, Costinot, and Rodríguez-Clare (2012) linked the gains from international trade in a range of models to bilateral trade flows (via the gravity equation) and showed that in a number of important cases, these gains are quite small. This conclusion stands in stark contrast to the large gains from trade estimated in regression studies such as Frankel and Romer (1999) and Feyrer (2009). As discussed above, Dave's work [1] on railways of India and his joint work with Arnaud [3] on US agriculture also found significant gains from economic integration. Could it be that [8] obtains different results because the effects of international trade on markups—the so-called "pro-competitive effects"—are ignored in the earlier papers? Indeed, one may have conjectured that the gains from trade could be much higher if greater trade intensifies competition.

Accordingly, [8] considers a richer class of models that allows for demand elasticities to vary with quantities, which implies that trade will alter equilibrium markups. Though this class of models is still restrictive, it nests many models with variable markups that have been used in the previous trade literature. Surprisingly, at the estimated demand parameters, in which demand elasticities decrease with the level of consumption (consistent with the common empirical finding of incomplete pass-through), the effect of trade on markups turns out to reduce, rather than increase,

the welfare effects of trade. More precisely, the welfare gains are bounded from above by the same macro-level elasticity-based estimates provided in the earlier paper.

There is a new and surprising insight underlying this seemingly paradoxical result. A reduction in barriers to trade has two opposing effects on monopoly markups: it reduces the markup of domestic firms (because they face tougher competition from abroad), but it increases the markups of foreign firms (because they now have lower costs of serving the domestic market and reductions in costs tend to be incompletely passed-through). Given their estimated demand parameters, the second effect turns out to dominate. Clearly, this result need not extend to richer market structures. Yet it qualifies the important conditions under which we may expect, and may not expect, the "pro-competitive" effects of trade to be present, and it provides yet another useful benchmark by encompassing the kinds of models that trade economists commonly use for studying the effects of trade liberalization.

Another important paper within the same research program is [9], joint with Rodrigo Adao, which develops a new methodology to construct nonparametric counterfactual predictions, free of functional-form restrictions on preferences and technology, in neoclassical trade models. To do so, Rodrigo, Arnaud, and Dave establish the equivalence between such models and "reduced exchange models" in which countries directly exchange factor services. This equivalence implies that, for an arbitrary change in trade costs, counterfactual changes in the factor content of trade, factor prices, and welfare only depend on the shape of a reduced factor demand system. They then provide sufficient conditions under which estimates of this system can be recovered nonparametrically while using the same data sources and exclusion restrictions that are typically invoked. Together, these results offer a strict generalization of the parametric approach used in so-called gravity models.

Finally, in [10], Arnaud and Dave join forces with Margaret Kyle and Heidi Williams to investigate the origins of the productivity differences that are at the root of the Ricardian models of international trade. One approach in the international trade literature to this question has been the "home market effect," which suggests that countries should be more productive and export more in sectors where they have a larger home market. The home market effect emerges as a source of endogenous comparative advantage in models of monopolistic competition and trade costs (or nonhomothetic preferences) because a larger home market is an advantage for domestic firms and incentivizes more of them to enter and serve this greater market. This entry then leads to more varieties or other sources of greater productivity. The home market effect is also related to the endogenous direction of innovation, since one of the factors increasing an economy's productivity in a specific sector might be greater innovative activity directed to that sector because of its greater importance, though the home market effect might also be due to other sources of industry-level economies of scale linking productivity to the level of domestic production.

The paper explores the home market effect in the context of drug exports. It builds on previous work on the effect of demographic changes on innovation and product entry such as the analysis of demographic change and new pharmaceutical

innovations in Acemoglu and Linn (2004). The bottom line result of Arnaud, Dave, Margaret, and Heidi's work is that countries that for demographic reasons are expected to have high demand for a certain type of drug are actually more likely to be net exporters of the same drugs, and in fact, the greater the home demand predicted, the greater the sales abroad. With a naïve model with fixed productivities, one would have expected the opposite—the more a country demands of a particular drug, the more of this drug it will need to import and the less it would be expected to sell to other countries. The home market effect is particularly plausible in the context of drugs because of the importance of endogenous innovation activity in this sector, a pattern consistent with the authors' evidence on the stronger response of nongeneric (relative to generic) drugs to the size of the home market.

Trade Policy

The insights of Donaldson's empirical work offer indirect lessons and insights about trade policy, but in [11], Dave and Arnaud work with Jonathan Vogel and Iván Werning to investigate more explicitly the implications of Ricardian comparative advantage for the design of optimal trade policy. The theory of optimal trade policy in models of trade based on differences in factor endowments is relatively well understood (for example, Dixit 1985). However, some basic questions in realistic cases of differences in comparative advantage had not been confronted. These include, for example, whether a country should protect more in importing sectors or whether it should subsidize more in exporting sectors that already have a strong comparative advantage. The paper has a very sharp answer to these questions: optimal import tariffs should be uniform across sectors, regardless of the pattern of comparative advantage, while optimal export subsidies should be nonincreasing with comparative advantage.

Though the theoretical results in [11] are sharp, questions about the extent to which these conclusions apply in realistic contexts continue. For example, many countries protect their least competitive sectors, and some countries tend to subsidize export sectors. This might be because of additional constraints on trade policy or because of political economy considerations. But as with so many results in optimal tax theory, we need to understand the benchmark problem of a benevolent and unconstrained planner in order to move on to a deeper understanding of actual political behavior.

Summary

Dave Donaldson together with colleagues and collaborators such as Arnaud Costinot and David Atkin have been at the forefront of the revival in empirical international trade. They have catapulted this field into one of the most dynamic areas of the last decade and a half.

The challenges facing empirical work in this field cannot be understated. Most questions in international trade have a major equilibrium component, and thus simple reduced-form strategies, comparing one economy, region, industry, or firm to another will not provide fully satisfactory answers. Turning to fully specified structural models is an option, but most models rely on a myriad of simplifying assumptions, and so that would not be fully satisfactory either. Dave's work has overcome these challenges by combining careful, credible, reduced-form work (based on new data and interesting institutional settings) with powerful modern trade theory to estimate not just local effects but full equilibrium impacts. In this fashion, he has spearheaded the analysis of within-country economic integration, especially owing to major advances in transport technology in the 19th and 20th centuries, and also contributed to bridging the gap between theory and empirics in the study of international trade flows.

The hallmark of Dave Donaldson's work is that it so readily combines a range of elements: construction of new datasets; credible reduced-form empirical analysis; sound economic theory targeted to the question at hand; and thoughtful estimation of structural parameters to carry out counterfactual analysis and estimate welfare effects. This style of work, combining careful reduced-form estimation together with state-of-the-art models to infer underlying structural parameters and estimate the welfare consequences of major policy changes, has not only set a high standard for empirical work in international trade, but is becoming the norm in many different areas of economics. Many young scholars in international trade and other fields will surely seek to emulate and build on Dave's approach and intellectual leadership in the years ahead.

■ *I am grateful to David Atkin, Dave Donaldson, Elhanan Helpman, and James Poterba for useful discussion and comments.*

References

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Innovation from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119(3): 1049–90.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains." *American Economic Review* 102(1): 94–130.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Costinot, Arnaud. 2009. "An Elementary Theory of Comparative Advantage." *Econometrica* 77(4): 1165–92.

David, Paul A. 1969. "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails." *Economic History Review* 22(3): 506–525.

Dixit, Avinash. 1985. "Tax Policy and Open Economies." Chap. 6 in *Handbook of Public Economics*, vol. 1, edited by Alan Auerbach and Martin Feldstein, 313–74. Elsevier

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70(5): 1741–79.

Eaton, Jonathan, and Samuel Kortum. 2012. "Putting Ricardo to Work." *Journal of Economic Perspectives* 26(2): 65–90.

Feyrer, James. 2009. "Trade and Income—Exploiting Time-Series in Geography." NBER Working Paper 14910.

Fogel, Robert W. 1964. *Railways and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.

Frankel, Jeffrey A., and David H. Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3): 379–99.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*. London: John Murray.

Rodrik, Dani. 1997. *Has Globalization Gone Too Far?* New York: Columbia University Press.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 209–222

Retrospectives
Adam Smith's Discovery of Trade Gravity

Bruce Elmslie

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact Joseph Persky, Professor of Economics, University of Illinois, Chicago, at jpersky@uic.edu.

Introduction

The "gravity" equation as applied to the determinants of the magnitude of international trade flows between countries has been called "one of the most empirically successful in economics. It relates bilateral trade flows to GDP, distance, and other factors that affect trade barriers" (Anderson and van Wincoop 2003, p. 170). The first full development of the trade equation is credited to Tinbergen (1962), although a similar probabilistic model was also developed by Savage and Deutsh (1960), as discussed in Head and Mayer (2014). Before these works, trade flows were empirically linked to distance by Isard and Peck (1954), who grounded their work to location theory as developed by Weber (1911), Lösch (1944), and others. These precursors are of some interest in the history of the gravity equation, but it is clear that the full empirical linking of trade volume to GDP, distance, and trade barriers came from Tinbergen.

■ *Bruce Elmslie is Professor of Economics, Paul College of Business and Economics, University of New Hampshire, Durham, New Hampshire. His email address is bte@unh.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.209

doi=10.1257/jep.32.2.209

This paper argues that the potential relationship between volume of trade and distance, along with deviation from expectations due to trade barriers, were actually first conjoined by Adam Smith. Smith's context was his critique of the mercantilist balance-of-trade doctrine and his call for increased trade with France. However, Smith's development of the relationship is important for three further reasons. First, the modern gravity equation was an empirical account of trade, but Smith's was based on theory. The modern gravity equation was drawn from simple observation of trade flows from its initial statement in 1962 to Anderson (1979), and more generally to Anderson and Van Wincoop (2003). Its lack of theoretical foundation was a major weakness that resulted in its limited acceptance as a core element of trade theory by top theorists (Head and Mayer 2014). For example, Anderson (2011, p. 134) called the early gravity model "an intellectual orphan, unconnected to the rich family of economic theory." Smith developed the key elements of gravity theory not from empirical observation, but by refining and applying David Hume's (1758) theory of trade and trade gains, along with his own theory of the differential productivity of capital employed in different sectors of the economy. The second important point is that once Smith laid out the basic model, he elaborated a theory of trade restrictions based on gravity considerations and the illogic of mercantile jealousy over the wealth of one's neighbors. Finally, Smith used this same framework to account for the relative "width" of national borders, which is the finding in the trade literature that national borders seem to reduce trade flows by more than would be expected based on distance or other observable variables (Anderson and van Wincoop 2003; Obstfeld and Rogoff 2001).

Smith's Gravity

The passage in which Adam Smith relates bilateral trade volume to each economy's size and distance from one another is buried in the far reaches of the *Wealth of Nations*, in Book IV, Chapter III, Part II, with the title of Part II being "Of the Unreasonableness of Those Extraordinary Restraints upon Other Principles." Here we find some of Smith's most memorable and penetrating critiques of the mercantile system's balance-of-trade doctrine, which he described as "absurd." What Smith (1776) found problematic in the doctrine was not just that it lacked sound theoretical foundations, but that it served as the linchpin for the more abhorrent aspects of mercantilism's "restraints . . . [and] almost all the other regulations of commerce" (p. 615).[1] And, generally, it is through the doctrine of the balance of trade that "nations have been taught that their interest consisted in beggaring all their neighbours" (p. 621).

Smith made his gravity statement within the context of comparing the potential gains from trade with the damage done by trade restrictions, using the commercial

[1] A few pages later, Smith (1776, p. 622) credits "the spirit of monopoly" for originally inventing and propagating the doctrine.

relations between France and Great Britain as his example. In reading the quotation below, it is important to understand that Smith linked the gains from trade with the volume of trade flows between countries. Specifically, in discussing the efficiency gains due to international extensions of the division of labor, Smith (1776) stated that the "mutually afforded [gains] will be greater or smaller in proportion to the extent of their [each countries' bilateral] dealings" (p. 616).

Given this strong relation, Smith (1776, pp. 624–25; emphases added) compares the trade that could take place between England and France if his system of natural liberty prevailed versus the forced, policy-driven trade between England and the North American colonies, and between France and its colonies:

> [T]he commerce of France might be more advantageous to Great Britain than that of any other country, and for the same reason that of Great Britain to France. *France is the nearest neighbor to Great Britain.* In the trade between the southern coast of England and the northern and north-western coasts of France, the returns might be expected, in the same manner as in the inland trade, four, five, or six times in the year. The capital, therefore, employed in this trade, could in each of the two countries keep in motion four, five, or six times the quantity of industry, and afford employment and subsistence to four, five, or six times the number of people, which an equal capital could do in the greater part of the other branches of foreign trade. . . . It would be, at least, three times more advantageous, than the boasted trade with our North American colonies . . . *France, besides, is supposed to contain twenty-four millions of inhabitants.* Our North American colonies were never supposed to contain more than three millions: *And France is a much richer country than North America*; . . . France therefore could afford a market at least eight times more extensive, and, on account of the superior frequency of the returns, four-and-twenty times more advantageous, than that which our North American colonies ever afforded. *The trade of Great Britain would be just as advantageous to France, and, in proportion to the wealth, population and proximity of the respective countries*, . . .[2]

Thus, Smith holds that if the trade volume between two countries is determined by each country's consideration of "their real interest, without either mercantile jealousy or national animosity" (p. 624), it will be in relation to the size of the

[2] In using the term "wealth," Smith sometimes uses it in per capita terms as in the passage above, and sometimes in aggregate terms as in describing the "real wealth" as "the annual produce of the land and labour of the society" (Smith 1776, p. 5). For another example of a per capita usage, Smith writes on the first page of his introduction in the *Wealth of Nations*: "Accordingly therefore, as this produce [overall output], or what is purchased with it [through international trade], bears a greater or smaller proportion to the number of those who are to consume it, the nation will be better or worse supplied with all the necessaries and conveniences for which it has occasion." Hence Smith's perspective is that wealth times population is national produce. He also did not distinguish clearly between annual produce and the value of annual produce (for example, p. 873).

national produce of each country and the distance or proximity between them. Smith also draws connections between distance or "proximity" and quantity of trade at a number of other points in the *Wealth of Nations*, although the connection is not as clear-cut. For example, in a passage discussing the benefits from England's allowing a "most perfect freedom of trade" between the American and West Indian colonies, Smith stated that the "colonies are now become so populous and thriving, that each of them finds in some of the others a great and extensive market for every part of its produce. All of them taken together, they make a great internal market for the produce of one another" (Smith 1776, p. 737). Here Smith uses the terms "populous" and "thriving" and leaves the impact of distance implicit in his use of the term "internal market," but the same basic relation of the effects of wealth, population, and distance seems apparent. Another gravity connection, combining distance and wealth, is found in Smith's discussion of the impact on England and Europe from the expansion of East Indies trade due to the discovery of the shipping route around the Cape of Good Hope (pp. 563–64).

How Smith Arrived at the Determinants of the Extent of Trade

The modern gravity equation of trade was developed as a purely empirical exercise, without theoretical foundations. Isard and Peck (1954), for example, found a strong negative relation between ocean-going freight (measured in tons) and shipping distance. They develop an opportunity cost analysis based on a simple model from Graham (1932) to make the prediction that "transport costs vary systematically with distance" (Isard and Peck 1954, p. 105). But they also lamented the lack of any current theory of the relationship connected to traditional trade theory: "[W]e have a complex of international trade doctrines which in some directions have been pushed to extreme refinement whilst in others [a trade model related to distance] left in a primitive stage of development" (pp. 104–05).[3] Although Tinbergen (1962) does not discuss how he developed the equation, it seems clear from his text that he was a student of actual trade flows.

In contrast, Adam Smith's discussion of distance and trade was developed in the context that actual trade between Great Britain and France was being carried out almost entirely by smugglers. He wrote that extensive "mutual restraints have put an end to almost all fair commerce between the two nations [because of] national prejudice and animosity" (1776, pp. 595–96).[4] Thus, Smith's arguments about relationships between distance, "mass" or national income, and volume of trade were derived as an application of more general theories.

[3]A similar critique of classical trade theory as developed from David Ricardo to Frank Taussig was famously formulated by Williams (1929).

[4]From 1771 to 1775, the average annual value of legal trade between England and France was 236,000 pounds, compared with 2,396,000 pounds between England and its North American colonies (Schumpeter 1960).

First, consider the effect of distance. In the long quotation above, Smith offers an argument that may sound unfamiliar to modern ears: specifically, he links the proximity of countries to the number of times capital can be turned over per year. This is an application of Smith's general arguments on the relative productivity of capital employed in different sectors such as agriculture, manufactures, intracountry trade, and international trade, which are found in Books II and III of the *Wealth of Nations*. Smith finds that the same quantity of capital employs the most productive labor in agriculture and the least in international trade.

Comparing the productivity of capital in trade at home and in international trade, he makes the argument in terms of the number of times a given quantity of capital can turn over per year. For Smith, the employment of productive labor is proportionate to the quantity of capital. Thus, a capital that turns over more quickly will result in a more extended division of labor and more employment of productive labor; and, for Smith, "the whole annual produce [is] the effect of productive labour" (p. 424). For example, Smith (1776) writes:

> … [a] capital … employed in the home-trade will sometimes make twelve operations, or be sent out and returned twelve times, before a capital employed in the foreign trade of equal consumption has made one. If the capitals are equal, therefore, the one will give four and twenty times more encouragement and support to the industry of the country than the other (pp. 469–70).

In general, "the capital … employed in the home-trade of any country will generally give encouragement and support to a greater quantity of productive labour in that country, and increase the value of its annual produce more than an equal capital employed in the foreign trade …" (p. 473).

Later, Smith applied this idea directly to the differential benefits of trade based on the distance between trading partners. He argues that "the quantity of productive labour which any capital employed in the foreign trade of consumption can maintain, is exactly in proportion … to the frequency of its returns. … A foreign trade of consumption carried on with a neighbouring [country], is, upon this account, in general, more advantageous than one carried on with a distant country" (p. 763).

The argument that domestic investment is more advantageous to the country is not necessarily the same as arguing that profit-interested capitalists will invest according to what is best for the country. As Smith (1776) states, "consideration of his own private profit, is the sole motive which determines the owner of any capital to employ it … and the different values which it may add to the annual produce of the land and labour of the society, according as it is employed in one or other of those different ways, never enter into his thoughts" (pp. 476–77). Smith contends, however, that the home trade will also be preferred by the capitalist as long as it is possible to achieve about the same overall rate of profit from home and foreign trade.

> [E]very individual endeavours to employ his capital as near home as he can … Thus, upon equal or nearly equal profits, every wholesale merchant naturally prefers the home-trade to the foreign trade …In the home-trade his capital is never so long out of his sight as it frequently is in the foreign trade … (p. 570).

Thus, Smith offers a reasonably complete analysis of the role played by distance in international trade. Trade with more proximate countries will result in greater gains for the country because it will employ greater numbers of productive laborers, but the self-interests of the capitalist will also result in the actual extent of trade between countries being proportionate to proximity. In a trading system based on natural liberty, the "mercantile stock of every country naturally courts … the near, and shuns the distant employment … It naturally courts the employment which is in ordinary cases most advantageous, and shuns that which in ordinary cases is least advantageous to that country" (1776, p. 798).[5] In contrast, by turning "a foreign trade of consumption with a neighbouring [country], into one with a more distant country," Britain's combination of a near-prohibition of trade with France and a monopoly on colonial trade turned trade "from a direction in which it would have maintained a greater quantity of productive labour, into one, in which it can maintain a much smaller quantity" (p. 771).

Next consider the effect of the mass component of the gravity equation, or what Smith calls "the real value of the annual produce of its [a country's] land and labour" (p. 873).[6] For this "real value," Smith's components are "wealth" and population. In a system of natural liberty, Great Britain will trade more with France than its colonies, and France will trade more with Great Britain than with its colonies, simply because the economies of the two countries have greater mass. Again, this idea was not based on empirical observation, given that only trade occurring between Great Britain and France at the time of Smith's writing was illegal—as Smith wrote, "smugglers are now the principal importers" (p. 595). Instead, Smith was applying Hume's (1758) theory that the increase of the wealth of one nation spills over onto other nations as the increased wealth of one makes their residents better customers for the other. Specifically, Hume states, "man can scarcely be industrious, where all his fellow-citizens are idle … They consume the produce of my industry, and afford me the produce of theirs in return" (p. 329). Similarly, Hume wrote that "the encrease of riches and commerce in any one nation, instead

[5]In making statements like these, Smith is implying an opportunity cost from such trade due to constrained resources such as capital. Elsewhere he makes this opportunity cost argument explicit when he argues that the "mercantile capital of Great Britain, though very great, yet not being infinite… [and] not being increased in the same proportion as the colony trade, that trade could not possibly be carried on without withdrawing some part of that capital from other branches of trade, nor consequently without some decay of those other branches" (1776, p. 758). The other branches he is referring to here are the more beneficial ones, "particularly of that to other parts of Europe."

[6]At times, Smith uses the term "mass" when abstracting away from individual items such as goods or money, as in "[t]he mass of commodities annually thrown into the great circle of European commerce" (1776, p. 751).

of hurting, commonly promotes the riches and commerce of all its neighbours; and that a state can scarcely carry its trade and industry very far, where all the surrounding states are buried in ignorance, sloth, and barbarism" (p. 328). In completing the thought, Hume wrote that as a loyal "BRITISH subject, I pray for the flourishing commerce of GERMANY, SPAIN, ITALY, and even FRANCE itself" (p. 331; original emphasis).

Notice that Hume, being highly influenced by a famous rich country–poor country debate with Josiah Tucker that dealt with convergence issues and the possibility for continuous economic growth, wrote in terms of "the increase of riches" of neighboring countries (as discussed in Elmslie 1995). Smith (1776) turns this growth argument into an analysis based on the size of economies, rather than their growth rates:

> The wealth of a neighbouring nation … is certainly advantageous in trade. … [I]t must likewise enable them to exchange with us to a greater value, and to afford a better market, either for the immediate produce of our own industry, or for whatever is purchased with that produce. As a rich man is likely to be a better customer to the industrious people in his neighbourhood, than a poor, so is likewise a rich nation. … The same maxims which would in this manner direct the common sense of one, or ten, or twenty individuals, should regulate the judgment of one, or ten, or twenty millions, and should make a whole nation regard the riches of its neighbours, as a probable cause and occasion for itself to acquire riches. A nation that would enrich itself by foreign trade, is certainly most likely to do so when its neighbours are all rich, industrious, and commercial nations (pp. 622–23).

In a system of natural liberty, trade between large and wealthy countries will be more extensive than between small and poor countries. As Smith states elsewhere in the book, imports and exports make up the "two distinct benefits" from trade (p. 561).[7] A richer neighbor generates a larger benefit to any nation because it generates more of both.[8] To bring this point home, Smith comes close to quoting Hume's "ignorance, sloth, and barbarism" phrase quoted earlier when he states that "rich and civilized nations can always exchange to a much greater value with one another, than with savages and barbarians" (1776, p. 564).

[7] The exact interpretation of Smith's "two distinct benefits" has been the subject of tremendous controversy and debate among historians of economic thought. However, Schumacher (2015) has recently made a strong and convincing case that Smith meant them to be the benefits of both exporting and importing, rather than the mercantilist idea of gains arising from exporting alone.

[8] This idea is also spelled out in an early draft of the *Wealth of Nations* where Smith compares the trade benefits expected from Great Britain's two closest neighbors, France and Portugal. Smith expected that the major share of the gains would come from France owing to "its superior opulence having more to give, would take more from us, and exchanging to a much greater value and in a much greater variety of ways, would encourage more industry in Great Britain … it is only passion and national prejudice which ever made any body think otherwise" (Smith *Lectures on Jurisprudence* 1982, p, 578).

In creating his version of what later became the gravity equation of trade, Smith applied two aspects of his overall theory of the determinants of the wealth of nations. First, he employed his theory of the differing productivity of capital along with his "invisible hand" notion that self-interested behavior leads to social benefits, to create the connection in which the self-interest of merchants leads to more national trade with neighbors than with distant countries and to greater benefits from doing so. Second, Smith drew on Hume's argument (influenced by Tucker) that the volume or "value" of trade will be positively related to the size of the markets of one's neighboring states to complete the argument that trade is positively related "in proportion" to mass and proximity. There is nothing *ad hoc* in Smith's development of this relation.

A Gravity Theory of Trade Restrictions

Tinbergen (1962, p. 262), in describing his original methodology for a gravity analysis of trade flows, stated that he assumed that the free trade pattern "coincides with the 'average' pattern actually prevailing; this means that we assume the impediments to be of a stochastic nature." Smith used his gravity logic to reach a different conclusion; namely, that the same forces that create the most natural trade partners also create the greatest mercantile interests against this trade. Thus, Smith presumed that the extent of trade barriers and restrictions would be proportional to a competitor country's size and proximity. As Smith (1776, p. 625) stated, "the very same circumstances which would have rendered an open and free commerce between the two countries [France and Great Britain] so advantageous to both, have occasioned the principle obstructions to that commerce."

In selecting Great Britain and France as his example for extolling the potential benefits of trade, Smith (1776) must have known that his argument would face some pushback, because the relationship between these two countries had been scarred by centuries of wars: in particular, Smith mentions the "four expensive French wars of 1688, 1702, 1742, and 1756…" (p. 440). However, Smith also seemed to realize that he had picked a particularly appropriate example, because in calling for extreme restrictions in trade with such a natural partner that "confounded the common sense of mankind," the "national animosity" needed to be "violently inflamed" in order to succeed when such claims were in such opposition to the interests of "the great body of the people" (p. 622). Merchants feared the tremendous competition that would come from free trade with wealthy and proximate countries, so they used this "national animosity" to their advantage to create an impression in the public's mind that was the opposite of reality. Free and open exchange "ought naturally to be, among nations, as among individuals, a bond of union and friendship [but through 'absurd' arguments of the mercantilists], has become the most fertile source of discord and animosity" (p. 621).

This argument implied that where Great Britain could find its greatest gains, through trade with nearby France, mercantile interests would make the loudest calls for protection. Smith (1776) wrote:

> Being neighbours, they [Great Britain and France] are necessarily enemies, and the wealth and power of each becomes … more formidable to the other; and what would increase the advantage of national friendship, serves only to inflame the violence of national animosity. They are both rich and industrious nations; and the merchants and manufacturers of each, dread the competition of the skill and activity of those of the other. Mercantile jealousy is excited, and both inflames, and is itself inflamed, by the violence of national animosity: And the traders of both countries have announced, with all the passionate confidence of interested falsehood, the certain ruin of each, in consequence of that unfavorable balance of trade, which, they pretend, would be the infallible effect of an unrestrained commerce with the other (p. 625).

The result of these arguments made with "all the passionate confidence of interested falsehood" was a set of restraints between France and Great Britain that allowed for almost no trade between these natural trading partners. In the case of Great Britain, the "duties are equivalent to a prohibition. The French in their turn have … treated our goods and manufacturers just as hardly; …Those mutual restraints have put an end to almost all fair commerce between the two nations …" (p. 595). Where the national gains from trade are the greatest, the calls for protection from vested interests will be the loudest, resulting in the most severe restrictions in trade.[9] Thus, gravity giveth and gravity taketh away.

How Wide Is the Border?

Economic theory suggests that after factors that affect trade costs—distance, trade barriers, language differences, and so on—are controlled for, trade between two areas that crosses a national boundary should be about the same as trade between two areas within a given country (Helliwell and McCallum 1996). However, one empirical finding from the gravity model that has caused much debate has been that political borders seem to have an unexpectedly large impact on the volume of trade between countries. McCallum (1995) found that the US–Canadian border had a dramatic negative impact on trade volume. Specifically, he found a 2,200 percent difference between trade across Canadian provinces and US–Canadian

[9] In *The Theory of Moral Sentiments*, Smith (1759, pp. 229–230) argued directly that national animosity is decreasing in distance: "National prejudices and hatreds seldom extend beyond neighbouring nations. We very weakly and foolishly, perhaps, call the French our national enemies; and they perhaps, as weakly and foolishly, consider us in the same manner. Neither they nor we bear any sort of envy to the prosperity of China or Japan."

trade, after controlling for distance. Such results led to a large empirical literature that attempted to understand the effect of national borders on trade, taking many factors other than distance into account (for a review, see Head and Mayer 2014). The results also led to efforts to specify stronger theoretical foundations for the gravity equation. Anderson and van Wincoop (2003) develop a theoretical structure that allows for multilateral trade effects and differences in country sizes to influence the empirical impact of borders. After controlling for these effects, they find that the US–Canadian border has a sizable but much smaller impact of 44 percent on international trade between these countries versus intranational trade. In tests with other industrialized countries, they estimate the border effects to be in the range of 29 percent.

The findings of Anderson and Van Wincoop (2003) would likely be unsurprising to Adam Smith. He thought that, even after controlling for distance and trade barriers, borders would still have a negative effect on trade between advanced countries. Moreover, Smith also made an argument that this border effect is dependent on the relative level of development of the countries. For advanced countries such as Great Britain, the border effects would be stronger in trade with the North American colonies or the West Indies than with the more advanced countries of Europe.

Smith makes several arguments accounting for the "width" of the borders, but they all relate to the same explanation: capital employed at home is more secure than capital employed abroad, and capitalists prefer more security to less. In making his arguments, Smith (1776) employs a gravity metaphor: "Home is in this manner the center … round which the capitals of the inhabitants of every country are continually circulating, and towards which they are always tending, though by particular causes they may sometimes be driven off and repelled from it towards more distant employments" (p. 571).

For example, capital in foreign trade involves greater risk, not just because of the risk of how bad weather can affect shipping, but also because of the difficulties of long-distance contract enforcement in a situation where the laws are different and where the investor often cannot personally assess either the actual situation that has occurred in a faraway land or the character of other parties involved.[10] For example, Smith (1776) wrote:

> The man who employs his capital in land, has it more under his view and command, and his fortune is much less liable to accidents, than that of the trader, who is obliged frequently to commit it, not only to the winds and the waves, but to the more uncertain elements of human folly and injustice, by giving credits in distant countries to men, with whose character and situation

[10] To give an idea of the differential risks associated with trade in the Americas versus Europe, we can observe insurance rates by location. From 1768 to 1770 for summer shipments originating in London, the rates per £100 to Amsterdam and Bilbao, Spain, were 20 and 25 Shillings respectively. For North American shipments the rate was 42 Shillings, while Jamaican shipments were 50 Shillings (John 1958).

> he can seldom be thoroughly acquainted (p. 483). … In the home-trade his capital is never so long out of his sight as it frequently is in the foreign trade of consumption. He can know better the character and situation of the persons whom he trusts, and if he should happen to be deceived, he knows better the laws of the country from which he must seek redress (p. 570).

Another issue is that lesser-developed countries face the "peculiar circumstances" that their capital is "always understocked" (p. 764). As a result, capital-starved traders in these countries will hold it to use for themselves for as long as they can. With regard to the North American colonists, Smith wrote:

> They have a constant demand … for more capital than they have of their own; … The most common way in which the colonists contract this debt… [is] by running as much in arrear to their correspondents, who supply them with goods from Europe, as those correspondents will allow them. Their annual returns frequently do not amount to more than a third, and sometimes not to so great a proportion of what they owe. The whole capital, therefore, which their correspondents advance to them is seldom returned to Britain in less than three, and sometimes not in less than four or five years (p. 764).

These slow returns are faster and more regular than those from the West Indies, but in either case the returns of both are "not only more distant, but more irregular, and more uncertain too, than those of the trade to any part of Europe, or even of the countries which lie round the Mediterranean sea …" (Smith, p. 765). For this reason, the size of the border effect is positively related to the differences in the levels of development of the countries involved.

Conclusion

Adam Smith looked at what appeared to be the fairly sizeable foreign trade that existed between Great Britain and its North American colonies, but what he perceived was the much larger trade that could have been happening between Great Britain and France. He sought to look beyond the "undiscerning eye of giddy ambition … [and the] confused scramble of politics and war" (1776, p. 797). Smith developed his own version of what would come to be known as the gravity theory of trade, based on ideas of proximity and mass, to give an account of the volume or extent of trade that could occur between nations in the absence of restrictive trade practices. Nearly 200 years later, Tinbergen's motivations mirrored those of Smith: "The purpose of the present analysis is to determine the normal or standard *pattern of international trade that would prevail in the absence of discriminating trade impediments*" (Tinbergen, 1962, p. 262; emphasis added).

Adam Smith has traditionally not received much credit for his understanding of the determination of trade flows or the gains from trade. David Ricardo (1817,

p. 295) stated that some of Smith's views on the gains from trade were "at variance with all his general doctrines"; a similar assessment came from John Stuart Mill (1848, p. 579). C. F. Bastable (1908, p. 168) complained of seeing no "special contribution" in Smith's theories of trade, and Jacob Viner (1937, pp. 108–09) concluded that "all the important elements in Adam Smith's free-trade doctrine had been presented prior to the *Wealth of Nations*." Even in an attempt to resurrect Smith's reputation, Bloomfield (1975, pp. 456–57) concluded that "Smith was not a great trade theorist." While these negative assessments of Smith's contributions to trade theory been called into question more recently (for example, Elmslie 1994; Elmslie and Sedgley 2002; Meoqui 2014; Rassekh 2015; Schumacher 2015), neither critics nor supporters have focused on how Smith developed a theory of natural trade volume that came directly from his overall system.[11] Any attempt to judge Smith as a trade theorist must now include his understanding of trade gravity. Indeed, I can find no other author before Tinbergen in 1962 who has offered as complete a statement of the elements of the gravity theory of trade as did Adam Smith.

Smith did not use the "gravity" terminology explicitly, but it is intriguing that for the determinants of the volume of trade Smith emphasized mass and distance, which is of course similar to Isaac Newton's theory of gravity. Smith gives no direct indication that he had Newton's gravity model in mind, but a connection is not implausible. Newton's work had a significant impact on Smith's methodology in the *Wealth of Nations* (Montes 2008; Hetherington 1983). In an earlier work written prior to 1758, Smith (1795) calls Newton's theory of gravity "the greatest discovery that ever was made by man" (p. 105), and describes the Newtonian mathematical connections between mass, distance, and gravitational force (pp. 103–104). One of the commonalities that the gravity theory of trade shares with Newton's theory of gravitation is the difficulty of perceiving and analyzing what will later seem obvious—once it has been explained.

■ *I thank Jim Anderson, Travis Friedman, Gordon Hanson, Doug Irwin, Andrea Maneschi, Soroush Marouzi, Farhad Rassekh, Reinhard Schumacher, Maya Shatzmiller, Timothy Taylor, and Harald Uhlig for helpful comments and suggestions.*

[11] Schumacher (2016) does discuss Smith's use of the relationship between wealth and distance in his analysis of trade, but this is a rare exception. Schumacher uses the relation to make a specific argument regarding the lack of an international division of labor in Smith's analysis of international trade rather than a general analysis of trade based on gravity.

References

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review* 69(1): 106–16.

Anderson, James E. 2011. "The Gravity Model." *Annual Review of Economics* 3(1): 133–60.

Anderson, James E., and Eric Van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93(1): 170–92.

Bastable, C. F. 1903. *The Theory of International Trade with Some Applications to Economic Policy,* 4th edition. London: Macmillan.

Bloomfield, Arthur. 1975. "Adam Smith and the Theory of International Trade." In *Essays on Adam Smith*, edited by Andrew S. Skinner and Thomas Wilson, 455–81. Oxford: Clarendon Press.

Elmslie, Bruce. 1994. "Positive Feedback Mechanisms in Adam Smith's Theories of International Trade." *European Journal of the History of Economic Thought* 1(2): 253–71.

Elmslie, Bruce. 1995. "Retrospectives: The Convergence Debate between David Hume and Josiah Tucker." *Journal of Economic Perspectives* 9(4): 207–16.

Elmslie, Bruce, and Norman Sedgley. 2002. "Vent for Surplus: A Case of Mistaken Identity." *Southern Economic Journal* 68(3): 712–20.

Graham, Frank D. 1932. "The Theory of International Values." *Quarterly Journal of Economics* 46(4): 581–616.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." Chap. 3 in *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North-Holland.

Helliwell, John F., and John McCallum. 1995. "National Borders Still Matter for Trade." *Policy Options* 13(July–August): 44–8.

Hetherington, Norriss S. 1983. "Isaac Newton's Influence on Adam Smith's Natural Laws in Economics." *Journal of the History of Ideas* 44(3): 497–505.

Hume, David. 1758 [1985]. "Of the Jealousy of Trade." In *Essays: Moral, Political, and Literary*, edited by Eugene Miller. Indianapolis: Liberty Fund.

Isard, Walter, and Merton J. Peck. 1954. "Location Theory and International and Interregional Trade Theory." *Quarterly Journal of Economics* 68(1): 97–114.

John, A. H. 1958. "The London Assurance Company and the Marine Insurance Market of the Eighteenth Century." *Economica* 25(98): 126–41.

Lösch, August. 1944. *Die räumliche Ordnung der Wirtschaft.* Revised edition. Jena G. Fischer Publisher.

McCallum, John. 1995. "National Borders Matter: Canada–U.S. Regional Trade Patterns." *American Economic Review* 85(3): 615–23.

Meoqui, Jorge Morales. 2014. "Reconciling Ricardo's Comparative Advantage with Smith's Productivity Theory." *Economic Thought* 3(2): 21–37.

Mill, John Stuart. 1848 [1987]. *Principles of Political Economy.* Fairfield, NJ: Augustus M. Kelly.

Montes, Leonidas. 2008. "Newton's Real Influence on Adam Smith and Its Context." *Cambridge Journal of Economics* 32(4): 555–76.

Obstfeld, Maurice, and Kenneth Rogoff. 2001. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In *NBER Macroeconomics Annual 2000*, edited by Ben S. Bernanke and Kenneth Rogoff. Cambridge, MA: MIT Press.

Rassekh, Farhad. 2015. "Comparative Advantage in Smith's *Wealth of Nations* and Ricardo's *Principles*: A Brief History of Its Early Development." *History of Economic Ideas* 23(1): 59–75.

Ricardo, David. 1817 [1981]. *On the Principles of Political Economy and Taxation.* Cambridge University Press.

Savage, I. Richard, and Karl W. Deutsch. 1960. "A Statistical Model of the Gross Analysis of Transaction Flows. *Econometrica* 28(3): 551–72.

Schumacher, Reinhard. 2015. "Adam Smith's 'Two Distinct Benefits' from Trade: The Dead End of 'Vent-for-Surplus' Interpretations." *History of Political Economy* 47(4): 577–603.

Schumacher, Reinhard. 2016. "Adam Smith, the Patterns of Foreign Trade and the Division of Labour: A Country as a Jack-of-all-Trades rather than a Specialist." Center for the History of Political Economy Working Paper no. 2016-22.

Schumpeter, Elizabeth Boody. 1960. *English Overseas Trade Statistics 1697–1808.* Oxford: Clarendon Press.

Smith, Adam. 1759 [1982]. *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L Macfie. Indianapolis: Liberty Classics.

Smith, Adam. 1776 [2003]. *An Inquiry into the Nature and Causes of the Wealth of Nations.* New York: Bantam Books.

Smith, Adam. 1795 [1982]. *The Principles Which Lead and Direct Philosophical Enquiries; Illustrated by the History of Astronomy.* In *Essays on Philosophical Subjects*, edited by William Persehouse Delisle Wightman and J. C. Bryce. Indianapolis: Liberty Classics.

Smith, Adam. 1982. *Lectures on Jurisprudence*, edited by R. L. Meek, D. D. Raphael, and P. G. Stein. Indianapolis: Liberty Classics.

Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

Viner, Jacob. 1937 [1975]. *Studies in the Theory of International Trade*. Clifton, NJ: Augustus M. Kelley.

Weber, A. 1911. "Die Standortslehre und die Handelspolitik." *Archiv für Sozialwissenschaft und Sozialpolitik* 32(May): 674–77.

Williams, John H. 1929 "The Theory of International Trade Reconsidered." *Economic Journal* 39(154): 195–209.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 223–230

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

The RAND Corporation has published a book-length literature review, *The Science of Gun Policy: A Critical Synthesis of Research Evidence on the Effects of Gun Policies in the United States*, by a team of 17 researchers led by Andrew R. Morral. Here are some main conclusions: "Conclusion 1. Available evidence supports the conclusion that child-access prevention laws, or safe storage laws, reduce self-inflicted fatal or nonfatal firearm injuries among youth. There is moderate evidence that these laws reduce firearm suicides among youth and limited evidence that the laws reduce total (i.e., firearm and nonfirearm) suicides among youth. … Conclusion 2. Available evidence supports the conclusion that child-access prevention laws, or safe storage laws, reduce unintentional firearm injuries or unintentional firearm deaths

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.32.2.223

doi=10.1257/jep.32.2.223

among children. In addition, there is limited evidence that these laws may reduce unintentional firearm injuries among adults. ... Conclusion 3. There is moderate evidence that background checks reduce firearm suicides and firearm homicides, as well as limited evidence that these policies can reduce overall suicide and violent crime rates. … Conclusion 4. There is moderate evidence that stand-your-ground laws may increase state homicide rates and limited evidence that the laws increase firearm homicides in particular. … Conclusion 5. There is moderate evidence that laws prohibiting the purchase or possession of guns by individuals with some forms of mental illness reduce violent crime, and there is limited evidence that such laws reduce homicides in particular. There is also limited evidence these laws may reduce total suicides and firearm suicides. ... Conclusion 7. There is limited evidence that a minimum age of 21 for purchasing firearms may reduce firearm suicides among youth. Conclusion 8. No studies meeting our inclusion criteria have examined required reporting of lost or stolen firearms, required reporting and recording of firearm sales, or gun-free zones. ... Conclusion 10. Research examining the effects of gun policies on officer-involved shootings, defensive gun use, hunting and recreation, and the gun industry is virtually nonexistent. … Conclusion 11. The lack of data on gun ownership and availability and on guns in legal and illegal markets severely limits the quality of existing research. ... Conclusion 12. Crime and victimization monitoring systems are incomplete and not yet fulfilling their promise of supporting high-quality gun policy research in the areas we investigated." March 2018, https://www.rand.org/pubs/research_reports/RR2088.html. There's also a nice accessible website with a summary of results and links to the more detailed studies https://www.rand.org/research/gun-policy.html.

The International Organization for Migration has issued the *World Migration Report 2018*. "The current global estimate is that there were around 244 million international migrants in the world in 2015, which equates to 3.3 per cent of the global population. … The great majority of people in the world do not migrate across borders; much larger numbers migrate within countries (an estimated 740 million internal migrants in 2009). ... Current data indicate that in 2016 there were 40.3 million internally displaced persons (IDPs) worldwide and 22.5 million refugees. Further, the total number of people estimated to have been displaced globally is the highest on record. ... Migration can generate very large benefits for migrants, their families and countries of origin. The wages that migrants earn abroad can be many multiples of what they could earn doing similar jobs at home. For example, a study conducted in 2009 found that the ratio of wages earned by workers in the United States to wages earned by identical workers (with the same country of birth, years of schooling, age and sex, and rural/urban residence) abroad ranges from 15.45 (for workers born in Yemen) to 1.99 (workers born in the Dominican Republic), with a median ratio of 4.11. … In addition to benefiting individual migrants and their families, there is a large research literature that evidences the wider beneficial effects that emigration can have for migrants' countries of origin. ... Globally, remittances are now more than three times the amount of official development assistance. Migration can also result in the transfer of skills, knowledge and technology—effects that

are hard to measure, but that could have considerable positive impacts on productivity and economic growth. … There is widespread agreement that migration can also generate economic and other benefits for destination countries. The precise nature and size of these benefits at a given time critically depends on the extent to which the skills of migrants are complementary to those of domestic workers, as well as on the characteristics of the host economy." https://publications.iom.int/system/files/pdf/wmr_2018_en.pdf.

Richard Damania, Sébastien Desbureaux, Marie Hyland, Asif Islam, Scott Moore, Aude-Sophie Rodella, Jason Russ, and Esha Zaveri discuss *Uncharted Waters: The New Economics of Water Scarcity and Variability*. "The future will be thirsty and uncertain. … Projections suggest that by 2050, global demand for water will increase by 30–50 percent, driven by population growth, rising consumption, urbanization, and energy needs. At the same time, water supplies are limited and under stress from negligent management, growing pollution, degraded watersheds, and climate change. As many as 4 billion people already live in regions that experience severe water stress for at least part of the year. ... Water stress is emerging as a growing and at times underappreciated challenge in many countries of the developed and developing worlds. One in four cities, with a total of US$4.2 trillion in economic activity, is classified as water-stressed. Moreover, 150 million people live in cities with perennial water shortages, defined as having less than 100 liters per person per day of sustainable surface water or groundwater. In coming years, population growth and continuing urbanization will bring a 50–70 percent rise in the demand for water in cities." World Bank, October 2017, https://openknowledge.worldbank.org/bitstream/handle/10986/28096/9781464811791.pdf.

Dan Andrews, Müge Adalet McGowan, and Valentine Millot confirm the worldwide zombie threat in "Confronting the Zombies: Policies for Productivity Revival." "There is growing recognition, however, that the productivity slowdown experienced over the past two decades is partly rooted in a rise of adjustment frictions that rein in the creative destruction process. One important dimension of this phenomenon is that firms that would typically exit or be forced to restructure in a competitive market—i.e. 'zombie' firms—are increasingly lingering in a precarious state to the detriment of aggregate productivity. … Main findings are reported under two main headings. First, the paper provides evidence for the conjecture that weak firms are stifling productivity growth and highlights the considerable scope for raising growth by spurring the orderly exit or restructuring of such firms. Second, it explores the potential for insolvency, financial and other reforms to revive productivity growth by addressing three inter-related sources of structural weakness in labour productivity: the survival of 'zombie' firms, capital misallocation and stalling technological diffusion." OECD Economic Policy Paper #21, December 2017, http://www.oecd-ilibrary.org/economics/confronting-the-zombies_f14fd801-en. For a discussion of zombie firms in China, see "Resolving China's Zombies: Tackling Debt and Raising Productivity" by W. Raphael Lam, Alfred Schipke, Yuyan Tan, and Zhibo Tan. "Nonviable 'zombie' firms have become a key concern in China. ... [T]his paper illustrates the central role of zombies and

their strong linkages with state-owned enterprises (SOEs) in contributing to debt vulnerabilities and low productivity." IMF Working Paper WP/17/266, November 27, 2017, http://www.imf.org/en/Publications/WP/Issues/2017/11/27/Resolving-China-Zombies-Tackling-Debt-and-Raising-Productivity-45432.

Ted Gayer, Robert Litan, and Philip Wallach discuss "Evaluating the Trump Administration's Regulatory Reform Program." "It is fair to say that the Trump administration has launched the most ambitious regulatory budgeting program in human history—just a tremendous undertaking. Whereas Canada and the United Kingdom have managed to get their [regulatory reform] programs up and running with some success thanks to relying on relatively simple metrics of cost, in the United States the regulatory budget will attempt to get much closer to real social costs, at the expense of adding considerable complexity. That makes it potentially more meaningful and deep reaching, but also more likely to bog down and create a massive bureaucratic headache to go with those that already exist … But if all that the Trump administration's regulatory budget turns out to be is an elaborate moratorium on new actions, that would represent a missed opportunity for would-be deregulators. The whole purpose of instituting a forcing mechanism is to confront the problem of accumulated and outdated regulatory requirements that burden U.S. businesses, thereby freeing Americans' energies for productive purposes and unleashing economic growth. If this administration's initiative ends up being nothing more than a pause in further accumulation—of both good and bad prospective regulations—it would stand as a harsh judgment on the likelihood that existing regulation would ever be seriously reformed." Center on Regulation and Markets at Brookings, October 2017, https://www.brookings.edu/wp-content/uploads/2017/10/evaluatingtrumpregreform_gayer-litanwallach_102017.pdf.

Sarah Cohodes reviews the evidence on "Charter Schools and the Achievement Gap." "The best estimates find that attending a charter school has no impact compared to attending a traditional public school. That might surprise you if you were expecting negative or positive impacts based on the political debate around charter schools. But using both lottery-based and observational estimates of charter school effectiveness in samples that include a diverse group of charter schools, the evidence shows, on average, no difference between students who attend a charter and those who attend a traditional public school. However, much of the same research also finds that a subset of charter schools has significant positive impacts on student outcomes. These are typically urban charter schools serving minority and low-income students that use a no excuses curriculum. … Attending an urban, high-quality charter school can have transformative effects on individual students' lives. Three years attending one of these high-performing charter schools produces test-score gains about the size of the black–white test-score gap. … One charter school practice stood out: high-quality tutoring. … As a strategy to close achievement gaps, adopting intensive tutoring beyond the charter sector may be less controversial than focusing explicitly on charter schools." *Future of Children*, Winter 2018, https://future-ofchildren.princeton.edu/resource-links/charter-schools-and-achievement-gap.

Essays on a Theme

Simon Evenett has edited a collection of 15 essays in *Cloth for Wine? The Relevance of Ricardo's Comparative Advantage in the 21st Century*. Evenett writes in the introduction: "Isaac Newton once wrote to a rival that 'If I have seen a little further it is by standing on the shoulders of giants.' Ricardo was such a giant, and it is testament to the enduring value of his insights that they remain the starting point for contemporary analysis of the world trading system. Given the overall goal of this volume is to assess the contemporary relevance of Ricardo's writings on international trade on the 200th anniversary of the publication of his *Principles*, the contributions in this volume have been organised around three themes: our contemporary understanding of Ricardo's insights and the manner in which they have been developed by researchers in recent years; the relevance of Ricardo's analysis in a world trading system far different from one where cloth was exchanged for wine; and the contemporary relevance of Ricardo's policy recommendations as they relate to rejecting protectionism in favour of unilateral free trade." November 2017, Center for Economic Policy Research (CEPR) Press, in association with the UK government Department for International Trade, https://voxeu.org/content/cloth-wine-relevance-ricardo-s-comparative-advantage-21st-century.

The *Russell Sage Foundation Journal for the Social Sciences* has published a double issue on the theme of "Anti-poverty Policy Initiatives for the United States," which includes 15 papers with a wide range of concrete proposals: focused on children in low-income families, the elderly, renters, food stamps, the earned income tax credit, the minimum wage, subsidizing or guaranteeing jobs, postsecondary training and higher education, contraception, and more. As one example, Luke Shaefer, Sophie Collyer, Greg Duncan, Kathryn Edin, Irwin Garfinkel, David Harris, Timothy M. Smeeding, Jane Waldfogel, Christopher Wimer, and Hirokazu Yoshikawa discuss "A Universal Child Allowance: A Plan to Reduce Poverty and Income Instability Among Children in the United States." "Part of the reason that other nations have fewer poor children than the United States is that they provide what the OECD terms a universal child benefit—a cash grant that goes to all families with children. Austria, Canada, Denmark, Finland, France, Germany, Ireland, Luxembourg, the Netherlands, Norway, Sweden, and the UK have all implemented a version of a child benefit. Some call their measures child allowances (CA). Others implement their CA through the tax code as universal child tax credits. A notable feature of these universal child benefit plans is that they are accessible to all: families with children receive them regardless of whether parents work and whatever their income. The level of these child benefits varies by country. The benefit in U.S. dollars for two children in Belgium and Germany is about $5,600 per year; in Ireland $4,000, and in the Netherlands $2,400. Canada has a base child allowance, in U.S. dollars, of roughly $5,000 per child under six and $4,300 per child age six to seventeen ..." February 2018, vol. 4, issue 2–3, February, https://www.rsfjournal.org/loi/rsf.

Daniel McFadden and Kenneth Train have edited an 11-chapter book called *Contingent Valuation of Environmental Goods*. From their introduction:

"Contingent valuation (CV) is a procedure that attempts to estimate the value to households of public goods. While CV can be used in many contexts, we consider its use for evaluating environmental goods. The method is implemented through a survey of households. … This book is born of our concern about the reliability of CV… There seems to be a view that supporting CV is pro-environmental and criticizing CV is anti-environmental. This is a deeply dangerous view. Importantly, results-driven science has an uncanny tendency to circumvent the instigators' intentions. CV can indeed be used to claim large damages against responsible parties (RPs), which seems, in itself, to be a pro-environmental outcome. But CV is used for restoration programs as well as environmental injury, and it gives large benefits for restoration programs. This side of CV provides an incredible boon to RPs, by allowing them to pay off their debts to society at pennies on the dollar. … In benefit–cost analysis, CV tilts the calculations against large environmental improvements. Small measures with relatively little environmental impact (e.g., repairing 15 acres of reef) obtain higher benefit–cost ratios than larger projects with substantial impact (preventing another Gulf spill) because, by CV, the former have about the same benefit as the latter but cost far less. Recognizing CV's unreliability—especially the form it takes—is not just scientifically responsible: it is ecologically responsible." Edward Elgar Publishing, 2017, https://www.elgaronline.com/view/9781786434685.xml. The Fall 2012 issue of this journal included a three-paper symposium on contingent valuation. For an overview of the contingent valuation study done of the BP Deepwater Horizon oil spill in 2010, see Richard C. Bishop and 19 co-authors in "Putting a Value on Injuries to Natural Assets: The BP Oil Spill" in *Science*, April 21, 2017, pp. 253–54, at http://science.sciencemag.org/content/356/6335/253.

Talking with Prominent Economists

Anthony Barnes Atkinson and Nicholas Stern collaborated to produce "Tony Atkinson on Poverty, Inequality, and Public Policy: The Work and Life of a Great Economist." Here are two lively snippets from Atkinson, among many. On how inequality was largely ignored for decades: "[C]learly, since about the early 1990s, I've been trying to get the government and other bodies to restore income distribution to being something that they actually publish data on. You have to remember, in this country—the UK—we dropped the income distribution statistics somewhere in the 1980s. After that, there were none. … The OECD, for example, after putting their toe in the water in the 1970s, didn't return to the subject for another 20 years. So the report that I did with Tim Smeeding and Lee Rainwater in 1995 for OECD (Atkinson et al. 1995) was the first time they'd had a publication on income distribution for 20 years." On understanding what's behind the data: "I think the other thing is that our understanding of data on the more macro side is much inferior to what it was. ... And I came across this when I wrote a review of how government output is measured, because the United States—still, as I understand it—measures government output according to the input. Some US economists say

this is a general policy, but it is not; the European Union, and the UK as part of it, has been using an output-based measure for quite a long time. When we looked at this issue, we discovered that about half the difference in the recorded growth rates between the UK and the US was due to this difference in method." *Annual Review of Economics*, 2017, pp. 1–20, https://www.annualreviews.org/doi/pdf/10.1146/annurev-economics-110216-100949.

The Hutchins Center on Fiscal and Monetary Policy at the Brookings Institution hosted "A Fed Duet: Janet Yellen in Conversation with Ben Bernanke." Here are some comments from Yellen: "So I guess what I do is I often compare the job of managing the [Federal Open Market] committee to the issue a designer would have to face who is trying to decide what's the right color to paint a room. You have 19 people around the table, and you want to come up with a decision we can all live with on what color to paint the room. And we'd go around the table. Ben, what would you like? You think baby blue is just absolutely ideal. David, what do you think? Chartreuse you think is a lovely color. (Laughter) And we go around the room like that. And the question is, are we ever going to converge? I would feel my job is get everybody to see that off-white is not a bad alternative. (Laughter) As brilliant as your choice was, maybe you could live with off-white, and it's not so bad. And we can converge on that and it's going to function just fine and maybe we can agree. So I felt I was often trying to get the committee to coalesce and decide. We'd come up with a good option that we could all agree on." February 27, 2018. Video, audio, and a transcript at https://www.brookings.edu/events/a-fed-duet-janet-yellen-in-conversation-with-ben-bernanke.

Aaron Steelman has an "Interview" with Douglas Irwin. "Did protectionism foster U.S. economic growth and development in the late 19th century? I'm not convinced that we can attribute America's industrial advance in the 19th century to high tariffs or protection. There are a couple points to make on this. ... A lot of the industrialization occurred prior to the Civil War, between 1840 and 1860 when we had low and declining tariffs. ... In addition, there are so many other things going on. We had open immigration, so there was a lot of growth in the labor force. We revamped our banking laws during the Civil War, finance became very important, and we got capital deepening. That's not because of the tariff; that's because the whole financial system of the United States was really developing. Another point to be made is that when you look at the high productivity growth sectors in the U.S. economy in the late 19th century, John Kendrick and others have shown they're mostly in the non-traded goods, service sector. Transportation and utilities were growing very rapidly. It's hard to see how the tariff would help the nontraded goods, service sector of the economy improve its performance. Also, Steve Broadberry has done some work showing that increasing productivity in the service sector was very important to the United States catching up with Britain in the late 19th century. That, too, doesn't seem to be tariff related. All of this doesn't lend itself to an easy story where the tariffs are the key factor behind U.S. growth and industrialization." *Econ Focus*, Federal Reserve Bank of Richmond, Third Quarter 2017, pp. 20–25, https://www.richmondfed.org/-/media/richmondfedorg/publications/research/econ_focus/2017/q3/interview.pdf.

Discussion Starters

The Ellen MacArthur Foundation has published *A New Textiles Economy: Redesigning Fashion's Future*. "[T]he way we design, produce, and use clothes has drawbacks that are becoming increasingly clear. The textiles system operates in an almost completely linear way: large amounts of non-renewable resources are extracted to produce clothes that are often used for only a short time, after which the materials are mostly sent to landfill or incinerated. More than USD 500 billion of value is lost every year due to clothing underutilisation and the lack of recycling. Furthermore, this take-make-dispose model has numerous negative environmental and societal impacts. For instance, total greenhouse gas emissions from textiles production, at 1.2 billion tonnes annually, are more than those of all international flights and maritime shipping combined. Hazardous substances affect the health of both textile workers and wearers of clothes, and they escape into the environment. When washed, some garments release plastic microfibres, of which around half a million tonnes every year contribute to ocean pollution—6 times more than plastic microbeads from cosmetics. Trends point to these negative impacts rising inexorably ..." November 2017, https://www.ellenmacarthurfoundation.org/assets/downloads/publications/A-New-Textiles-Economy_Full-Report.pdf.

The *OECD Review of Fisheries: Policy and Summary Statistics 2017* notes: "Production of wild-caught fish in OECD countries is considerably below its peak in the late 1980s and continues to decline. … Global aquaculture production already exceeds the volume of catch from wild fisheries, if aquatic plants are included. Annual average aquaculture growth in OECD countries has accelerated and now averages 2.1% per year. Globally, it is even more rapid, at 6% per year. Moreover, average prices of aquaculture products are increasing ..." At http://www.oecd-ilibrary.org/agriculture-and-food/oecd-review-of-fisheries-policies-and-summary-statistics-2017_rev_fish_stat_en-2017-en.

Journal of Economic Perspectives—Volume 32, Number 2—Spring 2018—Pages 231–231

Do You Use *JEP* Articles in Your Classroom? One More Chance to Share!

We had a nice response our request, in the previous issue of the *Journal of Economic Perspectives,* that readers send us feedback about their use of *JEP* journal articles in the classroom. If you have *JEP* articles on your syllabus, there is yet one more chance to help us out. One of the goals for *JEP* is to be useful for students and teachers. To facilitate and foster the use of *JEP* articles in the classroom, we want to collect and make available concrete examples of successful use of *JEP* articles on reading lists or in classroom settings. If you are willing to share your experience with *JEP*-related class material, please send an email to Timothy Taylor, Managing Editor of *JEP*, at taylort@macalester.edu.

If you know of colleagues who use *JEP* material in their classes, please help us in spreading the word. This invitation is meant broadly. If you are just using one or a few *JEP* articles in the classroom, and they are working well for you, let us know. If you sometimes assign *JEP* articles to groups of students and then have the students explain the articles to the rest of the class, tell us about it. If you are running a *JEP*-centric class with a substantial proportion of *JEP* articles on the reading list, we definitely want to hear from you. Our main focus is on undergraduate courses, but if you have recommendations at the graduate level, we are glad to hear about those, as well. If time permits, send along a few lines to let us know how long the articles have been on your reading list and what articles are working best for you and your students. Please feel encouraged to attach a copy of your syllabus, too.

Based on the responses, we are hoping to compile a short article for *JEP*, which would list some of the most widely used *JEP* articles for different courses. Along with that article, we could post on the *JEP* website a more detailed description, course by course, of what *JEP* articles are being used successfully.

Thank you for your help.

Enrico Moretti, Editor
Timothy Taylor, Managing Editor

2018 Application/Renewal for Membership

AMERICAN ECONOMIC ASSOCIATION
2014 Broadway, Suite 305
Nashville, TN 37203
Ph. 615-322-2595 fax: 615-343-7590
Federal ID No. 36-2166945
www.vanderbilt.edu/AEA

RENEWING MEMBERS, ENTER ACCT. NUMBER & EXP. DATE	**IF PAYING BY CREDIT CARD, PLEASE FILL OUT BELOW**
ACCOUNT NUMBER:	CARD NUMBER:
EXPIRATION DATE:	EXP DATE: CSC CODE: AMT. PAID:

FIRST NAME:	MI:	LAST NAME:
ADDRESS:		
CITY:		STATE/PROVINCE: ZIP:
COUNTRY:		☐ Check here if non-US
PHONE:		FAX:
PRIMARY FIELD OF SPECIALIZATION:		
SECONDARY FIELD OF SPECIALIZATION:		
EMAIL:		☐ Check here to exclude your email address from the public directory

Please include my email address to receive:
☐ Announcements about public policy affecting economists or the economics profession
☐ Surveys about economics or the economics profession

MEMBERSHIP DUES — Based on annual income. Please select one below.

☐ Annual income of $70,000 or less	$20	$
☐ Annual income of $70,000 to $105,000	$30	$
☐ Annual income over $105,000	$40	$

The AEA dues above include online access to all AEA journals and JSTOR archived issues.
For print or CD subscription(s) indicate preference below and add appropriate charge(s).

Journal		**Print**	**Int'l Postage***	**CD***	
AER	(12 issues, includes the AEA P&P)	☐ Add $25	☐ Add $35	☐ Add $25	$
AEA Papers & Proceedings	**(1 May issue)**	☐ Add $10	n/a	n/a	$
JEL	(4 quarterly issues)	☐ Add $15	☐ Add $15	☐ Add $15	$
JEP	(4 quarterly issues)	☐ Add $15	☐ Add $15	☐ Add $15	$
AEJ: Applied	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Policy	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Macro	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Micro	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$

* Int'l postage applies only to print journals mailed outside of the U.S. No additional postage is required for CDs or the AEA Papers and Proceedings.

				Sub Total	$
Check One:	☐ 1 Year	☐ 2 Years	☐ 3 Years	TOTAL AMOUNT	$

Make checks payable to: American Economic Association.
Must be drawn on a US bank.
Apply online at http://www.aeaweb.org/membership.php

Payments must be made in advance. We accept checks (in US dollars only, with correct coding for processing in US banks) and credit cards; online or by faxing or mailing the application. Please choose one method; it is the Association's policy NOT TO REFUND dues.

AEA's

JOE NETWORK

The Preferred Hiring Tool for the Economics Job Market

Employers	Candidates	Faculty
1,700+ Positions filled	**5,000+** From around the world	**142,000+** Reference letter requests

The JOE Network is designed for economists…by economists.

No other placement program offers a more comprehensive way to match high-caliber candidates with sought after economics positions—including secure management of the faculty reference letter writing process.

JOE Network automates all hiring tasks. Users share materials, communicate confidentially, and easily manage their files and data right from the AEA website.

Compliments of the American Economic Association

Webcasts of Selected Sessions from the 2018 AEA Annual Meeting . . .

Now available on the AEA Website

January 5, 2018

• Climate Policy and Trade
Presiding: *Mar Reguant*
Measuring Leakage Risk *Meredith Fowlie and Mar Reguant*
How Trade-sensitive are Energy-intensive Sectors? *Carolyn Fischer and Alan Fox*
Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Carbon Tariffs *Eva Lyubich, Joseph Shapiro, and Reed Walker*
Discussants: *Matilde Bombardini, Teresa Fort, and Thibault Fally*

• TrumpEconomics: a First Year Evaluation
Presiding: *Dominick Salvatore*
The Effects of Policy Uncertainty *Olivier Blanchard*
The Impact of the Trump Tax Reforms *Martin Feldstein*
What's Lacking in Trumpean Economic Policy *Edmund Phelps*
Trump and Globalization *Joseph E. Stiglitz*
Continuing Relevance of Secular Stagnation *Lawrence H. Summers*
Discussant: *Dominick Salvatore*

• AEA/AFA Joint Luncheon
Presiding: *David Scharfstein*
Raghuram Rajan

• New Measures of the Economy
Presiding: *Betsey Stevenson*
Estimating Changes in Well-being Using Massive Online Choice Experiments *Erik Brynjolfsson, Felix Eggers, and Avinash Gannamaneni*
Measuring Social Progress: Evidence From Advanced Economies *Daniel Fehder, Scott Stern, and Michael E. Porter*
Using Online Prices to Measure Standards of Living Across Countries *Alberto Cavallo, W. Erwin Diewert, Robert Feenstra, Robert Inklaar, and Marcel Timmer*
Internet Rising, Prices Falling: The Era of e-Commerce and its Macroeconomic Implications *Austan Goolsbee and Pete Klenow*
Discussants: *Hal Varian, Betsey Stevenson, Charles Hulten, and Dennis Fixler*

• AEA Richard T. Ely Lecture
Presiding: *David Laibson*

January 6, 2018

• Central Bank Communications
Presiding: *Andrew Haldane*

Future of Monetary Policy Communications *Alan Blinder*

Central Bank Forward Guidance and the Signal Value of Market Prices
Hyun Song Shin and Stephen Morris

Central Bank Communications and the General Public
Michael McMahon and Andrew Haldane

Discussants: *Ricardo Reis, Michael Ehrmann, and Refet Gurkaynak*

• Global Inequality and Policy
Presiding: *Emmanuel Saez*

The Elephant Curve of Global Inequality and Growth
Facundo Alvaredo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman

From Communism to Capitalism: Private Versus Public Property and Inequality in China and Russia
Filip Novokmet, Thomas Piketty, Li Yang, and Gabriel Zucman

Applying Generalized Pareto Curves to Inequality Analysis
Thomas Blanchet, Bertrand Garbinti, Jonathan Goupille-Lebret, and Clara Martinez-Toledano

Extreme Inequality: Evidence From Brazil, India, the Middle-East, and South Africa
Facundo Alvaredo, Lydia Assouad, Lucas Chancel, and Marc Morgan

Discussants: *Thomas Blanchet, Lucas Chancel, Emmanuel Saez, and Li Yang*

• Nobel Laureate Luncheon
Presiding: *Olivier Blanchard*

Daron Acemoglu and Luigi Zingales

• Economic Consequences of Artificial Intelligence and Robotics
Presiding: *Erik Brynjolfsson*

What Can Machines Learn, and What Does It Mean for the Occupations and Industries?
Erik Brynjolfsson, Tom Mitchell, and Daniel Rock

Demographics and Robots
Daron Acemoglu And Pascual Restrepo

Using Similarity And Uncertainty To Predict How Automation Will Effect Occupations
Rob Seamans and Edward W. Felten

Machine Learning by Doing
Joshua Gans, Avi Goldfarb, and Ajay Agrawal

Discussants: *Jason Furman, Ariel Burstein, Susan Helper, and Hal Varian*

• AEA Presidential Address
Presiding: Olivier Blanchard

Alvin E. Roth

2018 AEA Continuing Education webcasts also available

Visit www.aeaweb.org/conference/webcasts/2018

New Journal!

American Economic Review: INSIGHTS

ABOUT THIS JOURNAL

AER: Insights is designed to be a top-tier, general-interest economics journal publishing papers of the same quality and importance as those in the *AER*, but devoted to publishing papers with important insights that can be conveyed succinctly.

The aim of *AER: Insights* is to provide a high-quality outlet for important yet concise contributions to economics, both empirical and theoretical.

CALL FOR PAPERS

CONCISE PAPERS ⇄ QUICKER REVISION PROCESS

Turnaround time for first response is a maximum of three months. First responses will be limited to either Conditional Accept OR Reject with no lengthy responses to referees or second round comments. All requested revisions must be returned to the editor within eight weeks. **Length limits on submissions and revisions will be strictly enforced.**

Open Now! Submissions must be <=6,000 words and have a maximum of five exhibits (figures or tables).

ANTICIPATED LAUNCH 2019

New Journal!

AMERICAN ECONOMIC ASSOCIATION

Full Submission Guidelines
www.aeaweb.org/journals/aeri/submissions

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary *JEP* for your e-reader, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $20.00, $30.00, or $40.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of

Economic Perspectives

Spring 2018, Volume 32, Number 2

Symposia

Does the US Really Gain from Trade?

Risk in Economics and Psychology

Articles

Features

AMERICAN ECONOMIC ASSOCIATION