The Impact of an Epidemic: Experimental Evidence on Preference Stability from Wuhan

By Jason Shachat, Matthew J. Walker and Lijia Wei*

Are preferences stable or are they potentially influenced by formative events and experiences? Questions of this type have troubled economists for decades (Stigler and Becker, 1977). Preference stability underlies fundamental theories of decision-making. For example, Samuelson's (1937) proposition of utility as a discounted sum of rewards is based on the assumption that a rational agent's preferences are stationary.

We harness the exogenous shock of the Covid-19 outbreak to provide controlled evidence on the evolution of deep economic parameters including trust, risk and time preferences. Our study repeatedly applies incentivized decision tasks via the WeChat social media platform to a randomly selected sample of 396 student subjects in Wuhan, China - "ground zero" of the epidemic. The experiments were conducted from late January to early March of 2020. We compare behavioural measurements elicited during this period to those of a pre-epidemic baseline sample of 206 subjects recruited from the same population.

Our results suggest that the initial outbreak, coupled with the lockdown of Wuhan City, undermined trust and tempered the willingness of subjects to seek out unknown situations. The outbreak also led to a fall in risk aversion and there is marginal evidence subjects exhibited more present-biased in the early stages of the crisis. Over the rest of the period, notwithstanding several interesting transitory effects, we observe measurements return to baseline levels with the following exceptions: trust is elevated in March relative to January; risk aversion remains significantly below its baseline level; and subjects become more averse to lying. The results suggest that perturbations in behaviour and preferences from a public health crisis are sharp and can persist for at least several months.

I. Stability of Economic Preferences

Providing robust empirical evidence on the stability of economic preferences requires eliciting controlled behavioural measurements before and after an exogenous shock. By its nature, such a shock must be unanticipated and so it is impossible (certainly unethical) to design such an experiment ex ante. Instead, researchers must rely on natural experiments.

One way scholars have addressed this question is by evaluating the impact of personal economic experiences on preferences. A seminal paper in this literature is Malmendier and Nagel (2011), who use data from between birth cohorts to show that stock market returns experienced are good predictors of risk tolerance.

A separate literature has emerged around the effect of natural disasters - such as floods, earthquakes or hurricanes - on preferences.[1] Since natural disasters are by definition unanticipated, most prior studies of this type rely on measurements taken after the event, between groups with differential exposure.[2] By contrast, behavioural measurements from our control group were elicited six months before the first reports

* Shachat: Durham University Business School, Durham, UK, DH1 3LB; and Economics and Management School, Wuhan University, Wuhan, China, HUB 430072; jason.shachat@durham.ac.uk. Walker: Durham University Business School, Durham, UK, DH1 3LB, matthew.j.walker@durham.ac.uk. Wei: Economics and Management School, Wuhan University, Wuhan, China, HUB 430072, ljwei@whu.edu.cn. All experiments in this study were approved by the Academy of Humanities and Social Sciences, Wuhan University. The authors gratefully acknowledge Natural Science Foundation of China under grant number 71403192, and by the ESRC/FNR under grant number ES/P011829/1.

[1] For a survey, see Chuang and Schechter (2015).

[2] A notable exception is Beine et al. (2020).

of Covid-19 emerged, which offers a natural baseline for comparison.

II. Experimental Context and Design

The experiments in this study were implemented using Ancademy, an online platform for social science experiments based on the open interface of WeChat (https://www.ancademy.org/). This platform facilitates experiments in which participation and payments disbursed through (but not limited) to one's smartphone. In May 2019, we deployed a set of behavioural economics games and preference elicitation tasks to a random sample of 206 student subjects at Wuhan University.[3] The outbreak of Covid-19 at the turn of 2020 enabled us to conduct a fortuitous natural experiment. On January 23rd, the Chinese authorities imposed a strict lockdown of Wuhan city, followed shortly by lockdowns of other cities throughout Hubei province. By this time, many students at Wuhan University had returned home to celebrate the Lunar New Year, the beginning of which coincided with mass quarantine.[4]

Over the following six weeks, we monitored subjects' preferences and behaviour in their place of quarantine. We conducted five separate waves of sampling between late January and early March (Table 1). Data for Wave 1 were collected in the immediate aftermath of the lockdown. Waves 2 and 3 were implemented either side of the well-publicised death on February 7th of Dr. Li Wenliang, a Chinese Ophthalmologist who was widely considered a hero for early warnings of the Covid-19 outbreak (Green, 2020). Waves 4 and 5 followed at two week intervals after this event. In each wave, we recruited a random sample of (approximately) 80 subjects from our subject population. To ensure comparability with the baseline, all subjects completed the experiment using a mobile device.

[3]This was part of a separate project examining how the experimental interface affects economic decision-making.

[4]Media outlets at the time were dubbing this the largest quarantine in human history, see for example, https://apnews.com/7f7336d2ed099936bd59bf8cb7f43756.

The decision-making tasks were designed to capture deep economic parameters of trust and trustworthiness, time, risk and ambiguity preferences, and truth-telling.

To measure trust and trustworthiness, we use first-mover amounts sent and second-mover amounts returned (as a proportion) in a standard Trust game.[5] Subjects are randomly matched into pairs and to roles within a pair. First movers then decide how much of an 8 RMB endowment to send to their match; the amount sent is tripled; and the second mover decides how much of the tripled amount to return. Any money not returned is kept by the second mover.

To elicit time preference, subjects make a series of nine pairwise choices between 100 RMB today or $100(1+i)$ RMB one month later, where the interest rate i increases uniformly from 0 to 0.24. The switching point carries information about a subject's discount rate r, which we use to calculate her annualized rate of patience as $G = (\frac{1}{1+r})^{12}$.[6]

To elicit preference towards risk and ambiguity, we present subjects with a list of nine pairwise choices between a lottery and a sure amount of money. The lottery prizes are fixed across pairs: a chance to win 3 or 9 RMB. For the risk (ambiguity) task, the probability of winning each amount is 50 percent (unknown). The sure amount increases uniformly from 3 to 9 RMB. We use the switching point in each task to calculate risk and ambiguity aversion as an ordinal variable increasing in a subject's aversion.

To measure truth-telling, subjects are asked to privately choose an integer at random from 0-9, add this to the last number of their student ID and record the ones digit. The system then randomly generates a number in the same interval and subjects must report whether the two numbers match. If they match (unobserved by the

[5]The Trust game was only implemented in one half of sessions; in the other sessions, it was replaced with the Ultimatum game (not reported here). This was to prevent behavioural spillover effects between tasks.

[6]Raw switching points are in the Online Appendix, along with a second measure of time preference elicited between two future dates. For all elicitations, those subjects who never switch are assigned a switching point of 10 and we exclude those who switch more than once.

TABLE 1—EXPERIMENTAL WAVES AND DESCRIPTIVE STATISTICS OF THE MAIN OUTCOME VARIABLES.

Experimental wave:	Baseline	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Dates of sampling:	May 2019	Jan 24/26	Feb 4/6	Feb 7/8	Feb 21/22	Mar 6/7
Number of subjects:	206	80	78	80	78	80
Trust	3.39	1.75**	4.95**	3.6	3.66	4.42
$[0, 8]$	(2.59)	(1.80)	(2.48)	(2.44)	(2.73)	(2.59)
Trustworthiness	0.27	0.22	0.33	0.23	0.38	0.34
$[0, 1]$	(0.21)	(0.23)	(0.20)	(0.20)	(0.23)	(0.21)
Patience	0.43	0.39	0.37*	0.46	0.4	0.38
$[0, 1]$	(0.27)	(0.25)	(0.28)	(0.28)	(0.26)	(0.28)
Risk aversion	6.55	6.33*	6.08***	6.74	6.29	6.05***
{1,2,...,10}	(1.13)	(0.86)	(1.47)	(1.18)	(1.42)	(1.58)
Ambiguity aversion	6.51	6.85	7.07***	6.82***	6.59	6.53
{1,2,...,10}	(1.33)	(1.40)	(1.46)	(1.75)	(1.55)	(1.55)
Truth-telling[a]	0.66	0.65	0.62	0.66	0.6	0.51**
{0,1}	(0.47)	(0.48)	(0.49)	(0.48)	(0.49)	(0.50)

Note: Values are mean (standard deviation). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, two-tailed Wilcoxon rank-sum tests of equal means vs. Baseline, except for Truth-telling, which is based on Fisher's exact test. [a] Expected rate of 0.10.

experimenter), subjects' earn 5 RMB, else zero. The extent of lying is inferred from the difference between the reported matching rate and the expected rate of 10%.

Subjects completed all tasks sequentially with no feedback until all tasks were completed. Subjects were paid according to the outcome of each task, earning an average of 65.68 RMB. Payments were made via the WeChat pay facility. If the payoff-relevant date for a task is in the future, payment is received on the future date. Sessions lasted about 45 minutes. No subject participated in more than one session.

III. Results

A. *Trust and Trustworthiness*

As shown in Table 1, first-mover amounts sent in the Trust game are significantly lower in Wave 1 relative to the Baseline ($p = 0.02$). This is driven by a mass of subjects keeping their full endowment. Trust recovers sharply in Wave 2 ($p < 0.01$) and then remains significantly above the Wave 1 level.[7] There is some evidence of a short-term fall in amounts sent between Waves 2 and 3 ($p = 0.098$). There are few systematic differences in second-mover returns after adjusting for the amount received.

To gain individual-level insight, we conduct a Tobit regression of amounts sent on a categorical variable for the experimental waves, a dummy for being in Wuhan during 2020, and demographic control variables. There is evidence of a gender effect: the coefficient estimate on the female dummy is positive and highly significant ($p < 0.01$). The estimate on the Wuhan dummy, however, is negative and more than offsets the gender effect ($p < 0.01$). The fitted values from this analysis are presented in Figure 1, for all subjects (black squares), subjects based in Wuhan (orange circles) and subjects based outside of Wuhan (blue triangles). We observe lower trust among Wuhan-based subjects, although this is based on small sample size.

B. *Time, Risk, Ambiguity and Truth-telling*

The average annualized rate of patience in our Baseline sample is 0.43. This rate decreases to 0.37 in Wave 2 and the difference is marginally significant ($p = 0.09$). Again, there is a fluctuation between Waves 2 and 3, in which subjects temporarily place greater value on the future ($p = 0.09$). Our

[7] Detailed results of all statistical analyses in this section are contained in the Online Appendix.

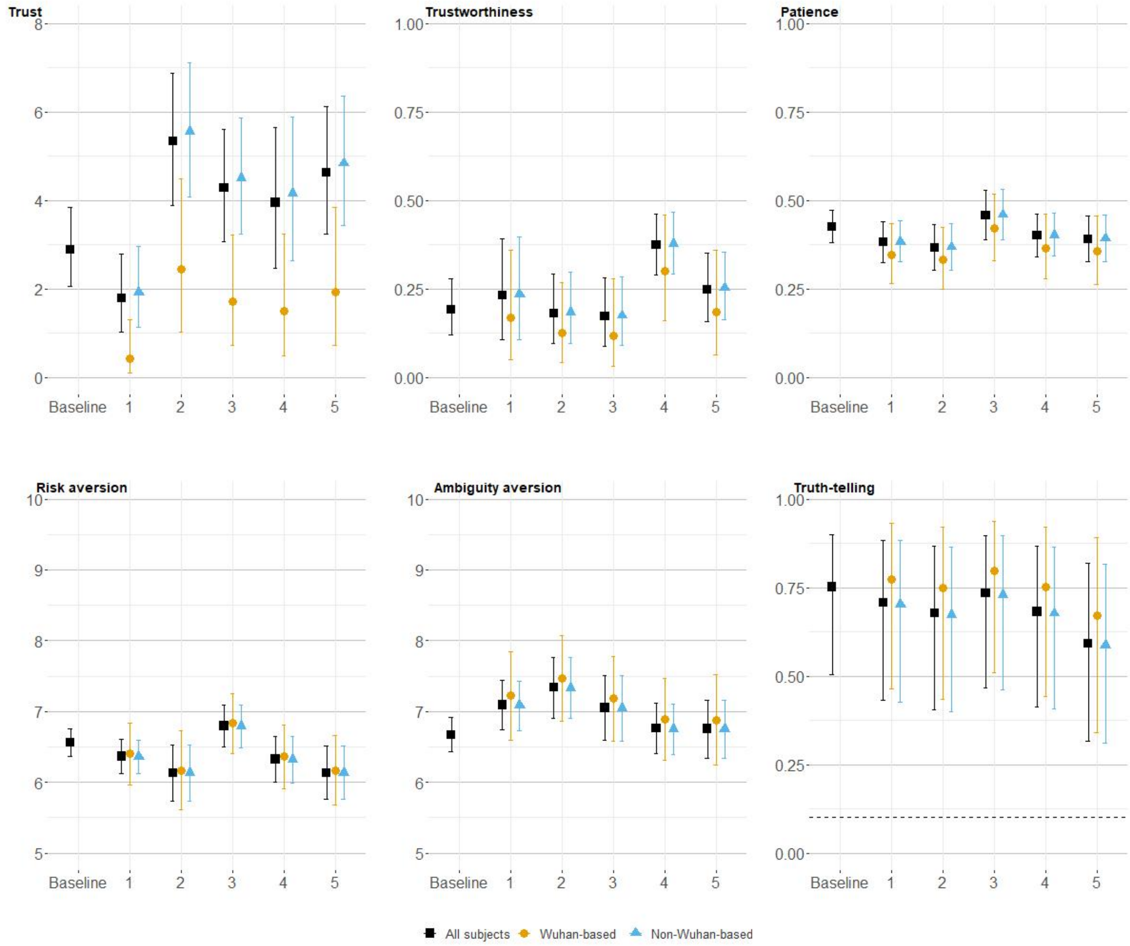

Figure 1. Fitted values of the main outcome variables by experimental wave.

Note: Fitted values are for all subjects (black squares), subjects based in Wuhan (orange circles) and subjects based outside of Wuhan (blue triangles), from censored regressions of the outcome on indicators for the experimental wave relative to the Baseline, and controls. Whiskers indicate 95% confidence intervals, based on robust standard errors. The dashed line in the Truth-telling panel indicates the expected rate of truth-telling given our design.

measure of time preference ends the 2020 sampling period where it began.

Risk aversion is significantly lower in Waves 1 and 2 than in the Baseline (respectively, $p = 0.06$ and $p < 0.01$). The downward trend is interrupted in Wave 3 with a pronounced spike upwards ($p < 0.01$). By Wave 5, risk aversion is back below the Baseline level ($p < 0.01$). Ambiguity aversion is significantly elevated in Waves 1, 2 and 3 (respectively, $p = 0.02$, $p < 0.01$ and $p < 0.01$). This bias does not persist. By Wave 5, ambiguity aversion is significantly below the Wave 2 level ($p = 0.02$). While the distribution of switching points in the Baseline appears to exhibit two modes, the aftermath of the Covid-19 outbreak shifts the lower mode to the right, creating a unimodal distribution.

Across all waves, reported matching rates are significantly above the expected rate given perfect truth-telling. There is no systematic difference in lying propensity between the Baseline and Wave 1 samples ($p = 0.89$). There is evidence that subjects become more averse to lying as the crisis unfolds. In Wave 5, 51% of subjects report a match, significantly below the Baseline frequency of 66% ($p = 0.03$).

Similar behavioural trends can be found in Figure 1 and are robust to controlling for observed covariates. Wuhan-based subjects are more impatient and more averse to risk and ambiguity, although these differences are not significant at the 10% level.

IV. Discussion and Conclusion

The results from the decision-making tasks suggest that preferences are liable to be influenced by formative events and experiences. The Covid-19 outbreak in China appeared to initially undermine subjects' trust, but this was followed by a sustained period of greater trust in others. There is precedent for this: Cassar, Healy and von Kessler (2017) observe a similar phenomenon after the 2004 tsunami in Thailand. That the outbreak induced lower patience is also consistent with prior event studies (Beine et al., 2020; Cassar, Healy and von Kessler, 2017; Voors et al., 2012).

We observe differential responses in the preference domains of risk and ambiguity. Whereas subjects exhibit a greater willingness to take risks after the crisis, they are less willing to seek out unknown situations. Such a divergence has been observed in other settings (Cavatorta and Groom, 2020). Those early days in Wuhan were a time of many unknowns. It is plausible that subjects were cautious in the face of uncertainty (such as taking preventative measures to avoid catching the virus), but after contracting the virus were willing to take significant risks to avoid the worst outcome (death).

Finally, we observe a sharp discontinuity in elicited measures of trust, patience and risk aversion before and after the death of Dr. Li Wenliang, an event that appeared to resonate in the collective unconsciousness of ordinary Chinese citizens.

Differential selection into the Baseline and post-Covid samples is clearly a potential confound in our design, and so we remain cautious in any causal interpretation. Nevertheless, the marked short-term fluctuations between our 2020 waves suggest that the time between measurements is an important factor to consider when making inferences as to the impact of natural events on preferences.

REFERENCES

Beine, Michel, Gary Charness, Arnaud Dupuy, and Majlinda Joxhe. 2020. "Shaking Things Up: On the Stability of Risk and Time Preferences." Working Paper.

Cassar, Alessandra, Andrew Healy, and Carl von Kessler. 2017. "Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand." *World Development*, 94: 90 – 105.

Cavatorta, Elisa, and Ben Groom. 2020. "Does deterrence change preferences? Evidence from a natural experiment." *European Economic Review*, 127: 103456.

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: 151–170.

Green, Andrew. 2020. "Li Wenliang." *The Lancet*, 395(10225): 682.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics*, 126(1): 373–416.

Samuelson, Paul A. 1937. "A Note on Measurement of Utility." *Review of Economic Studies*, 4(2): 155–161.

Stigler, George J., and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum." *American Economic Review*, 67(2): 76–90.

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.