Carrots, Sticks, and Consequences:
Gender Differences in Incentive Provision and Its Evaluation

Alexandra van Geen, Erasmus University
Olga Shurchkov, Wellesley College

American Economic Association Annual Meetings

January 4, 2016

Motivation and contribution

- Provision of appropriate incentives is an important attribute of leadership

- Women represent just 4.4% of CEOs and 19.2% of board members of Fortune 500 companies (Catalyst 2015)

- Can gender differences in incentive choice and the subsequent evaluation of this choice partly explain these gender gaps in leadership status?

- Contibution: incentive regime selection, impact, and evaluation in an experimental setting with a real-effort task

Specifically: 3 questions

1. What type of incentive regime (if any) are decision makers (DMs) more likely to choose to motivate others' productivity?

   ■ Possible regimes: no incentives; "carrots" vs. "sticks;" "competitive" vs. "self-reference"

2. How effective are different types of incentives for increasing productivity?

3. What factors influence the evaluation of the DMs?

   ■ Subquestion: What role (if any) does gender play in answering each of those three questions?

Real-world relevance

Experimental agenda aims to model the following types of real-world interactions:

- Experiment 1 (this talk):

    - an appointed team project leader chooses incentive type knowing that her profit depends on the output of worker colleagues;
    - workers evaluate the project leader;
    - someone else may lead the next project

- Experiment 2 (design stage):

    - third-party reviewer evaluates the project leader;
    - third-party decides whether to "select" the project leader for future decision-making

EXPERIMENTAL DESIGN AND METHODOLOGY

Experiment 1: Sessions and treatments conducted so far

- Erasmus School of Economics Econlab
- 2 pilots to test the experiment (data not included)
- 6 experimental sessions all in 1 week (end of September 2015)
- Average profit including show-up fee: €17

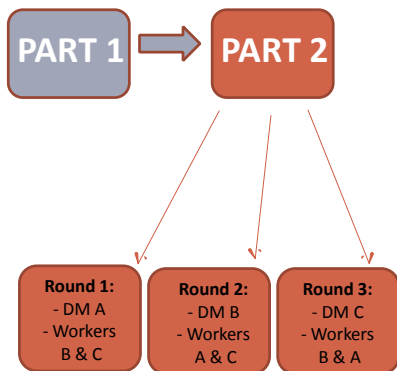| Treatment	# Sessions	#Subjects		
		Male	Female	Total
Gender	3	39	33	72
No gender	3	31	32	63
Total	6	70	65	135

Table 1: Summary of sessions and treatments

Schematic representation of Experiment 1

PART 1

- Questionnaire
  - **Control: NO GENDER**
  - **Treatment: GENDER**
- Baseline addition
  task (<u>sums of 5 2-
  digit numbers</u>)

Schematic representation of Experiment 1

PART 1 → **PART 2**

| Round 1:
- DM A
- Workers
B & C | Round 2:
- DM B
- Workers
A & C | Round 3:
- DM C
- Workers
B & A |

- DM views worker resume (no gender)
- Makes incentive choice that applies to both workers
- DM knows that workers will view the choice
- Workers view DM choice + resume (**Treatment: GENDER**) ➔ **rate DM competency**

Schematic representation of Experiment 1

PART 1 ➡ **PART 2**

Round 1:
- DM A
- Workers
  B & C

Round 2:
- DM B
- Workers
  A & C

Round 3:
- DM C
- Workers
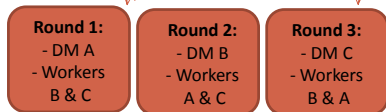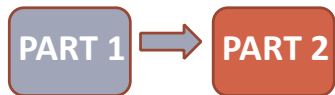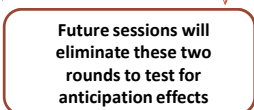 B & A

- DM views worker resume (no gender)
- Makes incentive choice that applies to both workers
- DM knows that workers will view the choice
- Workers view DM choice + resume (**Treatment: GENDER**) ➔ **rate DM competency**

Details of the DM's choice:

- DM is endowed with 100 "DM points"

- DM chooses 1 of the following 5 options:
  1. Worker does not get or lose any points [no incentive]
  2. + 50 points if outperform other worker [carrot; comp]
  3. + 50 points if exceed own past score [carrot; self-ref]
  4. - 50 points if underperform other worker [stick; comp]
  5. - 50 points if fall short of own past score [stick; self-ref]

- Incentive costs:
  - Option 1 costs 0 DM points
  - Any of the other 4 options costs 10 DM points
    (10 points = 50 Euro cents)

Schematic representation of Experiment 1

PART 1 ➡ **PART 2**

Round 1:
- DM A
- Workers B & C

Future sessions will eliminate these two rounds to test for anticipation effects

Details of the DM's choice:

- DM is endowed with 100 "DM points"

- DM chooses 1 of the following 5 options:
  1. Worker does not get or lose any points [no incentive]
  2. + 50 points if outperform other worker [carrot; comp]
  3. + 50 points if exceed own past score [carrot; self-ref]
  4. - 50 points if underperform other worker [stick; comp]
  5. - 50 points if fall short of own past score [stick; self-ref]

- Incentive costs:
  - Option 1 costs 0 DM points
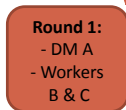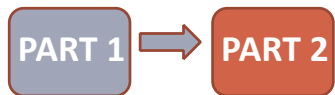 - Any of the other 4 options costs 10 DM points (10 points = 50 Euro cents)
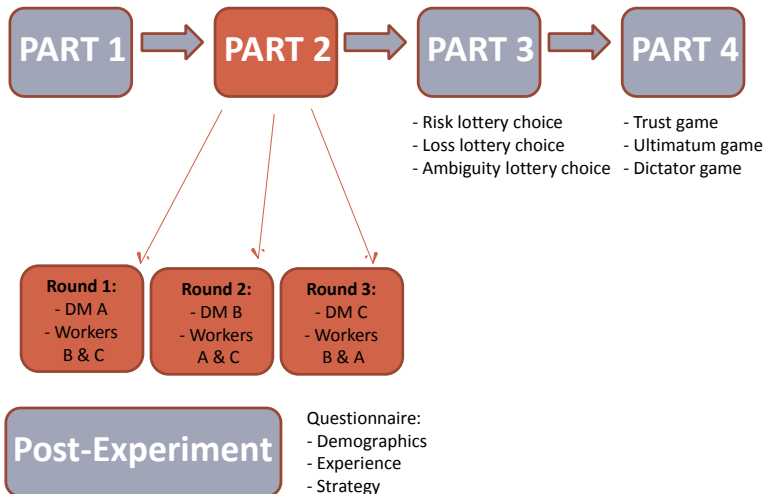
- DM views worker resume (no gender)
- Makes incentive choice that applies to both workers
- DM knows that workers will view the choice
- Workers view DM choice + resume (**Treatment: GENDER**) ➔ **rate DM competency**

Schematic representation of Experiment 1

PRELIMINARY RESULTS FROM THE FIRST BATCH OF SESSIONS OF EXPERIMENT 1

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Descriptive statistics

As compared to women in our sample, men are on average significantly

- more likely to live in Rotterdam, have higher household income, and be white

- more likely to favor transactional rather than transformational leadership style

- more trusting, less risk averse, and less loss averse

▸ Detailed descriptive statistics

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

DECISION-MAKERS: CHOICES AND THEIR DETERMINANTS

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Carrots are overwhelmingly chosen
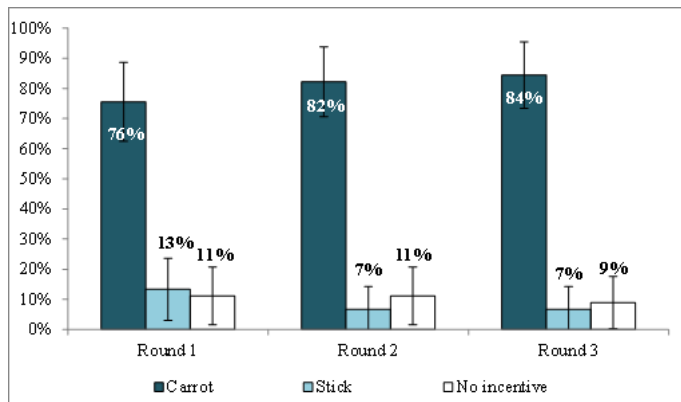
Figure 1: Percentage of DMs choosing different incentive types
by round (95% C.I.)

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Gender and other determinants of incentive choice

- Averaging across rounds and controlling for round and session FE, women are less likely to give positive incentives and more likely to choose no incentive

- However, the gender effect disappears after controlling for observable worker resume and individual DM characteristics
    - DMs who are more trusting & reciprocal and who are less risk & loss averse are more likely to choose carrots
    - DMs who are less trusting, less strategically altruistic, and who are more risk & loss averse are more likely to choose the no incentive option

- However, conditional on choosing a positive incentive, female DMs are 10% less likely to choose the competitive option, even after controlling for characteristics (consistent with Niederle & Vesterlund 2007; Shurchkov 2012)

▸ Determinants of incentive choice

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

DECISION-MAKERS AND WORKERS: IMPACT OF INCENTIVE CHOICES

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Impact of DM's choice & gender on worker effort & score

- Conditional on all worker and DM resume characteristics observable to workers (averaging across all rounds and treatments):

  - carrots increase score by 10%
  - carrots increase effort by 7%

- The knowledge of DM gender matters:

  - when DM gender is revealed, incentive type is not significant but having a female DM increases the score by 10% (effort by 6%);
  - when gender is NOT revealed, score is 13% higher with carrots (effort 9% higher); DM gender is not significant

▸ Estimates of the impact of incentives on worker outcomes   ▸ Do incentives pay off?

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Impact of DM's choice & gender on worker effort & score

- Decomposing the impact by carrots offered by female vs male DMs:

  - When DM gender is revealed:
    - carrots chosen by female DMs increase worker score by 13% relative to stick/no incentive
    - carrots chosen by male DMs do not (gender difference significant)

  - When DM gender is NOT revealed:
    - no significant gender differences in terms of impact of the incentive on worker score or effort
    - carrots increase worker score by around 11% relative to stick/no incentive regardless of DM gender

  - Analogous differential treatment effects on effort

▶ Estimates of the impact of incentives on worker outcomes   ▶ Do incentives pay off?

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

WORKERS: EVALUATION OF THE DECISION-MAKERS

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Women are rated as less competent overall and when choosing carrots

Rating for all rounds with:	Treatment	Male DM	Female DM	Difference	t-test p-value	N
Any incentive (overall)	Gender	2.82	2.48	0.336***	0.006	288
Carrot	Gender	3.06	2.63	0.424***	0.000	240
Non-carrot	Gender	1.20	1.93	-0.729**	0.020	48
Zero inc	Gender	1.38	2.00	-0.625*	0.053	40
Any incentive (overall)	No Gender	2.47	2.30	0.171	0.196	252
Carrot	No Gender	2.65	2.54	0.110	0.417	196
Non-carrot	No Gender	1.50	1.67	-0.167	0.563	56
Zero inc	No Gender	1.00	1.83	-0.83	0.166	16

Table 2: DM rating by incentive choice, treatment, and DM gender

Introduction & Motivation
Experimental Design and Methodology
Preliminary Results
Discussion

Descriptive Statistics
DMs: Choices and Their Determinants
Impact of incentive choices
Evaluation of the DM

Determinants of worker's rating of the DM in the first round

- When DM gender is known, female DMs are rated, on average, 0.65 (out of 5) points lower than male DMs

- When DM gender is NOT known, female and male DM ratings are not statistically different

- DMs choosing sticks are rated as significantly less competent compared to DMs choosing positive incentives

- The overall negative attitude toward sticks seems to stem from low ratings of DMs choosing self-reference type of incentive (caution: very small sample!)

- Finally, female DMs choosing carrots are rated as less competent relative to comparable male DMs also offering positive incentives (but very few obs!)

DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

Summary of results so far

- DMs overwhelmingly use positive incentives (consistent with Baker, Jensen, Murphy 1988; Luft 1994)

- Women and men are equally likely to use carrots or no incentives, conditional on individual characteristics

- However, women are significantly less likely to use competitive incentives (consistent with Niederle & Vesterlund 2007; Shurchkov 2012)

- Incentives significantly improve worker performance (consistent with non-real-effort studies Fehr & Schmidt 2007; Fehr, Klein, & Schmidt 2007); gender of DM matters

- Conditional on the type of incentive and on resume characteristics, female DMs are rated as less competent relative to otherwise comparable male DMs

Increasing internal validity of experiment 1

- The issue of very few observations of sticks and no incentive choices

  - Low incentive costs $\implies$ Very few zero incentive observations
  - Will conduct additional sessions to increase N
  - Future research: Control treatment with random incentive assignment

- The issue of multiple rounds

  - Control for social preferences to control for reciprocity
  - Do not find evidence that previous DM decisions affect DM decisions in future rounds
  - Use round 1 data only (after increasing N)
  - Future sessions with only 1 round; compare to test for anticipation effects

Increasing external validity: experiment 2 and other treatments

- DM evaluation not incentivized; done by workers rather than third-party observer
  - Experiment 2 will address this issue
  - Future research: evaluations by groups of observers/gender composition

- Role of expectations:
  - Explore further whether incentive assignment is "rational" given DM expectations of worker productivity (have these data, need to analyze)
  - Future research: Shed light on the reasons for why women are rated as less competent (Ask worker expectations prior to revealing DM choice)

- Future research: Effect of worker gender on choice of incentive regime

APPENDIX: DETAILED TABLES

Resume Characteristics

	Females		Males		Comparison
Variable	Value	# Obs.	Value	# Obs	(t-test p-val)
	Part 1 (Resume) Characteristics				
Mean resume addition score	9.97	63	9.83	69	0.226
% respondents living in Rotterdam	88%	65	74%	70	0.049**
% respondents studying at Erasmus	97%	65	97%	70	0.941
% respondents in 19-25 age range	87%	65	91%	70	0.507
% respondents in social sciences	94%	65	86%	70	0.123
% respondents in Masters program	43%	65	44%	70	0.889
% respondents with GPA >7.5/10	31%	65	30%	70	0.923

▸ Back to summary statistics overview

Demographics and preferences

Variable	Females		Males		Comparison	
	Value	# Obs.	Value	# Obs	t-test p-val	U-test p-val
Part 2 Decisions						
Demographic Characteristics						
Mean GPA (out of 10)	7.05	64	7.08	67	0.829	0.551
Mean age (years)	21.57	65	21.80	70	0.597	0.882
% citizen of the Netherlands	65%	65	76%	70	0.161	0.160
% White respondents	52%	65	71%	70	0.022**	0.023**
% with fluent or native English proficiency	58%	65	53%	70	0.516	0.514
% with family income >Eu65,000	20%	65	39%	70	0.018**	0.019**
% married	5%	65	6%	70	0.776	0.774
% with previous leadership experience	65%	65	60%	70	0.584	0.582
% with previous experience with games	58%	65	64%	70	0.491	0.489
% transformational leadership	85%	65	70%	70	0.044**	0.044**
% democratic leadership	82%	65	84%	70	0.674	0.673
% affiliative leadership	24%	65	35%	70	0.161	0.160
% commanding/pacesetting leadership	6%	65	13%	70	0.190	0.189
Preferences						
Mean pl. 1 trust transfer (trust) (Eu/10)	3.78	65	5.23	70	0.022**	0.045**
Mean pl. 2 trust transfer (reciprocity) (Eu/10)	4.71	65	5.07	70	0.410	0.351
Mean pl. 1 ultim. transfer (strategic) (Eu/10)	3.92	65	3.99	70	0.847	0.965
Mean dictator transfer (pure) (Eu/10)	1.98	65	2.07	70	0.838	0.578
Mean risk lottery choice (out of 6 lotteries)	2.72	65	3.43	70	0.008***	0.009***
Loss aversion (loss - risk choice)	1.02	65	0.44	70	0.076*	0.035**
Ambiguity aversion (risk - ambiguity choice)	-0.29	65	0.11	70	0.167	0.225

Determinants of incentive regime choice

Outcome variable:	Pr [Carrot]		Pr [Competitive]		Pr [No Incentive]	
	(1)	(2)	(3)	(4)	(5)	(6)
female dummy	-0.094***	-0.04	0.04	-0.096**	0.088***	-0.054
	(0.025)	(0.033)	(0.038)	(0.041)	(0.024)	(0.085)
reciprocity		0.020***		0.059***		(0.007)
		(0.007)		(0.011)		(0.009)
trust		0.020***		-0.008		-0.053***
		(0.002)		(0.008)		(0.013)
pure altruism		-0.005		-0.002		0.008
		(0.006)		(0.012)		(0.010)
strategic altruism		-0.008		0.027		-0.065***
		(0.007)		(0.017)		(0.022)
risk preference		-0.025**		-0.036*		0.027*
		(0.012)		(0.021)		(0.015)
loss aversion		-0.015**		-0.026**		0.048**
		(0.006)		(0.013)		(0.022)
ambiguity aversion		-0.008		0.023		-0.056***
		(0.011)		(0.014)		(0.018)
Worker Resume	NO	YES	NO	YES	NO	YES
DM Characteristics	NO	YES	NO	YES	NO	YES
Round	All Rounds		Carrots Only		All Rounds	
Observations	810	756	654	564	684	504
Pseudo R-squared	0.10	0.41	0.06	0.36	0.08	0.57

Notes: All regressions include round and session fixed effects; worker resume controls observable by the DM: score in Part 1 addition, GPA range, major, student status (Bachelors or Masters), age bracket, university, and city of residence; DM characteristics: race, GPA, income bracket, marital status, english language proficiency, citizenship, and past experience with various games. Sp. 5-6 exclude data from session 4 (no observations of no incentive). Robust standard errors are clustered by group-round (in parentheses). Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

OBSERVATIONS:

- On average, women are less likely to give positive incentives and more likely to choose no incentive (Sp 1 & 5)

- However, the gender effect disappears after controlling for individual characteristics (Sp 2 & 6)

- Sp 3 & 4 estimate the likelihood of choosing a competitive incentive (rather than self-reference), conditional on choosing a carrot. Women are less likely to choose a competitive type of carrot relative to men (consistent with previous lit)

▸ Back to determinants of incentive choice

Impact of DM's choice on worker effort and score

Outcome variable:	Score (number correct solutions)				Total number entered			
Dependent Variable Mean:	[11.93]	[12.49]	[11.93]	[12.49]	[14.34]	[14.80]	[14.34]	[14.80]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DM female	1.231***	-0.242			0.925***	-0.355		
	(0.302)	(0.272)			(0.278)	(0.292)		
carrot dummy	0.309	1.599***			0.585	1.275***		
	(0.484)	(0.336)			(0.443)	(0.342)		
carrot x DM female			1.526**	1.308***			1.372**	1.279**
			(0.649)	(0.365)			(0.624)	(0.559)
carrot x DM male			0.288	1.512***			0.39	1.910***
			(0.668)	(0.508)			(0.617)	(0.622)
stick or no incent x DM female			1.205	-0.361			0.68	0.52
			(0.814)	(0.515)			(0.828)	(0.649)
f-test p-value			**[0.001]**	**[0.639]**			**[0.003]**	**[0.080]**
worker female	-1.394***	1.196***	-1.396***	1.191***	-0.803**	1.102***	-0.815***	1.143***
	(0.392)	(0.380)	(0.399)	(0.388)	(0.336)	(0.415)	(0.339)	(0.412)
worker part 1 addition score	0.744***	0.784***	0.744***	0.784***	0.792***	0.826***	0.792***	0.825***
	(0.068)	(0.057)	(0.068)	(0.057)	(0.063)	(0.054)	(0.064)	(0.054)
Treatment (DM gender shown?)	Yes	No	Yes	No	Yes	No	Yes	No
Round	All	All	All	All	All	All	All	All
Other worker characteristics?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	405	354	405	354	405	354	405	354
R-squared	0.65	0.74	0.65	0.74	0.67	0.78	0.67	0.78

Notes: p-values for f-tests of the difference between male and female carrot choice reported in brackets below the estimates. All regressions include session and round FE. Other worker characteristics include social and risk preferences, self-reported GPA and age, and self-reported indicators for student status (Bachelors or Masters), major, race, city of residence, english language proficiency, and citizenship. Robust standard errors are clustered by group-round (in parentheses). Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

OBSERVATIONS:
• Score and effort are higher with a female DM when gender is revealed (Sp 1,5) and with a carrot when gender is not revealed (Sp 2,6)
• Carrots are more effective when provided by female DMs than when provided by male DMs in the gender treatment (Sp 3 vs 4; Sp 7 vs 8)
• Loss aversion increases effort and therefore score, but only in the no gender treatment (not shown)

DM and Worker Payoffs

- On average, DMs make 9% more than workers

- For DMs, pooling across over all treatments and rounds, the benefits of incentives do not significantly overweigh the costs of providing them. In fact, providing a stick marginally decreases the DM's profit, all else equal.

- For workers, carrots significantly increase payoffs (by 17.5%, on average).

▸ Back to main impact slide

Determinants of Worker's Rating of the DM

Outcome variable:	DM Competency Rating (0 - Not at all competent; 4 - Extremely competent)					
Dependent Variable Mean:	[2.52]	[2.25]	[2.51]	[2.25]	[2.51]	[2.25]
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	Gender	No Gender	Gender	No Gender	Gender	No Gender
DM gender (1 = Female)	-0.344*	-0.151	-0.650**	0.035	2.159**	0.624
	(0.189)	(0.192)	(0.269)	(0.376)	(0.901)	(1.624)
DM Resume Score	0.095***	0.029	0.048	0.116	0.07	0.106
	(0.034)	(0.027)	(0.056)	(0.084)	(0.062)	(0.115)
+ve incentive	1.331***	1.336***	-0.328	0.859	2.273***	1.614
	(0.444)	(0.420)	(0.722)	(1.144)	(0.669)	(1.396)
-ve incentive	-0.852	0.415	-2.545**	-0.044	0.183	1.803
	(0.551)	(0.472)	(1.092)	(1.103)	(0.972)	(2.354)
+ve incentive & female DM					-2.760***	0.113
					(1.030)	(1.757)
-ve incentive & female DM						-2.31
						(2.650)
f-test p-value	[0.000]	[0.000]	[0.023]	[0.077]	[0.000]	[0.501]
DM Resume and Worker Char	NO	NO	YES	YES	YES	YES
Worker Social Preferences	NO	NO	YES	YES	YES	YES
Date FE	NO	NO	YES	YES	YES	YES
Observations	92	80	90	76	90	76
R-squared	0.34	0.30	0.66	0.58	0.66	0.59

Notes: Round 1 data only (avoid reciprocity effects over rounds). p-values for f-tests of joint differences of the coefficients on the DM choice are reported in brackets below the estimates. DM resume controls include: GPA range, age bracket, student status (Bachelors or Masters), major, city of residence, and university. Worker characteristics include: race, GPA, income bracket, marital status, english language proficiency, and citizenship. Robust standard errors in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

‣ Back to main evaluation slide

Determinants of Worker's Rating of the DM

Outcome variable:	DM Competency Rating (0 - Not at all competent; 4 - Extremely competent)			
Dependent Variable Mean:	[2.51]	[2.25]	[2.51]	[2.25]
	(1)	(2)	(3)	(4)
Treatment	Gender	No Gender	Gender	No Gender
DM gender (1 = Female)	-0.790***	0.264	-0.656**	0.164
	(0.279)	(0.547)	(0.285)	(0.792)
DM Resume Score	0.038	0.07	0.069	0.065
	(0.061)	(0.092)	(0.077)	(0.153)
+ve competitive incentive	-0.352	1.606	-0.585	1.621
	(0.855)	(1.171)	(0.974)	(1.226)
-ve competitive incentive		0.558		0.534
		(1.176)		(1.354)
+ve self-ref incentive	0.03	0.624	-0.39	0.655
	(0.858)	(1.146)	(0.819)	(1.185)
-ve self-ref incentive	-3.075***	-0.261	-2.671**	-0.082
	(1.080)	(1.612)	(1.176)	(2.635)
f-test p-value	[0.001]	[0.012]	[0.104]	[0.060]
DM Resume and Worker Char	YES	YES	YES	YES
Worker Social Preferences	NO	NO	YES	YES
Date FE	NO	NO	YES	YES
Observations	90	76	90	76
R-squared	0.61	0.58	0.67	0.61

Notes: Round 1 data only (avoid reciprocity effects over rounds). p-values for f-tests of joint differences of the coefficients on the DM choice are reported in brackets below the estimates. All regressions include DM resume controls: GPA range, age bracket, student status (Bachelors or Masters), major, city of residence, and university. Worker characteristics include race, GPA, income bracket, marital status, english language proficiency, and citizenship. Robust standard errors in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

▸ Back to main evaluation slide

The real effort task: sums of 5 2-digit numbers