

The role of cognitive ability and personality traits for men and women in gift exchange outcomes

Emel Filiz-Ozbay¹ · John C. Ham² ·
John H. Kagel³ · Erkut Y. Ozbay¹

Received: 11 December 2015/Revised: 26 October 2016/Accepted: 1 November 2016
© Economic Science Association 2016

Abstract We examine the role of cognitive ability and personality traits in a gift exchange experiment. Controlling for cognitive ability and personality characteristics, men offer higher wages than women, and men and women with greater cognitive ability and greater agreeableness on the Big Five personality scale offer higher wages as well. Men provide greater effort than women do, and respond to higher wage rates with greater increases in effort. For both genders, one standard deviation increases in agreeableness and in wages generate similar increases in effort. Serious biases arise from omitting cognitive ability and pooling men and women.

Keywords Gift exchange experiment · Big five personality characteristics · SAT scores

JEL classification D03 · J16 · J22

Electronic supplementary material The online version of this article (doi:[10.1007/s10683-016-9503-2](https://doi.org/10.1007/s10683-016-9503-2)) contains supplementary material, which is available to authorized users.

✉ John H. Kagel
kagel.4@osu.edu

Emel Filiz-Ozbay
filizozbay@econ.umd.edu

John C. Ham
john.ham.econ@gmail.com

Erkut Y. Ozbay
ozbay@umd.edu

¹ Department of Economics, University of Maryland, College Park, USA

² Department of Economics, IFAU, IRP and IZA, National University of Singapore, Singapore, Singapore

³ Department of Economics, Ohio State University, Columbus, OH, USA

This paper contributes to the labor economics literature. The effects of personality traits, cognitive ability, gender, and ethnicity on economic outcomes has been of intense interest to labor economists using field data (see, for example, Borghans et al. 2008 and Heckman et al. 2006). In particular, cognitive ability, agreeability, and conscientiousness are strongly associated with positive labor market returns (Urzua and Veramendi 2012; Kern et al. 2013). Further, there are now numerous instances of differences in the behavior of men and women (see Croson and Gneezy 2009 and Niederle 2016 for reviews of the literature). In this paper we focus on the effects of personality characteristics, cognitive ability, and gender differences in the gift exchange game, a workhorse experimental paradigm for studying labor markets. With regard to gender differences, we follow the labor economics literature and examine how much of the substantial male–female differences we observe in the raw data are due to the genders having different characteristics and how much is due to their having different behavior. Studying these issues in a laboratory setting has a number of advantages relative to their study using “naturally” occurring field, as in the lab, one can isolate and control competing explanations for the behavior reported.

The paper also contributes to better understanding equilibrium outcomes in games with multiple equilibria, consistent with John Van Huyck’s interest in understanding behavior in games with multiple equilibria. From the viewpoint of standard economic theory, the equilibrium outcome for a gift exchange experiment with random rematching should approximate a one-shot game with workers providing minimum effort, which employers respond to with minimum wages. However, this is *not* the outcome observed in both laboratory and field experiments where, with random rematching, higher wages lead to greater effort levels (see Cooper and Kagel 2016 for a recent survey). It is generally recognized that other regarding preferences, of one sort or another, are behind this outcome. While different theoretical models of other regarding preferences would predict profitable gift exchange in equilibrium (Benjamin 2015; Bolton and Ockenfels 2000; Dufwenberg and Kirchsteiger 2004; Fehr and Schmidt 1999), coordination on any given equilibrium is not fully understood. This paper explores the role of personality characteristics, cognitive ability, and gender on the equilibrium outcome observed, which may have applicability well beyond the gift exchange game.

We focus on personality traits described by the Big Five personality characteristics, measured by the Big Five Inventory (BFI) (John et al. 2008). The BFI is a brief inventory that results in robust, and efficient measures of personality, gaining widespread acceptance in the psychology literature. It has been used in a number of labor market studies when available. The BFI provides measures of a person’s agreeableness, extroversion, conscientiousness, neuroticism, and openness. SAT scores are used to measure cognitive ability since they are highly correlated with the AFQT measure used in many labor market studies,¹ and are measured with relatively little error.

We find substantial gift exchange with the usual reported pattern: higher wages result in substantially higher average effort levels, which are mutually profitable for both “managers” and “workers.” We estimate separate models for men and women since the null hypothesis that they have the same coefficients is decisively rejected.

¹ The AFQT test is not publically available and thus cannot be used in experimental analysis.

Indeed, we find that the difference in male–female coefficients explains all of the substantial gender differences in the raw data, since the two groups have very similar characteristics. Moreover, estimating separate equations allows us to compare the variance of their permanent unobserved component, and we find that this variance is much higher for men. Previous work has not investigated the role of different coefficients versus different endowments in the differences in behavior by gender, nor has it investigated gender differences in terms of permanent variability.

SAT scores and agreeableness have important effects for wages offered and effort supplied for both men and women. However, if we drop SAT scores from the wage offer equations, agreeableness is not found to impact behavior, indicating the need to control for cognitive ability in assessing the impact of personality characteristics. The importance of allowing for gender differences over the full set of personality characteristics is demonstrated by (i) conscientiousness not playing a role in the pooled data, but does so when looking at men and women separately, and (ii) cognitive ability affects men’s, but not women’s, effort responses. The impact of personality characteristics on behavior can be quite substantial. A one standard deviation increase in agreeableness increases the effort by 87.4 percent for men and 106.5 percent for women, comparable to the effect of a one standard deviation increase in wages, the principal driver of increased effort levels in past experiments. Cognitive ability has a significant quantitative impact on wage offers, and conscientiousness *increases* the wage offers for men, but *lowers* it for women. Consistent with the gift exchange literature, we find a large amount of persistent unobserved heterogeneity across subjects, but adding the Big Five and SAT significantly reduces it in both the wage offer and effort response equations. Interestingly, this variance is much higher for men than women.

We briefly contrast these results for cognitive ability and personality characteristics with earlier experimental results. The focus is on experiments where, similar to gift exchange, standard economic theory calls for selfish and non-cooperative behavior, but these outcomes are not reported. Kurzban and Houser (2001) look at the role of the Big Five personality characteristics, along with other personality measures, on behavior in a voluntary contribution mechanism (VCM) public good game. They find no statically significant relationship between contribution levels and any of the Big Five characteristics, which they attribute to their relatively small sample size (57 subjects). Pothos et al. (2010) investigate individual correlations between the Big Five components on cooperation in a simultaneous move, prisoner’s dilemma game with random rematching, and find that more agreeable types were less likely to defect.² Becker et al. (2012) study direct correlations between the Big Five personality characteristics and behavior in a variety of games with random rematching. Agreeableness had the largest and most significant correlations in their study, being positively correlated with second mover returns and first mover allocations in the trust game, along with giving in the dictator game, and negatively correlated with punishment in the VCM game. Kagel and McGee

² They employed loaded terms (e.g., “defect” or “cooperate”) with no financial incentives. Jones (2012) reports similar results for a random end point, repeated PD game using generic terms and financial incentives.

(2014) investigate the role personality plays in a finitely repeated prisoner's dilemma game, finding that a one standard deviation increase in agreeableness increases the predicted probability of cooperation from 67.9 to 80.6 percent.

Our paper has some overlap with Ben-Ner and Halldorsson (2010) who study a gift exchange game with random rematching. They measured cognitive ability from the Wonderlic Personnel Test, including some of the Big Five as control variables in their analysis. They find that agreeableness has a significant effect on wage offers, as we do. However, their measure of cognitive ability does not have a significant effect on wage offers, in contrast to our results. We attribute this difference to measurement error given the limited nature of the Wonderlic Test compared to the SAT, which would bias the coefficient value towards zero. Their effort response equation is considerably simpler than ours, and they do not consider separate offer or effort equations for men and women. Schwieren (2012) reports the results of a gift exchange experiment focusing on gender effects. In her experiment firms and workers know each other's gender, which is not accounted for in the analysis. Hence the lower wages for women could be the result of gender stereotyping, while their lower effort levels could be a response to feelings regarding discriminatory wages from men. In our experiment, gender is anonymous, and we control for personality characteristics and cognitive ability, which Schwieren (2012) does not.

Anderson et al. (2012) use a large sample of (mostly male) truck driver trainees, measuring individual risk and time preferences, and obtaining scores for the Big Five personality characteristics and cognitive ability based on a Cognitive Skill Index. Among other things, they look at a modified (one-shot) trust game where first movers could send either \$0 or \$5, and second movers responded via the strategy method. More agreeable types, and those scoring higher on cognitive skills, were more likely to send the \$5, while more conscientious types were less likely to do so. More agreeable and more neurotic types were more likely to send money back in response to either a \$0 or \$5 transfer, with higher cognitive ability types sending less back in response to a \$0 transfer. Their work differs from ours in that we do not include measures of risk aversion and time preferences; one problem with these measures is they are likely to have significant measurement error (Gillen et al. 2015). Anderson et al. also use a different ability measure and different sample pool (truck drivers), and are unable to measure male–female wage differences given that the vast majority of subjects are male.

1 Big five personality measures and SAT scores

1.1 Measurement of personality traits

Prior to the start of each session, subjects filled out the 44-item Big Five Inventory (BFI) questionnaire (John et al. 2008).³ The Big Five personality characteristics

³ These were computerized using zTree (Fischbacher 2007), the software used to program the experiment. The BFI questionnaire is available at <http://www.outofservice.com/bigfive/>. The material in this section on the Big Five is taken from John et al. (2008) and Borghans et al. (2008).

represent a consensus among personality psychologists on a general taxonomy of personality traits. These personality characteristics do not represent a particular theoretical perspective but are based on natural language terms people use to describe themselves and others. The focus of the Big Five is on internal consistency rather than predictive ability. The psychology literature does not propose that these are of any intrinsic importance, or that personality differences are reducible to five traits; rather, the five dimensions represent personality at a very broad level of abstraction with each dimension summarizing a large number of distinct, more specific, personality characteristics. When more factors than the Big Five have been identified across cultures and studies, they are rarely replicated across multiple studies by independent investigators. In particular, we do not employ measures of risk preferences in the analysis as (i) there is the question of what measure to employ⁴ and (ii) Gillen et al. (2015) show that there is substantial measurement error in risk aversion measures, which will bias its coefficient towards zero in absolute value unless an instrumental variable approach is used.⁵ In addition, there is substantial evidence that much of the impact of risk preferences are captured in measures of cognitive ability (e.g., Dohmen et al. 2010).

The BFI measure consists of 44 short phrases prototypical of the five characteristics. For example, the characteristic *openness* includes “is original, comes up with new ideas” in the BFI. There are between eight and ten phrases associated with each of the five subscales, which are self-scored, and designed to disguise the nature of the characteristics.⁶ Further, subjects are not told either the purpose of the questionnaire or their scores on it. The BFI is used when time is at a premium, as it typically takes between 10 and 15 min to complete. The personality traits consist of:

1. Agreeableness: contrasts a pro-social and communal orientation to others, and includes traits such as altruism, tender-mindedness, trust, and modesty.
2. Extroversion: implies an energetic approach toward the social and material world, including traits such as sociability, activity, assertiveness, and positive emotionality.
3. Conscientiousness: describes socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks.
4. Neuroticism: contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense.

⁴ It is clear that rank order correlations of risk preferences across different domains are far from 1.0 (Einav et al. 2012).

⁵ This would explain why risk aversion measures are rarely statistically significant in experimental work (e.g., Ben-Ner and Halldorsson 2010, report including measures of risk preferences to account for the amount of money first movers send in a trust game has no significant effect).

⁶ For example, two of the agreeableness subscales questions consist of question (2) “Tends to find fault with others” which is reverse-keyed so that a higher rating on this scale counts as less agreeable, and question number (7) “Is helpful and unselfish with others” where a higher rating counts as more agreeable.

5. Openness: describes the breadth, depth and complexity of an individual's mental and experiential life.

Scoring higher on the scale of each characteristic is associated with the more positive elements of the traits, except for neuroticism, where the high pole is associated with poorer ability to cope with life.

1.2 Measures of cognitive ability

As noted above, SAT scores are used as a proxy measure for cognitive ability (denoted by g) because they are readily available through the university's Registrar's office, are measured with relatively little error, exhibit relatively wide variation in our sample, and are heavily weighted in university admissions, where they are considered to be a strong measure of ability. They are highly correlated with AFQT scores used in labor market field studies, and are relatively highly correlated with Raven's Advanced Progressive Matrices scores (hereafter Raven's measure). Like AFQT scores, SAT scores can be affected by learning (and thus by the Big Five characteristics), as well as by cultural factors. Moreover, Raven's measures tend to exhibit high variance for the same individual across time (Bors and Vigneaub 2001), suggesting that they contain much more measurement error than SAT scores, which is likely to bias downward coefficient values for cognitive ability.⁷

1.3 Summary statistics

The Big Five personality characteristics are on a scale of 1–5, with SAT scores on a scale of 400–1600. In the data analysis, we convert these scores to the percent of maximum possible score (POMP). Specifically, for individual i ,

$$POMP_i = \frac{Observed_i - Minimum}{Maximum - Minimum}$$

where $Observed_i$ is the observed score for individual i , $Minimum$ is the minimum possible score on the scale, and $Maximum$ is the maximum possible score on the scale. Since POMP is a linear transformation of the original scores, statistical evaluation of the data remains unchanged, but the regression coefficients are on a normalized scale, making them easier to interpret (Cohen et al.1999). Table 1 reports average POMP scores, their ranges and their standard deviations for men and women. *Men and women have very similar POMP scores; hence any differences between them in the behavior reported below are not due to differences in characteristics.*

⁷ Measurement error in cognitive ability will also bias the Big 5 coefficients if they are correlated with cognitive ability, as in our data.

Table 1 POMP scores for the big five and SAT scores

Group	N	Statistic	Scale					
			SAT	Open	Con	Ext	Agr	Neu
Males	105	Mean	74.0	67.6	65.1	58.2	68.2	38.4
		SD	12.4	14.6	14.2	17.7	16.5	18.5
		Range	44–98	22–100	25–95	21–97	25–100	3–94
Females	87	Mean	71.1	67.8	67.8	60.2	66.4	44.5
		SD	10.9	15.5	13.5	19.5	15.6	18.9
		Range	45–97	25–90	30–98	12–97	13–95	9–91
P values ^a			0.087	0.892	0.185	0.455	0.459	0.027

Open Openness; *Con* Conscientiousness; *Ext* Extroversion; *Agr* Agreeableness; *Neu* Neuroticism

^a P values for equality of mean values between men and women

2 Experimental design and procedures

After the completion of the BFI questionnaire, subjects were randomly divided into two equal size groups: “managers” and “employees.”⁸ Subjects played the same role for 12 periods, announced in advance. Managers were randomly matched with an (anonymous) employee in each period. There were 16 subjects in each session, with no employee was re-matched with the same manager more than twice, and never re-matched in two consecutive periods. The anonymity was designed to generate a sequence of games so that subjects could not develop individual reputations.⁹

In stage 1 of each period, managers chose a wage, an integer w from the interval $[0, 100]$. In stage 2, each employee, after seeing w chose an effort level e , an integer from the interval $[0, 100]$, after which managers observed the effort level of their employee.¹⁰ Payoffs were symmetric and calculated as follows for managers π_M and employees π_E ¹¹

$$\pi_M = 100 - w + 5e$$

$$\pi_E = 100 - e + 5w$$

Subjects were asked to calculate the payoffs of both managers and employees in five examples before starting play for cash. Assuming that players care only about their own income, the unique subgame perfect Nash equilibrium (SPE) of the game

⁸ The instructions are available at www.econ.umd.edu/~filizozbay/FHKO_instructions.pdf.

⁹ Subjects only learned the outcome of their own match, designed to limit group reputations, confirmed by the fact that adding session specific dummy variables to the regressions showed no significant effects.

¹⁰ Computer screens showed payoffs for the current period, along with the history of that subject’s choice and payoffs.

¹¹ To our knowledge, a scaled version of these payoff functions were first employed in Brandts and Charness (2004), and have been employed in a number of studies since then.

is zero effort for any wage offer, in anticipation of which wage offers are zero. Alternatively, the efficient wage and effort levels which maximize total surplus equal 100. Usually the SPE prediction fails, and higher wages met with higher average effort responses.

Twelve sessions were run with undergraduates at the Experimental Economics Laboratory at University of Maryland. All sessions were computerized using the zTree software (Fischbacher 2007). Sessions lasted about 60 min, including the BFI questionnaire. Subjects were paid at a rate of 250 experimental currency units (ECUs) to 1 US dollar along with a \$6 participation fee. Average total earnings were approximately \$21.75 for employees and \$14.40 for managers.

3 Expectations regarding impact of cognitive ability and personality characteristics

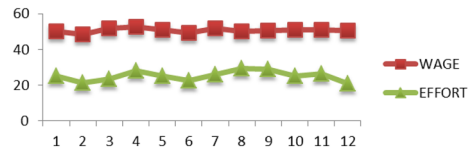
Higher scores for cognitive ability, agreeability, and conscientiousness are highly correlated with higher earnings in field data. Cognitive ability could have a negative or positive effect on wage offers. On the negative side, those with higher cognitive ability might be expected to offer lower wages, anticipating zero or low effort levels resulting from the absence of individual reputation effects. However, past experiments show that higher wage offers typically result in higher effort levels and greater profits for managers, which higher cognitive types could anticipate. Further, since the extant literature indicates that men and higher cognitive ability types tend to be less risk averse (Dohmen et al. 2010; Burks et al. 2009), male managers, as well as managers with greater cognitive ability, might also be more willing to bear the risk inherent in offering higher wages.¹² There are potential positive and negative effects of greater cognitive ability with respect to effort levels as well. Again, the implications for lower or zero effort as a result of workers inability to establish individual reputations should be more transparent to workers with greater cognitive ability. On the other hand, workers with greater cognitive ability are more likely to be sensitive to the social norms of reciprocity in work relations, holding the Big Five values constant.

One unambiguous effect anticipated here is that more agreeable types will offer higher wages and provide greater effort levels. This follows directly from the agreeability characteristic, which contrasts a pro-social and communal orientation with antagonism towards others, and includes traits such as altruism and trust. Similar results have been reported for the trust game where more agreeable types transfer more and return more (Becker et al. 2012), and are more cooperative in repeated prisoner dilemma games (Jones 2012).¹³ *The real question here is what will be the relative impact of greater agreeability on outcomes compared to increased wages— the primary driver of increased effort levels in past gift-*

¹² What is in question is the magnitude of these differences in risk aversion between men and women (see Niederle, 2016).

¹³ Of course, this is subject to the caveat that we include SAT scores and previous studies do not, so at the very least the quantitative impact of agreeableness in our study may differ from other studies.

Fig. 1 Average wage and effort level per period: wage and effort levels between [0, 100]



exchange experiments. We believe that most economists would predict a weaker effect of increased agreeability on effort responses compared to comparable wage increases. However, social psychologists and management science types might well predict similar quantitative effects for the two.

Conscientiousness describes individuals who tend to follow norms and rules, with higher scorers having better job performance, the implications of which would appear to be somewhat ambiguous with respect to wage offers and effort levels in the gift exchange game. However, to the extent higher effort in response to higher wages is the accepted social norm, more conscientious types would be more likely to respond that way. We have no priors on the impact of the remaining Big Five personality characteristics.

Investigations of gender differences in the gift exchange game have been limited and have not allowed for interactions between gender and measures of cognitive ability or personality characteristics. The trust game, which has many characteristics in common with gift exchange, shows limited gender differences. Reviews of outcomes show no gender differences in sending behavior (trusting), or that women are more trusting than men (Croson and Gneezy 2009). With respect to money returned (reciprocity), papers report no gender differences, or that women are more reciprocal than men. Few, if any, of these trust experiments have accounted for personality characteristics and cognitive ability. Given this caveat, either these results suggest no gender differences, or that women will offer higher wages and higher effort levels than men for the gift exchange game.

4 Basic experimental results

4.1 Basic results for the pooled data

Given our focus on gender differences, we discuss the pooled data to reassure readers that our experiment produces results similar to those found in previous studies, as readers may wonder whether using a consent form to obtain SAT scores and demographic data created subject selection bias or that answering the BFI questionnaire affected subjects' behavior. Figure 1 shows average wage offers and effort levels over time for the pooled data. Average wage offers are persistently higher than effort levels, with neither close to the SPE equilibrium, and are relatively constant over time, with at most a small end of session effect with respect to effort levels. Figure 2 shows effort levels at different wage rates (error bars indicate the 95% confidence intervals). Wage offers of zero occurred 12.0% of the time, with the typical effort response being zero; 29.9% of the time all non-zero

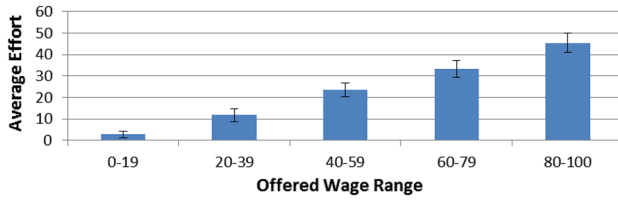


Fig. 2 Effort level over each wage interval

wages were met with zero effort, and 20% of the time wage offers in the three highest categories were met with zero effort.

Figure 3 shows managers' payoffs for different wage rates (error bars indicate the 95% confidence intervals). Interestingly managers' payoffs are monotonically increasing in offered wages. In short, our pooled results are quite similar to those reported in the previous literature.

4.2 Basic results for men and women

We now turn to gender differences in behavior. Figure 4 shows that although there are minimal differences in effort responses between men and women at lower wages, at middle and higher wage rates men consistently provide greater average effort than women do. Figure 5 shows that men tend to offer higher wages than women, with most of this difference accounted for by the higher frequency of wage offers in the interval 80–100 (39.8 percent of all men's wage offers versus 16.9 percent for women).

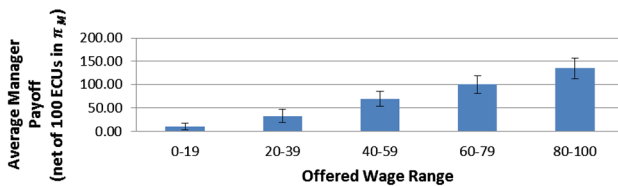


Fig. 3 Average income of managers at each wage interval (net of the 100 ECUs included in π_M)

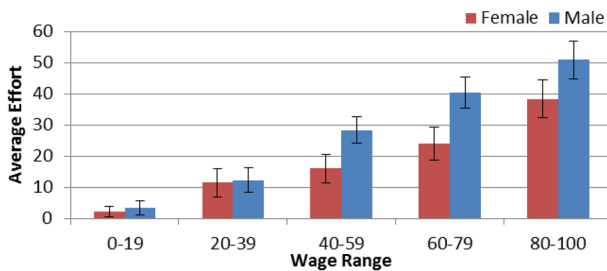


Fig. 4 Effort level over wage intervals for men and women

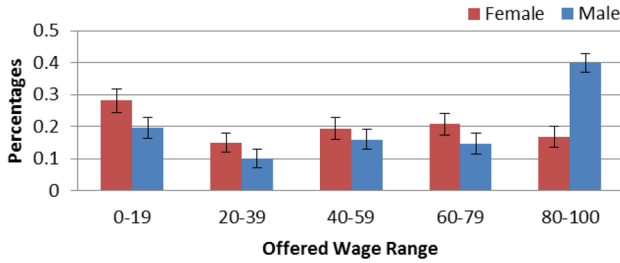


Fig. 5 Distribution of wage offers for men and women

Conclusion 1 *Relative to women, men tend to offer higher wages, and respond with greater effort to higher wages. Importantly, we find that these differences persist after controlling for BFI and SAT scores, as shown below.*

5 Econometric analysis

5.1 Statistical analysis including the big five and SAT scores: wage offers

Since wages are drawn from the closed interval $[0, 100]$, a random effect, two-limit Tobit model is used for the statistical analysis. Specifically, we assume that the offered wage index score (OWIS) for individual in period takes the random effects form

$$w_{ip}^* = \beta X_{ip} + \alpha_i + e_{ip}. \tag{1}$$

Further, we assume that observed wage offers are determined by

$$w_{ip} = 0 \quad \text{if } w_{ip}^* < 0, \tag{2}$$

$$w_{ip} = 100 \quad \text{if } w_{ip}^* > 100,$$

$$w_{ip} = w_{ip}^* \quad \text{otherwise.}$$

Our estimates of the OWIS coefficients for men and women are reported in Table 2.¹⁴ Specifically, columns (1) and (3) present the results for men with and without SAT included, while columns (2) and (4) present the corresponding results for women. Results excluding SAT scores are reported because past studies have often looked at the impact of the Big Five with no information on cognitive ability. Thus, by the standard Theil-Griliches specification error result, this will produce biased coefficients if (i) any of the Big Five variables has a nonzero partial correlation with SAT scores and (ii) SAT scores affect behavior (as shown below).

¹⁴ For the pooled data, a likelihood ratio test of the null hypothesis that men and women have the same coefficients conditional on them having different intercepts is strongly rejected ($p < 0.01$) (see Table A1 in the Online Appendix).

Table 2 Random effects estimates of the wage index function: men and women (standard errors are in parentheses)

	Men (1)	Women (2)	Men (3)	Women (4)
SAT	1.530*** (0.360)	1.667*** (0.464)		
Agreeableness	0.648** (0.280)	0.453* (0.255)	0.461 (0.324)	0.300 (0.284)
Openness	-0.046 (0.296)	0.157 (0.223)	0.061 (0.347)	0.228 (0.251)
Neuroticism	-0.093 (0.265)	0.190 (0.206)	-0.115 (0.312)	0.085 (0.231)
Conscientiousness	1.147*** (0.355)	-1.042*** (0.339)	1.477*** (0.406)	-1.23*** (0.378)
Extroversion	-0.032 (0.272)	0.343 (0.247)	-0.414 (0.298)	-0.012 (0.256)
Period	0.0035 (0.386)	-0.143 (0.295)	0.007 (0.387)	-0.140 (0.294)
Constant	-142.7*** (50.53)	-91.40* (54.60)	-29.23 (49.96)	63.11* (38.03)
Log-likelihood	-2167.0	-2214.5	-2174.7	-2220.2

Based on 50 men and 46 women

In this and the following tables: * denotes statistically significant at the 10 percent level. ** denotes statistically significant at the 5 percent level and *** denotes statistically significant at the 1 percent level

Comparing the results with and without SAT scores indicates the magnitude of these biases, if any, from omitting SAT scores.

SAT scores and agreeableness are statistically significant for both men and women, and are of comparable magnitude as well. Subjects with higher SAT scores offer higher wages, as do more agreeable types. The strong positive relationship between SAT scores and wages may indicate that managers with higher cognitive ability are better attuned to the large potential profits associated with higher wages. One very striking result is that if we omit SAT scores, agreeableness loses its statistical significance for both men and women; these results are in sharp contrast to the results in the labor market studies cited in the introduction. This strongly suggests exercising care in interpreting results from studies that only include the Big 5 (and no measure of cognitive ability). Interestingly, dropping the Big Five variables does not affect the SAT coefficients for both men and women. One of the more interesting results here is that conscientiousness is statistically significant for both men and women, but *opposite* in sign: Positive for men but negative, with a comparable absolute value, for women. Although this difference for increased conscientiousness is unexpected, an immediate justification for it can be found in the estimated effort response index functions reported below. There, other things equal, greater conscientiousness in men results in a modest but positive increase in

the effort response, but for women it results in a modest *decrease* in the effort response. Although we did not anticipate different signs with respect to conscientiousness by gender, at least there is an internal consistency to these results, with more conscientious men, thinking from their own perspective, being more likely to offer higher wages, while women, thinking from their perspective, would not.¹⁵ Interestingly, when we pool the data, the effects for men and women (approximately) cancel out, with a conscientious coefficient of 0.191 and a standard error of 0.245.

Adding the Big Five and SAT scores to the Tobit models reduces the variance of the persistent individual unobserved heterogeneity σ_a by 31 and 19 percent in wage offers (OWIS) for men and women respectively. Finally, it is interesting to note that after controlling for SAT scores and the Big Five, the variance of the persistent individual heterogeneity in the OWIS is 50 percent larger for men than women; a gender difference not noted previously.

To obtain an idea of the magnitudes of the effects implied by the estimates in columns (1) and (2) of Tables 2, 3 presents the effect on OWIS of a one standard deviation (1sd) increase in SAT scores, agreeableness, and conscientiousness, the variables that are statistically significant in columns (1) and (2). A 1sd in SAT scores increases the OWIS for men and women by 19.2 and 17.5 percent respectively, while a 1sd increase in agreeability increases the OWIS for men and women by 9.9 and 6.6 percent respectively. Further, 1sd increase in conscientiousness increases the OWIS for men by 23.4 percent but decreases the OWIS by 32.8 percent for women. Thus for both genders, a 1sd increase in conscientiousness has a larger effect on the OWIS than a 1sd increase in SAT score, which in turn has a larger effect than a 1sd in agreeability.

Conclusion 2 *Men and women differ substantially in their mean wage offer index functions, but show comparable effects in terms of a one standard deviation increase in SAT scores and agreeableness. The impact of conscientiousness is positive for men and negative for women, with both effects statistically significant and of comparable absolute value. Adding SAT scores and the Big Five substantially reduces the permanent unobserved heterogeneity in the OWIS for both men and women.*

5.2 Statistical analysis including the big five and SAT: effort responses

Since actual effort levels are bounded between $[0, 100]$ we again use a two-sided random effects Tobit model for our statistical analysis. The effort response index score (ERIS) of individual j in period p , who receives a wage offer of from manager i , is given by.

$$E_{jp}^* = \delta_1 X_{jp} + \delta_2 W_{ip} + \delta_3 (W_{ip} * X_{jp}) + \gamma_j + \varepsilon_{jp} = \delta Z_{jip} + \gamma_j + v_{jp}. \quad (3)$$

Further, observed effort response is given by

¹⁵ A potential explanation for this difference is discussed in the concluding section of the paper.

Table 3 The effect of a one standard deviation increase in key explanatory variables on the OWIS (change as a percent of the mean value of the wage index function in parentheses)

	Mean OWIS	Change in mean OWIS		
		SAT	Agreeableness	Conscientiousness
Men	62.4***	19.2*** (30.8%)	9.9** (15.9%)	14.6*** (23.4%)
Women	40.8***	17.5*** (42.9%)	6.6* (16.2%)	-13.4*** (-32.8%)

$$E_{jp} = 0 \quad \text{if } E_{jp}^* < 0, \quad (4)$$

$$E_{jp} = 100 \quad \text{if } E_{jp}^* > 100,$$

$$E_{jp} = E_{jp}^* \quad \text{otherwise.}$$

In Eq. (3) γ_j is a random effects error term, which is *iid* across j and distributed as, $N(0, \sigma_\gamma^2)$ while ε_{jp} is an idiosyncratic error term, which is *iid* (over j and p) and distributed as $N(0, \varepsilon_{jp}^2)$. Again the variance of γ_j σ_γ^2 represents a measure of the persistent unobserved subject heterogeneity in effort responses. We allow for interaction terms between the explanatory variables and the offered wage, since the null hypothesis of no interactions was decisively rejected ($p < 0.01$ in all cases).

Table 4 reports the estimated effort response index functions in the same format as those for wage offer index functions.¹⁶ The analysis is restricted to the case where a positive wage is offered, since zero wage offers are overwhelmingly met with zero effort. Cognitive and non-cognitive characteristics essentially play no role in responses to zero wage offers; including them would bias the estimates.¹⁷

The estimates indicate that men with higher SAT scores provide lower effort responses throughout the range of possible wages than men with lower SAT scores, but higher wages moderate this difference, so that the major impact is confined to lower wages. For example, based on Table 4, the effect of a one-unit increase in the mean SAT at wage of 85.0 is -0.384 , which is considerably smaller than the effect of -0.687 at a wage of 62.4. Further, and more importantly, it is considerably more profitable, on average, to offer a higher wage to a man with a high SAT score, than to offer an average wage to a man with an average SAT score (124 ECUs versus 42 ECUs).¹⁸ In short, increases in cognitive ability result in lower effort, holding wages

¹⁶ Here too we considered a version of (3) that contained session fixed effects, but this did not have any effect on the standard errors, and the session effects were not close to being significant at standard levels.

¹⁷ The bias arises from mixing the effort index function (3) with the very different ERIS that applies to a zero offered wage. Since wage offers are exogenous to the responders; omitting the responses for zero wage offers does not create selection bias.

¹⁸ This is after subtracting out the 100 ECUs in the manager's payoff function: high wage is one standard deviation above the average wage; high SAT is one standard deviation above the average SAT.

Table 4 Random effects estimates of the effort response index function (standard errors in parentheses)

	Men (1)	Women (2)	Men (3)	Women (4)
Wage	-0.042 (0.621)	0.277 (0.579)	0.665 (0.557)	0.150 (0.417)
SAT	-1.525*** (0.550)	0.097 (0.527)		
Agreeableness (Agr)	0.565 (0.408)	0.314 (0.367)	0.327 (0.402)	0.289 (0.344)
Openness (Open)	0.416 (0.463)	-0.013 (0.323)	0.291 (0.465)	0.001 (0.316)
Neuroticism (Neu)	0.319 (0.324)	-0.146 (0.269)	0.242 (0.326)	-0.140 (0.269)
Conscientiousness (Con)	0.770* (0.403)	0.230 (0.378)	0.769* (0.406)	0.234 (0.377)
Extroversion (Ext)	-0.294 (0.379)	0.476* (0.266)	-0.287 (0.382)	0.466* (0.260)
Period	-0.516 (0.389)	-0.157 (0.303)	-0.558 (0.390)	-0.153 (0.303)
Wage*SAT	0.0134*** (0.0050)	-0.0145 (0.0046)		
Wage*agr	0.0087** (0.0036)	0.0139*** (0.0033)	0.0109*** (0.0035)	0.0143*** (0.0030)
Wage*open	-0.0002 (0.0042)	-0.0051* (0.0028)	0.0012 (0.0041)	-0.0054* (0.0028)
Wage*neu	0.0012 (0.0030)	0.0060** (0.0024)	0.0018 (0.0029)	0.0059** (0.0024)
Wage*con	-0.0092** (0.0036)	-0.0053* (0.0031)	-0.0088** (0.0035)	-0.0053* (0.0031)
Wage*ext	0.0020 (0.0033)	-0.0028 (0.0023)	0.0016 (0.0033)	-0.0026 (0.0023)
Constant	-59.27 (65.76)	-64.36 (66.12)	-139.1** (59.60)	-56.22 (47.64)
Log likelihood	-1821.3	-1443.1	-1826.0	-1443.2

constant, but at the same time high SAT types are much more reactive, in a positive way, to increased wages. The latter is similar to what was reported with respect to conscientiousness in that men with higher SAT scores effort responses is consistent with their higher wage offers. It is also inconsistent with the idea that higher SAT types offer lower effort at any given wage out of a better understanding of the absence of individual reputation effects, or there are mixed motives at work. In contrast, for women, neither of the two SAT variables is individually significant, nor

are they jointly significant at conventional test levels.¹⁹ It is not clear why women with comparable SAT scores do not behave the same way as the men, as women with higher SAT scores also offer higher wages than those with lower scores.²⁰

The coefficients on agreeableness are positive but not statistically significant for men and women, with the coefficients on the interaction of wage and agreeableness statistically significant in both cases. The overall effect of agreeableness at the mean wage rate is to increase effort (the ERIS) by 1.06 (.337) and 1.12 (.316) for men and women, with both statistically significant (standard errors in parentheses).

The coefficient on conscientiousness is positive for both men and women, but the value is considerably larger for men relative to women, and is significant at the 10% level for the men. The coefficient on the interaction of wage and conscientiousness is significantly negative for both men and women, but again the effect is bigger (in absolute value) for men. At the mean wage rate, the overall impact of conscientiousness is to raise the effort response of men by 0.245 (.329) while for women there is a slight reduction, -0.078 (.324), with neither effect statistically significant at conventional levels. For men, at lower average wages, an increase in conscientiousness results in a statistically significant *increase* in effort levels. At higher wages, more conscientious women have significantly *lower* effort levels. Focusing on the signs for conscientiousness in Tables 3 and 5, this differential effort response with respect to conscientiousness is internally consistent with the effect of conscientiousness on male and female wage offers. That is, both genders may accurately predict own gender effort response with respect to conscientiousness, and act accordingly in setting wages, even though they do not know the gender of the person they are interacting with in any given play of the game. Note that there is independent evidence for this sort of effect, referred to as “consensus bias” in the psychology literature: the overuse of self-related knowledge in estimating the prevalence of attributes in a population (Ross et al. 1977; Kruger and Clement 1994).

The coefficient on the interaction of wage and neuroticism is statistically significant for women, but the overall effects are relatively small and insignificant for both genders at the mean wage rate. The coefficient on extroversion is positive for women, but again the overall effects are relatively small and insignificant for both genders at the mean wage rate. The coefficient on the interaction of wage and openness is negative and significant at the 10 percent level for women, but including the interaction effect with wages, openness is statistically insignificant for both men and women.

Finally, adding the Big Five and SAT scores reduces the variance of the persistent individual unobserved heterogeneity by 33 percent for men and 35 percent for women. In addition, the variance of the persistent individual unobserved heterogeneity for women is twice as large as it is for men.

¹⁹ The test statistic is 0.10 with a p value of 0.951. Moreover, at the mean wage rate, the overall SAT response for women is tiny 0.013 (.447; standard errors in parentheses).

²⁰ This result is, however, consistent with the evidence that it takes women longer to learn to act strategically in some game theoretic settings than men (see Casari et al. 2007).

Table 5 Change in the eris resulting from one standard deviation increases in the key explanatory variables (change as a percent of the mean value of the effort index function in parentheses)

	Mean ERIS	Change in mean ERIS			
		Wage	SAT	Agreeableness	Conscientiousness
Men	19.8***	26.8*** (135.4%)	-8.9* (-44.9%)	17.3* (87.4%)	3.8 (19.2%)
Women	16.9***	16.9*** (100.0%)	0.1 (0.6%)	18.0* (106.5%)	-1.1 (-6.5%)

Effects are calculated at mean values for all variables

Table 5 presents the effect on the effort response of a one standard deviation (1sd) increase in offered wages, SAT scores, agreeableness and conscientiousness on the ERIS. A 1sd in the offered wages increases the ERIS for men and women by 26.8 and 16.9 units respectively. Further, a 1sd in SAT scores reduces the ERIS for men by -8.9 units but has no effect on women, while a 1sd increase in agreeableness increases the ERIS for men and women by 17.3 and 18.0 units. Finally, a 1sd increase in conscientiousness increases the ERIS for men by 3.8 units but reduces it for women by 1.1 units.

It is quite impressive that a 1sd increase in agreeableness has a comparable effect on the ERIS as does a 1sd increase in the offered wage. This result is inconsistent with purely selfish economic man, but is consistent with the growing economics literature on other-regarding preferences. It also sheds some light on the nature of agreeableness on wages. Some may argue that people are agreeable in field data because it pays to be more agreeable. However, here more agreeable people actually pay a price (i.e. lose money) for being more agreeable as the higher effort level reduces their earnings, even though there is at best, very limited payoff to them in terms of receiving future higher wage offers because of the random rematching. This is consistent with the Big 5 measure of agreeability having strong empirical content labor markets.

Unlike the wage offer index function, dropping SAT scores does not have an important effect on the Big Five variables for any of our samples. However, it is interesting to note that there is a second order effect of dropping SAT scores in the race by wage interaction effects (which are not reported in the table but are included in the specifications). With the inclusion of SAT scores, the only statistically significant race effect is the dummy variable for Asian men, which is positive and significant at the 10 percent level. With the exclusion of SAT scores, the coefficient on the wage and the African-American dummy is negative and statistically significant at the 5 percent level for men. (The African-American dummy variable is not significant at conventional levels when SAT scores are included.) These ethnicity effects suggest that it will be important to control for SAT scores in any study examining ethnic (as opposed to gender) differences in experiments.

Conclusion 3 *There are major differences in effort responses of men and women: For men, the marginal effect of a higher SAT score is to reduce effort responses*

particularly at the low end of the wage scale, but they have higher average effort responses than men with lower SAT scores. In contrast, SAT scores have minimal impact on effort responses of women. More conscientious men supply greater effort, particularly at lower wages, while more conscientiousness women provide less effort at higher wages. A one standard deviation increase in average wages increases the effort response index function for men by 135 percent compared to 100 percent for women, while a one standard deviation increase in agreeableness has comparable effects. Dropping SAT scores does not affect the Big Five coefficient estimates, but has a substantial impact on the ethnicity coefficients.

6 Power calculations

A referee raised the question of getting some idea of the increases in sample size that would be needed to determine if some of the Big 5 characteristics that had large, but statistically insignificant values, were in fact significant. Power calculations of this sort are less frequent in advance in laboratory experiments than in field experiments because it is much easier to run additional lab sessions. However, they can be done ex post to determine if it is worthwhile, given the question at hand, to conduct additional sessions.²¹

To make an explicit power calculation, recall that power equals the probability of rejecting a null hypothesis when it is false, so depends on the true model. In a simple comparison of means between a randomly assigned treatment and control groups, if one specifies the variance of the outcome variable and the treatment effect, one can obtain a closed form expression for power given different sample sizes and significance levels. However, with our random effect Tobit model, power calculations require simulations (as described in Online Appendix).

Given space constraints, we carry out a ‘typical’ power calculation for the openness coefficient in the women’s OWIS. We choose this example since openness has a substantial, but insignificant, coefficient in our estimates. We calculate power at the 95% confidence level for the null hypothesis that the openness coefficient is zero when we estimate an OWIS that is a function of ability and the Big 5, assuming that the true model is given by the coefficients we obtained when the OWIS includes only these variables. At the actual sample size of 46 women, power equals 0.19; i.e., we only have a 0.19 chance of finding that it is significantly different from zero using a t test, even though it is different. If we triple the sample to 138 women, this increases to 0.25 chance of finding that it is significantly different from zero, and if we increase the sample size by a factor of five, it increases to 0.26. These results imply that the insignificant coefficient on openness in column (2) of Table 2 may reflect a lack of power, but that very large increases in the sample size are needed to increase power substantially. One can do power calculations for other variables with large but insignificant coefficients in an analogous manner.

²¹ For example, Ham, Kagel and Lehrer (2005) estimated a model of bidding in a private values auction, where cash balances were treated as an endogenous variable in a two stage least squares procedure using an instrument they generated within the experiment. They found that their results were quite noisy with 50 subjects, but precisely estimated once they increased the sample size to 100 subjects.

7 Summary and conclusions

We report results from a gift exchange game with random rematching of “firms” and “workers” accounting for the effects of gender, cognitive ability, and the Big Five personality characteristics on outcomes. We find substantial impacts on behavior for each of these factors, which are typically neglected in gift-exchange experiments and in a variety of other contexts.²² First, women offer lower wages than men. Women also offer less effort in response to the same wage offers. These differences result from different behaviors, not differences in cognitive ability or the Big 5 characteristics. Second, we find that the variances of persistent individual heterogeneity in the wage offer and effort response index functions are respectively 50 percent and 100 percent larger for men than for women. Finally, there are important gender differences in the effects of cognitive ability and personality traits.

The major impact of cognitive ability on wage offers is that both men and women with higher SAT scores offer higher wages than their counterparts with lower SAT scores. We conjecture that those with higher cognitive ability are better attuned to large potential profits associated with higher wages (Fig. 2), and are better able to tolerate the risk associated with offering these higher wages, which is captured in the SAT scores (Dohmen et al. 2010; Burks et al. 2009). Dropping SAT scores from the wage offer regressions results in the coefficient for agreeableness going from positive and statistically significant (when SAT scores are included) to being insignificant, for both men and women. Agreeableness is significant in field data where measures of cognitive ability are included in the analysis (see Heckman et al. 2006). Dropping SAT scores from the effort equation also affects the size and significance of ethnicity effects.

At times, the Big Five personality characteristics have as large an impact on the wage offers and effort responses as cognitive ability and wages. As in most gift-exchange experiments higher wage offers are met with higher effort responses. Here, a one standard deviation increase in agreeableness has the same impact on women’s effort response index function as does a one standard deviation increase in wages; the impact on men’s effort response index function is just under two-thirds the impact of a comparable wage increase. On the wage side, for men, the impact of a one standard deviation increase in conscientiousness increases the wage offer index function by about the same amount as a one standard deviation increase in SAT scores. Adding the Big Five and SAT scores significantly reduces the permanent unobserved heterogeneity for both men and for women.

One surprising result is that, for women, conscientiousness has the *opposite* impact on the wage offer index function (of roughly the same absolute value) as it does for men. This differential effect of conscientiousness on wages is consistent with best responding to its effect on effort where, since at low wages, increased conscientiousness leads to increases in the effort response index for men but has essentially the same, or a modest negative effect, for women. One possible

²² In turn, these hidden personality and cognitive ability factors may help account for the high degree of individual subject variability in effort responses and wage offers reported in the typical gift exchange experiment.

explanation for this differential effect of conscientiousness is as follows: One element of the conceptual definition of conscientiousness is “following norms and rules” (John et al. 2008, Table 4.2). With this in mind, note that there is some evidence suggesting that explicit monetary payments tend to drive out social preferences for women more so than for men.²³ In this case, women that are more conscientious would be more likely to have lower responsiveness to wage offers, with women in the role of “managers” best responding to these beliefs. In contrast, if men are either less sensitive or immune to this crowding out effect, and more accepting of the notion of explicit monetary benefits for reciprocal responses, more conscientious men would be more likely to take account of this fact and offer higher wages.

We have made a number of conjectures regarding the beliefs underlying the mechanisms through which cognitive ability and specific Big 5 characteristics impact behavior. As our referees have correctly pointed out, these conjectures can be directly investigated by collecting subjects’ beliefs in relationship to SAT scores and the Big 5 characteristics. We agree that this can, and should be, done but doing so lies well beyond the scope of the present paper, requiring an entirely new series of experimental sessions, and no doubt even larger sample sizes than reported on here.

It is also worthwhile comparing the successful role of the Big 5 and cognitive ability in organizing experimental outcomes here, to recent attempts to use outcomes in one set of economic games to organize behavior in a seemingly related set of games. For example, Dreber et al. (2014) compare giving in a dictator game to cooperation in an indefinitely repeated prisoner dilemma game. They report no systematic relationship. Davis et al. (2016) investigate whether a variety of individual characteristics in economics (e.g., risk attitude, time preference, altruism as measured by giving in the dictator game, etc.) are related to behavior in repeated play games. They note that “overall, the number of systematic relationships is surprisingly small”. This contrasts with the success of the Big 5 characteristic and cognitive ability in organizing behavior here, and their growing popularity in studying labor outcomes in field data. One explanation for these differences in success rates is that psychologists have vetted the Big 5 characteristics for a number of years with the explicit goal of summarizing a large number of distinct, more specific, personality characteristics identified consistently across cultures and studies. Moreover, when more factors than the Big Five have been identified, they rarely hold up cross multiple studies conducted by independent investigators. This is in contrast to outcomes in dictator games that are quite sensitive to small changes in experimental procedures (Cooper and Kagel 2016) and measures of risk preferences across different domains that are far from having rank order correlations close to one across individual subjects.²⁴

²³ Mellström and Johannesson (2008) found that paying people to donate blood reduced women’s donations, while men’s donations were unaffected. However, others have failed to find any evidence of gender effects in blood donations in large-scale natural field experiments (see Lacetera, Macis, and Slonim 2011).

²⁴ Lest one attribute these different success rates to random rematching versus repeated play games, Kagel and McGee (2015) report that in a finitely repeated prisoner dilemma games, agreeableness as measured by the Big 5, has a systematic (positive) relationship to stage-one cooperation rates in their game.

The results of this experiment have obvious and immediate implications for equilibrium outcomes in the social preferences literature in economics, as well as for all experiments that consider the economic effects of cognitive ability, personality traits, and gender, as well as for the labor economics literature.²⁵ Moreover, it is important to extend the analysis of the role of the Big Five personality characteristics, gender, and cognitive ability to gift exchange games in which agents can develop reputations through repeated or longer-term contracts. In this case, we would expect an even bigger impact of cognitive ability on effort responses for both men and women, as agents with greater cognitive ability, motivated by the potential for cooperation inherent in repeated interactions, provide greater effort at higher wages. It will be very interesting to see whether gender differences conditional on cognitive ability and personality traits persist, or even increase, in these richer models.

Finally, we discuss how to measure power for specific versions of the random effects wage offer and effort response equations. While space constraints prevented us from providing an exhaustive documentation of the power issues in our models, we hope that this discussion will encourage other authors to investigate the power properties of their models.

Acknowledgements We gratefully acknowledge helpful comments from Hal Arkes, Jennifer Cheavens, Sahiba Chopra, Laura Naumann, Albert Park, Noah Smith, Sergio Urzua, and participants at Binghamton University, Duke University, Hong Kong Institute of Science and Technology, Keio University, Ohio State University, the University of Arizona, the University of North Carolina at Chapel Hill, and Stanford University. We thank Kristian Lopez-Vargas for programming the experimental software. Ham's research was partially supported by NSF Grants SES-0136928 and SES-0627968. Kagel's research was partially supported by NSF Grants SES-1226460 and SES-0924764. Opinions, findings, conclusions, or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are responsible for any errors.

References

- Anderson, J., Burks, S., De Young, C., & Rustichini, A. (2012). Toward the integration of personality theory and decision theory in explanation of economic behavior. Institute for the Study of Labor (IZA), Bonn, Discussion paper 6750.
- Becker, A., Thomas, D., Thomas, D., Armin, F., and Fabian, K., 2012. "The Relationship between Economic Preferences and Psychological Personality Measures," Working Paper, University of Bonn.
- Benjamin, D. (2015). Distributional preferences, reciprocity-like behavior, and efficiency in bilateral exchange. *American Economic Journal: Microeconomics*, 7, 70–98.
- Ben-Ner, A. & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them and what affects them. *Journal of Economic Psychology*, 31, 64–79.
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90, 166–193.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43, 972–1059.
- Bors, D., & Vigneau, F. (2001). The effect of practice on raven's advanced progressive matrices. *Learning and Individual Differences*, 13, 291–312.

²⁵ For an extended discussion of the role of gift exchange studies in relationship to field data see Kagel (2015).

- Brandts, J., & Charness, G. (2004). Do labour market conditions affect gift exchange? some experimental evidence. *Economic Journal*, *114*, 684–708.
- Burks, S. V., Carpenter, J., Götte, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, social awareness, and job attachment. *Proceedings of the National Academy of Science*, *106*, 7745–7750.
- Casari, M., Ham, J. H., & Kagel, J. H. (2007). Selection bias, demographic effects and ability effects in common value auction experiments. *American Economic Review*, *97*, 1278–1304.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*, 315–346.
- Cooper, D. J., & Kagel, J. H. (2016). Other-regarding preferences: A selective survey of experimental results. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (Vol. 2). Princeton University Press.
- Crosno, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*, 448–474.
- Davis, D., Ivenov, A., & Korenok, O. (2016). Individual characteristics and behavior in repeated games: An experimental study. *Experimental Economics*, *16*, 67–99.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*, 1238–1260.
- Durber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated prisoner dilemma games: The role of altruism, inequity aversion and demographics. *Journal of Economic Behavior & Organization*, *98*, 41–55.
- Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*, 268–298.
- Einav, L., Finkelstein, A., Pascu, I., & Cullen, M. R. (2012). How general are risk preferences? choices under uncertainty in different domains. *American Economic Review*, *102*(6), 2606–2638.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready made economic experiments. *Experimental Economics*, *10*, 171–178.
- Gillen, B., Snowberg, E., & Yariv, L. (2015). *Experimenting with measurement error: Techniques with applications to the caltech cohort*. California Institute of Technology: Mimeo.
- Ham, J. C., Kagel, J., & Lehrer, S. (2005). Randomization, endogeneity and laboratory experiments: The role of cash balances in private value auctions. *Journal of Econometrics*, *125*, 175–205.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.
- John, P. O., Naumann, L. P., & Soto, J. C. (2008). “Paradigm shift: To the integrative big five”. In O. P. John, R. W. Robbins, & L. A. Pervin (Eds.), *Handbook of personality, theory, and research* (pp. 114–158). New York: Guilford Press.
- Jones, Matthew T. 2012. “Strategic complexity and cooperation: An experimental study,” SSRN Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2025659.
- Kagel, J. H., Frechette, G., & Schotter, A. (2015). “Laboratory experiments: The lab in relationship to field experiments, field data, and economic theory” in *methods of modern experimental economics*. Cambridge: Cambridge University Press.
- Kagel, J. H., & McGee, P. (2014). Personality and cooperation in finitely repeated prisoner dilemma games. *Economic Letters*, *124*, 274–277.
- Kern, M., Duckworth, A., Urzua, S., Loeber, R., Loeber, M., & Lynam, D. (2013). *Mind your manners! aspects of agreeableness and early adult outcomes for inner-city boys*. Mimeo: University of Pennsylvania.
- Kruger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, *67*(4), 596–610.
- Kurzban, R., & Houser, D. (2001). Individual differences in cooperation in a circular public goods game. *European Journal of Personality*, *15*, 37–52.
- Lacetera, N., Macis, and R. Slonim, 2011. “Rewarding altruism: A natural field experiment.” NBER Working Paper 17636.
- Mellström, C., & Johannesson, M. (2008). Crowding out in blood donations: Was titmuss right? *Journal of the European Economic Association*, *6*, 845–863.
- Niederle, M. (2016). “Gender effects in economic experiments,”. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics*. Princeton: Princeton university press.

- Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., & Busemeyer, J. R. (2010). Understanding cooperation in the prisoner's dilemma game. *Personality and Individual Differences, 51*, 210–216.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception. *Journal of Experimental Social Psychology, 13*, 279–301.
- Schwieren, C. (2012). The gender wage gap in experimental labor markets. *Economics Letters, 117*(3), 592–595.
- Urzua, S., & Veramendi, G. (2012). *Empirical strategies to identify the determinants and consequences of skills*. Mimeo: University of Maryland.