

# Which is the More Predictable Gender? Public Good Contribution and Personality

Marco Perugini, Jonathan H W Tan and  
Daniel John Zizzo<sup>1</sup>

## ABSTRACT

*Questionnaires can and have been used to predict behaviour in economic settings. Using two sets of measures from personality psychology (the Big Six) and social psychology (Social Value Orientation), we find that the behaviour of men is predictable in the first half of a public good contribution experiment, whereas that of women is not. This result agrees with the reinterpretation of Carol Gilligan's (1982) view that women are more sensitive to the context in which decisions are made. In practice, questionnaires such as those used in human resource management settings may fail to capture women's preferences.*

## 1. INTRODUCTION

THERE HAS BEEN A GROWING INTEREST over whether and how women behave differently from men in economic settings (e.g., Cadsby and Maynes, 1998; Eckel and Grossman, 2001; Andreoni and Vesterlund, 2001; Englestad et al., 2006). Theories of social preferences and of cooperation have dramatically different implications according to the mixture of types in the population, and it is natural to think of gender as a possible marker of inter-individual differences in behaviour.<sup>2</sup> Such differences are possibly attributable to explanations based on gender differences in personality. Economists have, of late, considered the predictive value of questionnaires drawn from social psychology, such as the Social Value Orientation (SVO), or from personality psychology, such as the Big Five personality factors (for example, Caplan, 2003). Personality questionnaires are, in practice, systematically used

in human resource management and other organisational settings, where running a complex behavioural experiment to capture preferences is not possible but administering a simple questionnaire is. The questionnaires we use in this paper to predict behaviour in a public good contribution setting are the Big Six (Lee and Ashton, 2004), a more sophisticated version of the Big Five, and the SVO;<sup>3</sup> therefore, our results are not specific to the use of a single methodology.

Our key finding is a new, and fairly striking, gender effect. While questionnaire measures have predictive power on the average public good contribution of men (at least in the first half of the experiment), women's behaviour cannot be predicted, or can be predicted much less well, on the basis of their questionnaire responses. We attribute this to a greater *context sensitivity* of women (see Gilligan, 1982; Gilligan and Attannucci, 1988; Cadsby and Maynes, 1998), and therefore to the more tenuous nature of the link between questionnaire items and actual public good contribution responses. While this result may be specific to public good contribution settings, it bodes pessimism for the ability of simple questionnaire instruments to capture women's preferences, and the need to think hard of ways to improve on them if they are to be of any use in achieving what they are meant to achieve. Section 2 briefly provides some background. Sections 3 and 4 describe the experimental design and results, respectively. Section 5 provides a discussion and concludes.

## 2. BACKGROUND

### 2.1 *Gender Effects on Contribution*

Gilligan (1982) claimed that women act more in terms of care and connectedness to other people, while men act more in terms of abstract justice, rights and obligations. Faced with a moral dilemma, an agent with a 'care perspective'<sup>4</sup> will consider it on its own merits, in a contextualised fashion that takes into account how the agent is related to the individuals who are involved in the dilemma. As a consequence, women may display different behaviour in different contexts as a function of the contextualised features involved, whilst men will tend to display behaviour that is less context-sensitive and more rule-based.

In public good contribution or similar settings, a common interpretation of Gilligan's work has been that women should contribute more and/or free ride less since they care more about other people (Stockard et al., 1988; Brown-Kruse and Hummels, 1993; Seguino et al., 1996). Empirical results have, however, been surprisingly ambiguous.<sup>5</sup> Boone et al. (1999) and Clark and Sefton (2001) found no statistically significant difference in mean cooperation due to gender in Prisoner's Dilemmas. Kruse and Hummels (1993) found less contribution by women than by men; Rapoport and Chammah (1965) found that women and men contributed as much in the initial round of Prisoner's Dilemmas, but with repetition women contributed less, particular-

ly when playing with other women; in contrast, Nowell and Tinkler (1994) found more contribution in all-female groups in a linear public good contribution experiment. Results are ambiguous and often depend on the treatment (e.g., Solow and Kirkwood 2002, and Stockard *et al.* 1988). In a one-shot public good contribution setting, women contributed more than men (Seguino *et al.* 1996); this is broadly consistent with findings of greater initial cooperation by women that tend to disappear in a repeated setting (Mason *et al.*, 1991; Ortmann and Tichy, 1999; the linear public good contribution treatment of Cadsby and Maynes, 1998). There are many factors that may explain some of the differences in the results, such as the greater risk aversion of women, the use of classroom subjects and more generally the social distance among participants and with the experimenter, the knowledge of the gender of the co-players, and the potential of greater sensitivity by women to being let down.<sup>6</sup>

Does Gilligan's theory really imply that we should expect more cooperation from women in experiments?<sup>7</sup> The alternative interpretation by Cadsby and Maynes (1998) is that women are more sensitive and responsive to others' play, and, as such, they are better able to coordinate around a selected equilibrium.<sup>8</sup> They support this with evidence from a discrete threshold public good contribution experiment, and claim that the linear public good contribution setting should instead lead women to coordinate more quickly to the only available equilibrium which is one with complete free riding. What seems clear is that, in linear public good contribution games, if women *do* tend to cooperate more – other things being equal – this may act as a brake to any tendency to coordinate on an equilibrium of complete non-cooperation, leading to ambiguous predictions.

Greater context-sensitivity does seem to imply, however, that women should be more sensitive to social reinforcement in the form of deviations from the other players' contributions. This prediction, also made by Cadsby and Maynes, is tested in our experiment.

Context-sensitivity also implies that, if women are asked personality questionnaires (only indirectly linked to a behavioural task they are faced with at some other point in time), a lower correlation between measures and behaviour is to be expected. In other words, questionnaires should be less able to predict the behaviour of women than they are of men. We test this hypothesis below.<sup>9</sup>

There has been very little work related to this hypothesis: ours is the first attempt to test systematically the hypothesis of predictability of behaviour, particularly in a strategic setting. On the basis of a post-experiment feedback questionnaire, Brown-Kruse and Hummels (1993) suggested that, although women 'talk' more altruistically, they do not behave as such. They had no formal test of this hypothesis, however.<sup>10</sup> Brown and Taylor (2000) found three times as large a difference between actual and stated difference in contribution to public goods for men than for women, suggesting that any discrepancy between personality measures and behaviour *cannot* be attributed to

a greater tendency of women to misrepresent in words what they want to do. In relation to dictator game data, Ben-Ner *et al.* (2004b) might be interpreted as implying (in their abstract) that women's behaviour is better explained by a Big Five questionnaire than that of men, but this does not appear to be the case if one looks more in detail at their analysis on this point.<sup>11</sup> In current work, Englestad *et al.* (2006) do find predictability of behaviour using a survey instrument (the domain-specific risk attitude scale) for men but not for women in a gambling setting, and their work may be considered in spirit as the most closely related to our paper.

## 2.2 Big Six and Social Value Orientation

*Big Six.* The case for the relevance of personality dimensions for the analysis and prediction of economic behaviour has recently been made by Caplan (2003). Caplan convincingly argued that the developments in personality psychology allow economists to enrich their theories of preferences. In fact, personality dimensions can be considered as preferences, variable (heterogeneously) across individuals while stable within individuals. They can allow for more refined theoretical analyses as well as better predictions of economic choices and interactions. Caplan focused his attention especially on the so-called Big Five.

The Big Five provides a (multi-dimensional) taxonomy of stable human personality characteristics. Several independent studies in a range of countries and languages have repeatedly produced variants<sup>12</sup> of the following personality factors (dimensions), namely (I) Extraversion, (II) Agreeableness, (III) Conscientiousness, (IV) Emotional stability (or Neuroticism), and (V) Intellect or Openness to Experience (for reviews, see De Raad 2000, Goldberg 1993, and Wiggins 1996). New empirical evidence has, however, consistently pointed out that six factors – the so-called Big Six – provides a more comprehensive and cross-culturally generalisable taxonomic structure of basic personality traits than the Big Five (e.g., Ashton and Lee, 2001; Ashton *et al.*, 2004). The differences between Big Six and Big Five include: i) the introduction of a new additional factor referring to the broad domain of morality and integrity that has been labelled Honesty/Humility, and ii) the identification of (slight) variants of Agreeableness and Emotional Stability.<sup>13</sup>

Lee and Ashton (2004) have developed an item questionnaire, the HEXACO-PI, specifically designed to measure the Big Six, in which each of the six factors is identified with four facets (more specific sub-dimensions), as listed in Appendix B. Agreeableness, specifically, has been associated to economic behaviour. Agreeable people are more highly motivated to maintain positive relations with other people (Graziano *et al.* 1996) and adopt a prosocial and communal orientation toward others rather than an antagonistic one (John and Srivastava 1999). Agreeableness is also associated with indicators of altruism (e.g., Ashton *et al.* 1998) because, being motivated to maintain positive relations with others, agreeable persons are likely to help. These motives

have been linked to a willingness to suspend one's personal interest for the good of one's social group (Buss, 1991), and high Agreeableness should be associated with higher public good contribution (Van der Zee and Perugini, 2006).<sup>14</sup> In both the Big Five and the Big Six, the key markers of Agreeableness are aspects such as gentleness, good-naturedness, agreeableness, tolerance, and patience. However, Agreeableness as defined in the Big Six includes content related to even temper, lack of irritability, and sensitivity, whereas these elements are part of the low Emotional Stability factor in the Big Five framework.<sup>15</sup> In the Big Six framework, Agreeableness is defined in terms of four facets with self-explanatory labels: Forgiveness, Gentleness, Flexibility, and Patience.

*Social Value Orientation.* While the Big Six is a state-of-the-art questionnaire from personality psychology, the SVO is in the tradition of social psychology, which has shown the relevance of considering one's social value orientation or preference for particular patterns of distributions of outcomes for self and others (Messick and McClintock, 1968; Van Lange *et al.*, 1997). A distinction has been made between individuals with (1) a prosocial (*cooperator*) orientation (i.e., enhancing joint outcomes and enhancing equality in outcomes); (2) an *individualistic* orientation (i.e., enhancing outcomes for self with no regard to other's outcome); and (3) a *competitive* orientation (i.e., to enhance the relative advantage over another's outcome).<sup>16</sup> Cooperators tend to expect other people to cooperate and perceive social dilemmas as an occasion for cooperative exchanges. The relevance of this orientation for economic behaviour has already been underscored by Offerman *et al.* (1996). In a step-level public good experiment, they found that cooperators (as defined by the SVO) systematically contribute more than individualists, albeit this difference tended to decrease in later contribution periods.

To sum up, two personality factors appear to be the best candidates to predict systematic differences in contribution in a public good environment. The first, Agreeableness, is one of the basic personality dimensions and features both in the Big Five and in the Big Six. Between the two, the cleanest definition of Agreeableness is provided by the Big Six framework. The second, cooperativeness (as defined by the SVO), is a more specific dimension that is not part of the basic personality framework but that nonetheless appears ideally suited for predictions of economic behaviour.

### 3. EXPERIMENTAL DESIGN

The experiment took place in the experimental economics laboratory at the European University Viadrina, Frankfurt-Oder, Germany. A total of 108 subjects, with nine subjects in each session, participated in twelve sessions, six in November 2003 and six in November 2004. Subjects were recruited via email from a database of potential subjects, from the faculties of business administration and economics, cultural science, and law. We had 37 men and

71 women in our sample.<sup>17</sup> None had prior experience of public good experiments. The interactive parts of the experiment were fully computerised. Standard experimental procedures were followed: computer terminals in the laboratory were individually partitioned, preventing subjects from seeing each other's computer screens, or communicating by means of audio, facial, or body signals; neutral phrasings were used, such as 'co-participant' instead of 'co-player', 'group' instead of 'team'; subjects could only participate in one session, and were duly paid at the end of it.

The experiment entailed two *stages*, each with 10 *rounds* of play. Subjects played in anonymously and randomly formed groups of three. Groups were re-matched for stage 2. Groups were composed of the same subjects within each stage. Group members were different across stages. In each round, subjects were allocated 20 'points' (each worth € 0.20) which they could either keep or invest. Each member in the group received 0.5 points for each point invested by any member of the group. Therefore, the design was one of a standard linear public good contribution game with 'partners' (Isaac and Walker, 1988), and, if we label  $x_i$  the contribution of player  $i$ , the payoff to player  $i$  each round was equal to

$$20 - x_i + 0.5 \sum_{j=1}^n x_j \quad (1)$$

Before stage 1 began, subjects completed instruction sheets and control questionnaires to check their understanding. They were allowed to begin only after the experimenters had checked the answers to these control questions; in the event of any incorrect answers, subjects were advised individually. At the top of the screen, the relevant stage, round, and participant number were displayed. At the end of each round, subjects received feedback on the investments and earnings of everyone in the group.

After completing stages 1 and 2, subjects answered a set of psychological questionnaires, inclusive of a Big Six and a SVO questionnaire.<sup>18</sup> At the end of the experiment, the computer informed subjects of their winnings, based on the points earned in two *winning rounds* – one from each stage 1 and 2 – randomly chosen by the computer. They were paid a fixed fee of € 6 for completing the psychological questionnaires. There was no show-up fee. Subjects were informed of the winning rounds for each stage, and their payments after completing all the questionnaires. They were then paid in private. The mean payment was €15.04 for about 1.5 hours of work.

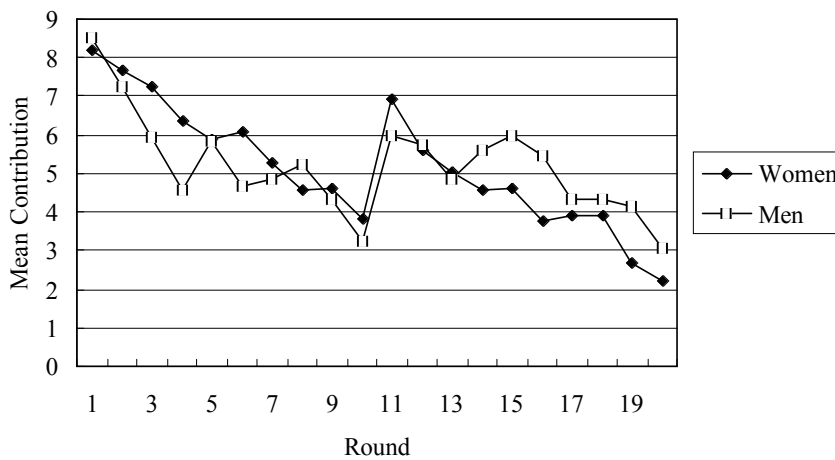
No emphasis was placed on gender in the recruitment process or while running the experiment. All our experimental sessions had both men and women, and subjects had no information about the gender of their co-players.<sup>19</sup> The experiment was conducted in German. English translations of the instructions are found in Appendix A, and of material relevant to the Big Six and SVO questionnaires in Appendix B.<sup>20</sup>

4. EXPERIMENTAL RESULTS

4.1 Behavioural Patterns

Figure 1 plots mean contribution values for men and women participating in the experiment, and Table 1 contains the relevant data. The overall pattern of contributions is commonly observed in similar linear public good contribution games, displaying both a negative trend in contributions through time and a restart effect as subjects started stage 2 (Cookson, 2000). Round 1 contributions were higher in stage 2 for women than for men ( $P < 0.05$  in a Mann-Whitney test), though they were not statistically different in round 1 of stage 1. Overall, in stage 1 women contributed 5.97 tokens on average, about half a point more than men ( $P < 0.1$  in a Mann-Whitney two-tailed test); in stage 2 men contributed slightly more, although not statistically significantly so in a Mann-Whitney test.

**Figure 1: Average contribution by men and women**



**Table 1: Evolution of contributions over time**

Round	Men		Women	
	Stage 1	Stage 2	Stage 1	Stage 2
1	8.18	6.93	8.51	5.97
2	7.68	5.61	7.24	5.76
3	7.28	5.06	5.92	4.84
4	6.37	4.56	4.59	5.59
5	5.90	4.62	5.86	5.97
6	6.10	3.76	4.68	5.46
7	5.28	3.89	4.84	4.35
8	4.55	3.90	5.24	4.35
9	4.61	2.70	4.35	4.16
10	3.80	2.20	3.24	3.05
Mean	5.97	4.32	5.45	4.95

Table 2 reports the results of four of the random effects regressions used to test for whether women converge to the equilibrium prediction of zero contribution more quickly.<sup>21</sup> The dependent variables are, for models 1 and 2, the amount contributed, and, for models 3 and 4 (estimated with a logit regression), a dummy variable equal to 1 if there was zero contribution and 0 otherwise. The independent variables are a gender dummy (equal to 1 for men), a round variable (between 1 and 20), and, in models 2 and 4, the stage number (1 or 2). The stage variable is useful to control for the restart effect.

**Table 2: Regressions of contribution (models 1 and 2) and probability of zero contribution (models 3 and 4)**

<i>Contribution in round</i>				<i>Probability of zero contribution</i>			
Model 1	B	S.E.	p	Model 3	B	S.E.	p
Constant	7.79	0.33	0.00	Constant	-4.17	0.35	0.00
Gender	-1.53	0.50	0.00	Gender	1.23	0.46	0.00
Round	-0.24	0.02	0.00	Round	0.19	0.02	0.00
Gender x Round	0.11	0.03	0.00	Gender x Round	-0.03	0.35	0.30
Model 2				Model 4			
Constant	5.72	0.41	0.00	Constant	-3.31	0.38	0.00
Gender	-1.52	0.49	0.00	Gender	1.25	0.46	0.01
Round	-0.46	0.03	0.00	Round	0.30	0.03	0.00
Gender x Round	0.11	0.03	0.00	Gender x Round	-0.03	0.03	0.31
Stage	2.92	0.35	0.00	Stage	-1.36	0.25	0.00

Notes: Sample size: n = 2160. Random effects regressions were used controlling for both subject level and session level non independence of observations. The log likelihood for models 1, 2, 3, and 4 is -6263.74, -6229.72, -928.77 and -913.63 respectively. All p values provided are two-tailed for the sake of comparability.

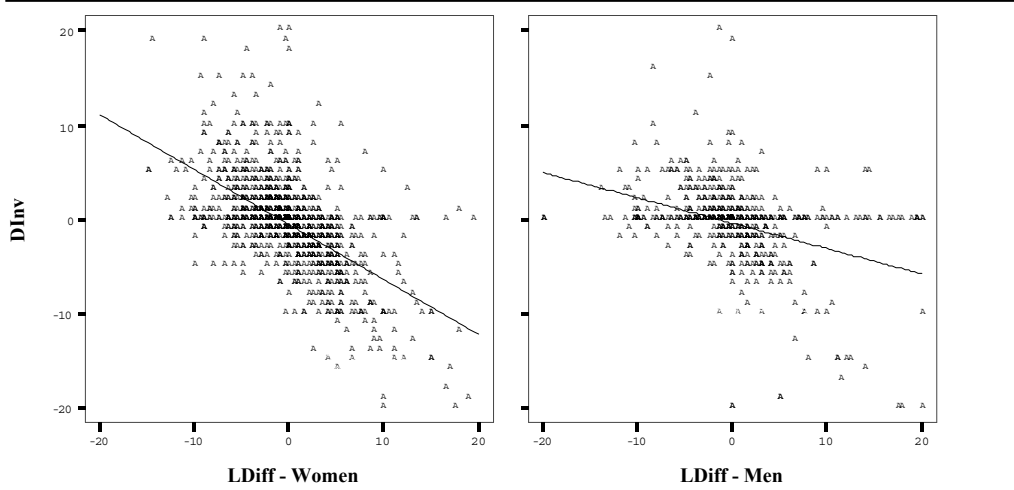
The picture from the regressions (Models 1 and 2) of contributions is one where men contribute significantly less than women in the initial round; subjects contribute significantly less with time, and this effect is more pronounced for women than men. This result holds when we control for the restart effect. The logit regressions (Models 3 and 4), show us that men initially free ride more, and over time the contributions of both genders converge towards the dominant strategy of zero contribution. This result is robust when we control for the restart effect.

Are women more sensitive to social rewards and punishments?<sup>22</sup> The Spearman correlation between the change in contribution per round (*DInv*) and the difference from the mean contribution of the two other players in the previous round (*LDiff*) is equal to - 0.60 for women ( $P < 0.01$ ) but only to - 0.41 for men ( $P < 0.01$ ). Table 3 presents the results of two regressions testing for sensitivity effects.<sup>23</sup> Model 1 is simply a regression of *DInv* on the gen-



der dummy,  $LDiff$  and  $gender \times LDiff$ . Model 2 is a robustness test, which replaces  $LDiff$  with  $LDiff'$  (the difference between own contribution and the mean contribution of all *three* players in the previous round). It gives a similar picture. Figure 2 plots  $DInv$  against  $LDiff$  for men and women. While both men and women increase their contribution if they are socially rewarded, and decrease it if they are punished, a difference in 1 point from the mean contribution of the other (or all) subjects leads to a mean contribution change of -0.35 (or -0.53) points for men, but -0.64 (or -0.97) for women.

**Figure 2: Change in contribution against deviation from other players' mean contribution**



Notes:  $DInv$  = change in contribution relative to previous round.  $LDiff$  = difference of own contribution from the mean contribution of the two other players in the previous round.

**Table 3: Regressions of the change in contribution by gender across rounds**

Model 1	B	S.E.	p
Constant	-0.56	0.16	0.00
Gender	0.16	0.27	0.54
$LDiff$	-0.64	0.03	0.00
$Gender \times LDiff$	0.29	0.04	0.00
<b>Model 2</b>			
Constant	-0.56	0.16	0.00
Gender	0.16	0.27	0.54
$LDiff'$	-0.97	0.04	0.00
$Gender \times LDiff'$	0.44	0.06	0.00

Notes: Sample size:  $n = 1944$ . Random effects regressions were used controlling for both subject level and session level non independence of observations. The log likelihood for both models 1 and 2 is -5338.25.  $LDiff$  = difference of own contribution from the mean contribution of the two other players in the previous round.  $LDiff'$  replaces mean contribution of the other players with the mean contribution of all three players in the group.

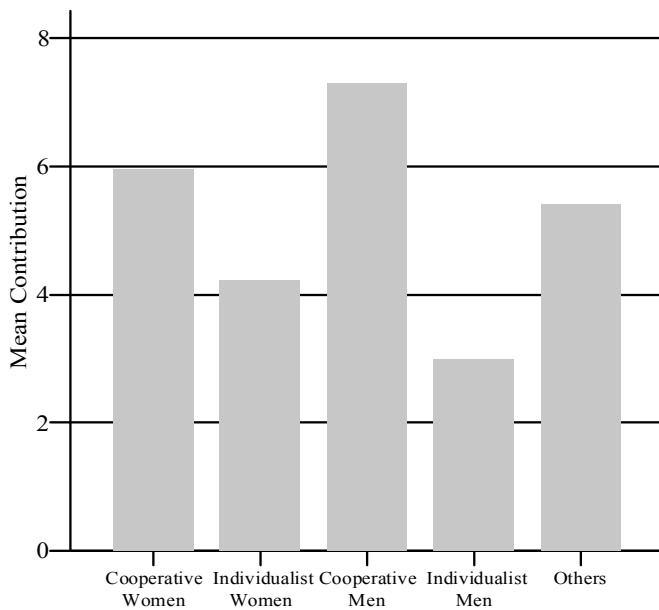
4.2 Personality and Behaviour

*Social value orientation.* Using the method described in section 2, we can classify 47 subjects (44 per cent) as cooperators, 43 subjects as individualists (40 per cent) and 4 (4 per cent) as competitive. Overall, some 87-88 per cent of the subjects' choices were consistent enough to be classifiable in one of these categories. The criterion for consistency is that at least 6 out of the 9 choices can be classified in the same category. The fraction of unclassified subjects is in line with the results found by Offerman *et al.* (1996) and Van Lange *et al.* (1997). Out of 14 subjects that were left unclassified, 13 were women and only 1 was a man: we return to this point below.

As in Offerman *et al.* (1996), given the small number of competitive subjects, our focus is on the comparisons between cooperators and individualists; we have fewer individualists, however, than the 65 per cent of Offerman *et al.*'s sample. In the overall sample cooperators contribute 6.41 points on average against the 3.71 of competitors ( $P < 0.001$ , Mann-Whitney test), seemingly confirming Offerman *et al.*'s finding that SVO can be used to predict behaviour in public good contribution settings.

However, if we analyse the data separately by gender, we find that the predictive power is driven largely by the sample of men (see Figure 3 and Table 4).<sup>24</sup> Men cooperators contributed a mean 7.30 points, over twice as much as men individualists (2.99,  $P < 0.01$  in a Mann-Whitney test); conversely, women cooperators contributed 5.95 points, less than two points more than women individualists (4.22,  $P < 0.05$  in a Mann-Whitney test).

**Figure 3: Mean contribution levels by cooperative and individualist types**



Another way of looking at the predictive power of SVO measures, which avoids discarding any observations from our sample, is to verify whether cooperators' contributions, or individualists' contributions, are different from those of everyone else pooled together. From Figure 3, this is clearly the case for men ( $P < 0.005$  in Mann-Whitney tests) but less clearly so for women (only  $P < 0.1$  in Mann-Whitney tests).

**Table 4: Mean contribution levels**

	Mean Public Good Contribution			Sample Size
	Stage 1	Stage 2	Whole Sample	
Cooperative Women	6.7	5.2	5.95	31
Individualist Women	5.07	3.38	4.22	25
Other Women	5.98	4.09	5.03	15
Cooperative Men	7.77	6.83	7.3	16
Individualist Men	3.16	2.82	2.99	18
Other Men	6.83	7.7	7.27	3

Table 4 also reports mean contribution values in stage 1 and stage 2 of the experiment. There is less predictive power in stage 2 than in stage 1, but men cooperators still appear to contribute significantly more than men individualists, and more than everyone else ( $P < 0.1$  or better in Mann-Whitney tests). Conversely, in stage 2 SVO does not have statistically significant predictive power for women.

In conclusion, the SVO classification method allows us to pick up the behaviour of men much better than that of women: virtually all the subjects whose SVO choices did not lend themselves to classification in a preference type were women, and SVO appears a much better predictor of public good contribution choices with men than with women.

*Big Six.* Table 5 summarises mean values for each of the Big Six dimensions and, in more detail, for the Agreeableness facets.<sup>25</sup> To what extent is it possible to use the Big Six to predict public good contribution for men and for women?

Table 6 reports Spearman correlation coefficients between mean contribution values and the Big Six dimensions and Agreeableness facets. A first finding is that, insofar as there is predictive power of personality variables, this is driven largely by stage 1. A second finding is that, in aggregate and largely in relation to stage 1, more honest and agreeable subjects are likely to contribute more to the public good (see section 2); of the four facets of Agreeableness, Flexibility and Patience have the greatest predictive power, but all four facets are significant at  $P < 0.07$  or better in two-tailed tests or  $P < 0.05$  in one-tailed tests.<sup>26</sup> Spearman  $r$  coefficients, where significant, hover around the 0.2-0.3 range.

As we decompose the analysis by gender we move to our third finding from Table 6: the moderate size aggregate correlation coefficients hide a stark

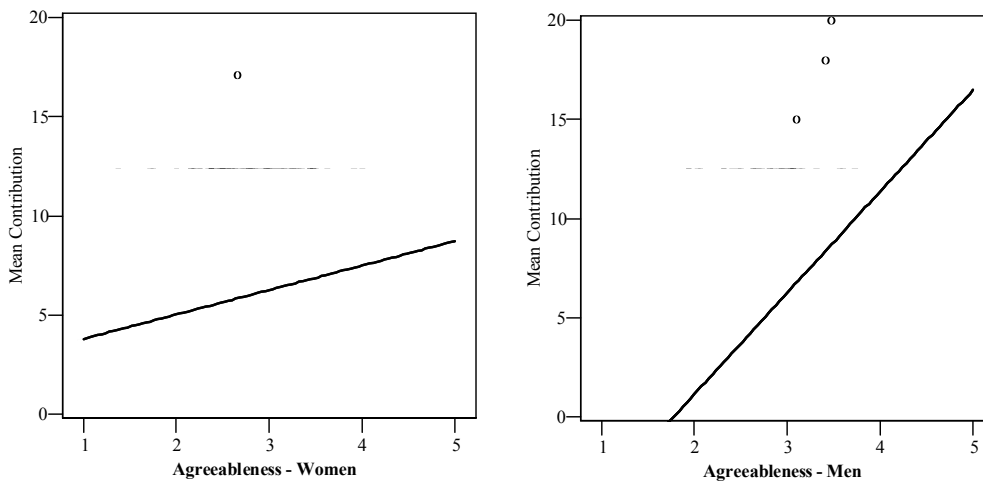
**Table 5: Mean big six and agreeableness facets values**

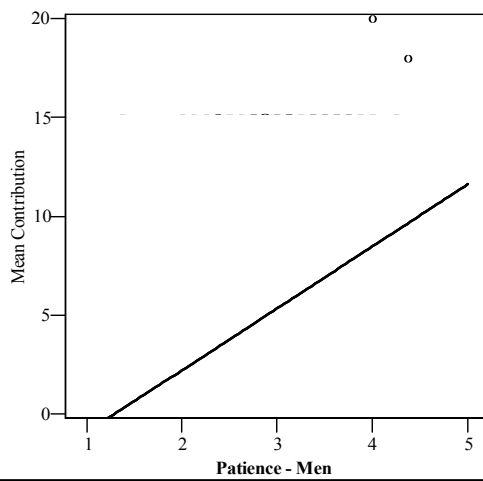
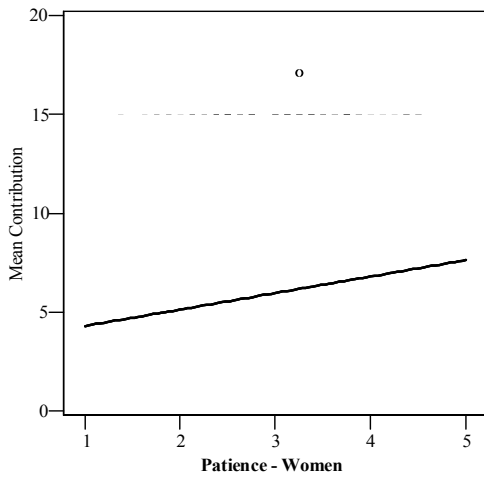
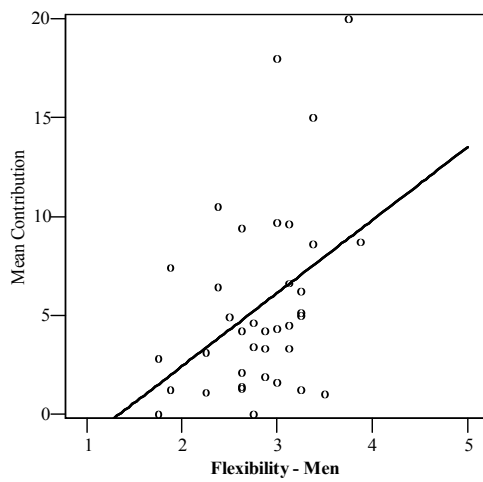
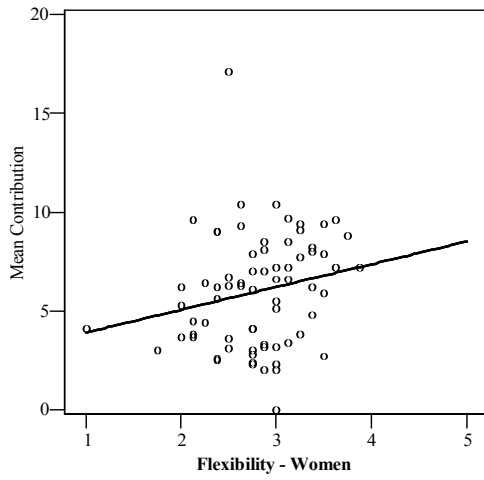
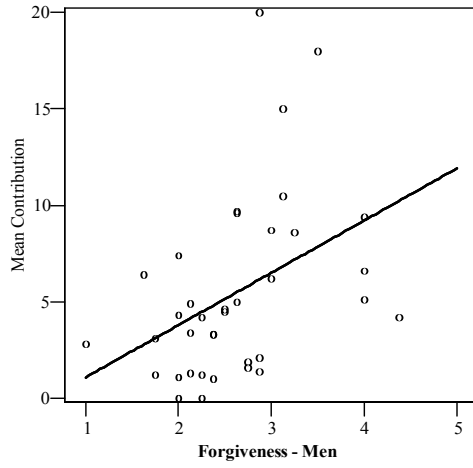
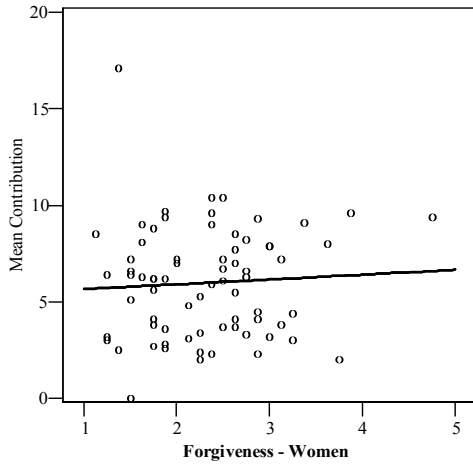
	Women	Men	Total
Agreeableness	2.78	2.84	2.8
Honesty	3.56	3.26	3.45
Emotionality	3.4	3	3.26
Extraversion	3.28	3.48	3.35
Conscientiousness	3.4	3.41	3.4
Openness to Experience	3.39	3.41	3.4
Forgiveness	2.32	2.61	2.42
Gentleness	2.95	2.91	2.94
Flexibility	2.8	2.82	2.81
Patience	3.03	3.04	3.03

Notes: Forgiveness, Gentleness, Flexibility and Patience are facets (aspects) of Agreeableness.

difference between the predictive powers of personality measures for men relative to women. Personality measures can be powerful predictors of the public good contribution of men in the first stage, but they are consistently less useful as predictors of behaviour for women. More forgiving, flexible and patient men are likely to engage in more public good contribution in Stage 1, with Spearman  $r$  coefficients equal to 0.525 ( $P = 0.001$ ), 0.368 and 0.359 ( $P < 0.05$ ), respectively; the overall Agreeableness score has Spearman  $r = 0.496$  ( $P < 0.01$ ) for stage 1. Figure 4 depicts these relationships and contrasts them with the weaker ones found for women.

**Figure 4: Predictability of stage 1 mean contributions using agreeableness measures**





**Table 6: Non-parametric correlations between personality measures and mean contribution**

	Full Sample			Men			Women		
	Total	Stage 1	Stage 2	Total	Stage 1	Stage 2	Total	Stage 1	Stage 2
Agreeableness	0.252***	0.289***	0.088	0.402**	0.496***	0.087	0.183	0.218*	0.093
Honesty	0.134	0.191**	0.034	0.188	0.249	-0.034	0.02	0.07	0.022
Emotionality	-0.027	-0.004	-0.022	-0.057	-0.113	0.013	-0.135	-0.058	-0.132
Extraversion	-0.151	-0.125	-0.15	-0.159	-0.081	-0.218	-0.067	-0.121	-0.055
Conscientiousness	-0.002	-0.045	0.02	-0.188	-0.202	-0.162	0.155	0.093	0.142
Openness to Experience	0.027	0.054	0.054	0.207	0.206	0.167	-0.033	-0.012	-0.009
Forgiveness	0.156	0.201**	0.053	0.287*	0.525***	0.009	0.132	0.084	0.106
Gentleness	0.169*	0.180*	0.061	0.321**	0.229	0.234	0.068	0.18	-0.055
Flexibility	0.194**	0.278***	0.014	0.26	0.368**	-0.21	0.166	0.249**	0.033
Patience	0.288***	0.268***	0.166*	0.349**	0.359**	0.08	0.239**	0.222*	0.19

Notes: \*\*\*:  $P > 0.01$ ; \*\*:  $P < 0.05$ ; \*:  $P < 0.1$ . All correlations in the table are Spearman correlations. All significance levels indicated are two-tailed for the sake of comparability.

**Table 7: Regressions of personality measures on mean contribution in stage 1**

	B	t	p		B	t	p
Constant	6.00	15.19	0	Constant	6.00	14.54	0
Gender	-0.79	-1.17	0.25	Gender	-1.08	-1.51	0.13
Agreeableness	1.23	1.58	0.12	Forgiveness	0.25	0.43	0.66
Gender x Agreeableness	3.89	2.68	0.01	Gender x Forgiveness	2.46	2.52	0.01
$R^2 = 0.17$				$R^2 = 0.11$			
				Constant	5.98	14.74	0
				Gender	-0.58	-0.83	0.41
				Flexibility	1.15	1.46	0.15
				Gender x Flexibility	2.55	1.89	0.06
$R^2 = 0.12$				$R^2 = 0.14$			
				Constant	5.98	14.93	0
				Gender	-0.55	-0.81	0.42
				Patience	0.84	1.57	0.12
				Gender x Patience	2.30	2.33	0.02

Notes: Sample size:  $n = 108$ . All p values provided are two-tailed.

Table 7 reports the results of four regressions conducted to test formally whether the predictive power of the above mentioned personality dimensions was significantly different for men and women in predicting the mean contribution in Stage 1. Each regression included the dummy gender (1 for males), a personality factor, and the interaction term gender by personality.<sup>27</sup> A significant interaction term implies a statistically different relation between

the personality factor and the mean contribution for men and women. Three of the four interaction terms are statistically significant at  $P < 0.05$  (Agreeableness, Forgiveness, Patience), whereas the fourth is marginally significant ( $P < 0.07$ , Flexibility). A simple effect analysis, as depicted in Figure 4, shows that the effects are both consistent and sizeable. For instance, considering Agreeableness, a change of one unit in the score for this trait (range 1-5) implies a change in tokens invested in the public good of 5.12 for a male but only 1.23 for a female; a ratio that is more than 4 times greater for males.

To summarise, we find that, particularly in relation to stage 1, personality dimensions based on Agreeableness have predictive power on the average contribution of men. The same is much less true for women.

## 5. DISCUSSION AND CONCLUSIONS

The aggregate behavioural findings are not particularly surprising, as they only show a slightly higher cooperation rate for women relative to men, especially in the first half of the experiment. Figure 1 also showed some negative trend in contributions, not unlike that found for example in Mason et al. (1991).

Our key finding was that, using personality questionnaires, we could predict the behaviour of men to some extent, especially in the first half of the experiment, but much less that of women. Another dimension in which personality questionnaires seemed to work less well was that, with only one exception, all of the subjects whom we were not able to classify as belonging to a SVO personality type were women. Both findings suggest that personality questionnaires are less useful for economic decision-making with women than with men. These findings have been replicated by Englestad et al. (2006) in the context of a gambling experiment with a different personality questionnaire measure.

A possible criticism may be that our study shows a more general inadequacy of personality measures to predict behaviour. However, this criticism would neglect a considerable body of evidence that supports the empirical validity and usefulness of considering personality dimensions. For instance, hundreds of studies analysed with meta-analytical statistical techniques have provided robust evidence that personality traits play an important role in predicting effective leadership (Bono and Judge, 2004), job performance (Hogan and Holland, 2003; Judge and Ilies, 2002) and job satisfaction (Judge *et al.*, 2002). Within the economic literature, the analysis of the role played by personality traits is still in its infancy, but there are examples that might pave the way for a broader application of personality analysis. For instance, besides the already mentioned contribution of Caplan (2003), where he advocates the adoption of personality traits to model heterogeneous preferences, Bowles et al. (2001) have argued for considering some personality traits as the basis of incentive-enhancing preferences that can usefully be applied to model some specific principal-agent relationships. Personality traits have also started to be

considered in experimental economics. For example, besides the already mentioned work of Offerman *et al.* (1996) with the SVO, the influence of personality traits has been investigated by Kurzban and Houser (2001) in a circular public good setting, by Büchner *et al.* (2007) in a solidarity game, by Brandstätter and Königstein (2001) in an ultimatum game with advanced production, and by Ben-Ner *et al.* (2004b) in a dictator game.

It is, nevertheless, fair to note that the results have not always shown supporting evidence. It is an open question why our results differ from those in Ben-Ner *et al.* (2004b), where the Big Five do not predict the behaviour of men any better than that of women, and less variance appears to be explained by the Big Five anyway. We believe that our results may be less in conflict than may appear at first glance, since they would also have been weaker had we used the Big Five, as Ben-Ner *et al.* did, rather than the Big Six.<sup>28</sup> Our conjecture is that, had Ben-Ner *et al.* used the Big Six, with its cleaner measure of Agreeableness and its facets, they would have found stronger results, as we do.<sup>29</sup> Of course, it is also possible that our results are, in some relevant sense, specific to public good contribution settings.

A second objection is that economists need not be interested in personality as obtained through questionnaire measures but simply in behaviour. This objection fails to recognise that it is exactly behaviour we are interested in and, given the heterogeneity of behavioural attitudes in important settings such as public good contribution or bargaining games, it makes sense to use questionnaires to measure personality dimensions that can help to predict behaviour. Gender is an obvious marker for possible differences in behaviour, and one of considerable importance in the light of economic applications such as labour market differences in wages and career prospects (Ginther and Hayes, 2003), where ‘personal characteristics’ such as cultural background, and tastes for home and work also play a role (Antecol, 2001). Empirical studies have shown that the majority of women can be classified as ‘adaptives’, a term coined by Hakim (1998; 2000) to describe those whose decisions on labour market participation, cast in the context of family planning and activities, are highly dependent on social policy. Indeed, Hakim proposes a dynamic ‘Preference Theory’ that is applicable to the employment choices of women, in general, and in response to such structural changes. Of course, alternative methods to capture heterogeneity of preferences exist and should be used.<sup>xxix</sup> We do not see this issue as an ‘either-or’ choice. Questionnaire-based personality measures should be considered as a complement and not a substitute of other behavioural measures.<sup>31</sup>

One may make sense of the lesser validity of questionnaire measures in the case of women in at least four ways: (1) women are more likely to misrepresent their preferences; (2) women are more likely to be indifferent about questionnaire responses; (3) women are more likely to behave in socially desirable ways, and this may distort responses; (4) women are more likely to be context-sensitive in their preferences. There are various reasons that militate



in favour of the last interpretation, and against the others. First, against the first interpretation, contingent valuation experiments have failed to find any evidence of greater misrepresentation of preferences by women, and, if anything, the study by Brown and Taylor (2000) shows exactly the reverse. Second, against the second interpretation, computation of Cronbach  $\alpha$  values shows that women's Big Six questionnaire responses were as reliable and consistent as those of men,<sup>32</sup> which they would not be if they were simply bored and answered at random or (due to reverse coding of about half of the questionnaires items) if they always chose the same number. Third, mean Agreeableness ratings were no higher for women than for men, showing that there was no gender-specific distortion due to differential social desirability. Finally, the greater context-sensitivity of preferences fits with the evidence of Cadsby and Maynes (1998) and with the finding of greater sensitivity of women to social rewards and punishments by the other players, which we find in our data. They are also consistent with the results of Kurzban and Houser (2001), who were able to classify as conditional cooperators a higher fraction of women than of men in a circular public good game setting. In other words, women were more likely to be sensitive to the other players' choices and cooperate according to the emerging aggregated behaviour. But, of course, being more sensitive to other players' choices implies that one is relatively less prone to follow one's own personal inclination, as emerging from the responses to questionnaire measuring basic personality dimensions. The empirical consequence is that for women these personality measures will be less predictive of behavioural choices.

In short, these reasons converge with our data in supporting an interpretation of women's behaviour in terms of greater contextual sensitivity, at least in a public good contribution setting. Our results imply the need to think hard of ways to design questionnaire instruments to make them more directed towards predicting economic behaviour in relations to *both* women and men.

*Accepted for publication: November 2008*

## APPENDIX A. EXPERIMENTAL INSTRUCTIONS

### **Stage 1 Instructions**

You are about to participate in an experiment on decision-making. The experiment is divided into 2 stages, and a set of questionnaires has to be filled after the two stages. You will be paid your earnings from two *winning rounds randomly chosen by the computer*, one from each stage (1 and 2), plus 6 euros for the set of questionnaires. You will not know which rounds are winning rounds until the end of the experiment.

Stage 1 and 2 consists of 10 rounds of the same task. In each *stage*, the computer will randomly match you and the other experimental participants into 3 *groups*

of 3 participants each, so there will be 2 co-participants in your group for each stage. The other 2 co-participants in your group *stay the same* from round to round during the stage. However, co-participants will change across stages: put it differently, you will be a member of a *different* group in Stage 2 relative to Stage 1. In Stage 2, you will not be matched again with any co-participant that you have been matched with in Stage 1.

**Task:** You are given 20 points. You may decide to invest any amount from 0 to 20 points; in addition, you will keep whatever is not invested. For each point that you and the co-participants from your group invest, every one of you will receive 0.5 points for the total amount invested. In general, you will receive:

$$\text{Payment} = 20 - \text{Your Investment} + 0.5 (\text{Total Investment in your Group})$$

**Example:** *If you invested 9 points, co-participant A invested 11 points, and co-participant B invested 10 points, the total amount invested is 30 points. For this round, you will have 26 points, co-participant A will have 24 points, and co-participant B will have 25 points as the final earnings for the round.*

Once you have decided how much to invest, please type in the number (from 0 to 20) into the text box provided. To confirm your decision, click on Confirm, where a message box will appear asking if you are sure of your decision. If so, click on OK in the message box, and then on Confirm again. If not, you may change your decision either by double clicking on the text box, or clicking on Cancel, and then entering your new decision. You may change your decision anytime before the 2<sup>nd</sup> click on Confirm.

After all your co-participants have made their decisions, you will be told how much each of you invested, and how much each of you earned. Each point that you have earned is worth 0.20 euros. So for the above example, if it is chosen as a winning round, you will earn 5.20 euros for this round. You may then move on to the next round by clicking on Continue.

At the end of the stage, you will be prompted by the computer to call for the experimenter. After reading these instructions, please answer the questions printed on the next page, and call for the experimenter before you begin. Many thanks and good luck!

## APPENDIX B. QUESTIONNAIRES

### a) Social Value orientation (SVO)

In this part of the booklet we ask you to make some choices. You can do this by circling the letters A, B, or C. Your choices will determine the number of points that you and someone else will receive. Imagine that this other person also makes choices in exactly the same task.

*Who is this other person?*

You must try to imagine that this other is a person who you do not know (who you have never met) and that you will never meet (or get to know) this person. So this other person is completely unknown to you.

*What is the meaning of these points?*

The points represent something which is valuable to you: the more you accumulate, the better. The same applies to the other person: the more he or she accumulates, the better for him/her.

An example of this decision task is displayed below.

---

	A	B	C
you get	500	500	550
other gets	100	500	300

---

In this example, if you choose A, you would receive 500 points and the other would receive 100 points; if you choose B, you would receive 500 points and the other would receive 500 points; if you choose C, you would receive 550 points and the other would receive 300 points.

We shall ask you to make nine decisions.

Keep the following in mind:

- There are no right or wrong answers. You choose to circle A, B, or C, depending on which alternative you find most attractive, for whatever reason.
- Imagine that the points are valuable to you: the more you accumulate, the better.
- The same applies to the other person: the more points he or she accumulates, the better for him/her.

*For each of the following nine choice situations, circle A, B, or C, depending on which column you prefer most.*

		A	B	C
(1)	YOU	480	540	480
	Other	80	280	480
		A	B	C
(2)	YOU	560	500	500
	Other	300	500	100
		A	B	C
(3)	YOU	520	520	580
	Other	520	120	320

		A	B	C
(4)	YOU	500	560	490
	Other	100	300	490

		A	B	C
(5)	YOU	560	500	490
	Other	300	500	90

		A	B	C
(6)	YOU	500	500	570
	Other	500	100	300

		A	B	C
(7)	YOU	510	560	510
	Other	510	300	110

		A	B	C
(8)	YOU	550	500	500
	Other	300	100	500

		A	B	C
(9)	YOU	480	490	540
	Other	100	490	300

**HEXACO-PI**<sup>33</sup>

The following list contains descriptions that you can consider more or less true for yourself and with which you can agree or disagree. There are no right or wrong answers. We just ask you to answer how much you agree or disagree with each statement.

The answer choices are:

- STRONGLY DISAGREE**
- DISAGREE**
- NEUTRAL**
- AGREE**
- STRONGLY AGREE**

We would like to remind you:

It is very important that you answer to every description. Please, check at the end whether all questions are answered.

We know that people may behave in a given way in a situation and in another way in a different situation. When deciding whether the description applies to you, you may think "It depends". We therefore ask you to decide whether usually or on average the description applies to you.

Don't take too much time to think about your answers.

*Reported below are one example of items for each facet of each construct. Items preceded by a minus sign between brackets are reverse coded. For a definition of the facets see Lee and Ashton (2004, Table 1). For the full HEXACO-PI questionnaire, please contact Mike Ashton (mashton@brocku.ca) or Kibeom Lee (kibeom@ucalgary.ca).*

**Honesty-Humility**

**Sincerity:**

(-) If I want something from a person I dislike, I will act very nicely toward that person in order to get it.

**Fairness:**

I wouldn't cheat a person even if he or she was a real 'sucker'.

**Greed Avoidance:**

Having a lot of money is not especially important to me.

**Modesty:**

(-) Sometimes I feel that laws should not apply to someone like me.

**Emotionality**

**Fearfulness:**

(-) Where physical pain is involved, I'm a very tough person.

**Anxiety:**

Sometimes I feel nervous without really knowing why.

**Dependence:**

(-) I can 'tough it out' on my own through any kind of personal hardship.

**Sentimentality:**

When someone close to me is concerned about something, I feel concerned too.

**Extraversion**

**Expressiveness:**

My style of speaking is often quite dramatic.

**Social Boldness:**

(-) I rarely express my opinions in group meetings.

**Sociability:**

I prefer jobs that involve active social interaction to those that involve working alone.

**Liveliness:**

(-) People often tell me that I should try to cheer up.

**Agreeableness**

**Forgiveness:**

(-) If someone who has been unkind to me starts being nice, I remain suspicious of that person for a long time.

**Gentleness:**

I tend to be lenient in judging other people.

**Flexibility:**

I often cooperate with other people even when I think their plans don't make sense.

**Patience:**

(-) I react very angrily if I find that someone is trying to cheat me.

**Conscientiousness**

**Organization:**

(-) People often joke with me about the messiness of my room or desk.

**Diligence:**

People sometimes call me a 'workaholic'.

**Perfectionism:**

(-) I don't like to spend time perfecting work that is already good enough.

**Prudence:**

I usually stop myself before doing anything that I might later regret.

**Openness to Experience**

**Aesthetic Appreciation:**

Sometimes I like to just watch the wind as it blows through the trees.

**Inquisitiveness:**

(-) I find TV nature programs to be very boring.

**Creativity:**

(-) I prefer doing things the way I've always done them, rather than waste time looking for a new way.

**Unconventionality:**

I like hearing about opinions that are very different from those of most people.

## ENDNOTES

1. Perugini: Faculty of Psychology, University of Milan-Bicocca, 20126 Milan, Italy; Email: marco.perugini@unimib.it. Tan: Nottingham University Business School, University of Nottingham, Nottingham NG8 1BB, United Kingdom. Email: jonathan.tan@nottingham.ac.uk. Zizzo: School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom. Email: d.zizzo@uea.ac.uk. We thank Nick Bardsley, Roberto Burlando, Robin Cubitt, Francesco Guala, Ed Hopkins, Todd Kaplan, David Kelsey, Tatiana Kornienko, Chris Starmer, Robert Sugden, an anonymous referee and more generally participants at presentations at East Anglia, Exeter, Köln, Stirling and Turin for advice and encouragement, Ramin Farinpour and Jan Heufer for the translations, Jan Heufer, Jessica Kaehler, and Claudia Vogel for the experimental assistance, and Michael Ashton and Kibeom Lee for the HEXACO-PI. The usual disclaimer applies. A previous version of this paper appeared as a Social Science Research Network discussion paper (March 2005). The financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

2. The title of one of the papers in this literature (Andreoni and Vesterlund, 2001: 'Which is the Fair Sex? Gender Differences in Altruism') has provided the inspiration for that of this paper.

3. Our paper is the first to use a measure based on the Big Five to predict public good contribution. It is also the first that uses the Big Six in an economic experiment.

4. The terminology is from Gilligan and Attanucci (1988).

5. The ambiguity of the findings agrees with the complex picture also emerging from dictator and ultimatum games (Andreoni and Vesterlund 2001; Eckel and Grossman 2001; Ben-Ner et al. 2004a, 2004b). For a more complete review than the one provided here see Eckel and Grossman (2001). For a review of psychological studies testing Gilligan's hypothesis about moral reasoning, see Jaffee and Hyde (2000).

6. Eckel and Grossman (2001) have an excellent discussion of some of these factors.

7. One possible problem is that experimental groupings are short-lived affairs, in relation to which the kind of care that women would supposedly be better at would not have the chance to develop: Solow and Kirkwood (2002) address this issue by considering natural groups (in the form of student societies), but, as they carefully acknowledge, their results are then hostage of self-selection bias issues, which may dominate other effects.

8. One way of interpreting this is that women are more likely to follow *conditional* as opposed to unconditional rules, and therefore rules that depend on the specific context at hand.

9. A more general way of interpreting our work is as a test of the consistency between what people claim to be (on paper) and how they actually behave (in the laboratory context).

10. Stockard *et al.* (1988) made a similar point, but again almost entirely informally.

11. The two regressions in their Table 3 report coefficients and standard errors for each of the Big Five personality measures. Out of ten coefficient comparisons, at a glance there are five that are significantly different from each other. Two (their Extraversion coefficients) have relatively large coefficients for men but not for women; for one (Neuroticism in the ordered logit only), the reverse is true; for two (Conscientiousness), the coefficients are of the same magnitude. The reason why the regressions for women are statistically significant while those for men are not is plausibly to be attributed to the non-personality-related variables in the regressions (none of which are close to significance in the regressions for men).

12. Variations in the items that define certain dimensions give rise to such variants.

13. Another strength of the Big Six is that, by considering the sixth factor, the interpretation of the other five factors becomes cleaner than in the original five-factor framework (Ashton *et al.* 2004, 2006).

14. We may interpret this as a stronger positive weight placed on the payoffs of other individuals or the group as a whole in the subject's utility function.

15. Technically, the factorial space between Emotional Stability and Agreeableness in the Big Five is rotated approximately 30 degrees in the Big Six. This rotation implies that some markers of the factors are different in the two frameworks. Another consequence of this rotation is that Emotional Stability is defined as Emotionality in the Big Six framework. This rotation allows a better understanding of the interplay of the different personality factors that are involved in prosocial behaviour, as elegantly shown by Ashton and Lee (2001).

16. This latter orientation is usually relatively infrequent and so we shall not devote much space to discussing it.

17. Because of the demographical composition of the university population, and the fact that we chose not to emphasise gender in the recruitment process (a point we shall return to below), we were not able to get a more balanced sample.

18. This procedure of having subjects answer questionnaires after experimental play is common practice (e.g. Ben-Ner *et al.* 2004a; Brandstätter and Königstein 2001; Büchner *et al.*, 2007). The experimental session also had a questionnaire to probe rec-

iprocity motivation (Perugini et al. 2003) and an alternative questionnaire probing for Big Five Agreeableness (the IPIP questionnaire, 2001). Neither was correlated systematically with behaviour and they are not discussed in this paper.

19. We followed Andreoni and Vesterlund (2001) on this matter. As they note, the lack of emphasis on gender and the use of mixed groups may avoid additional psychological influences (such as male or female bonding) and may reflect more the gender differences that would occur naturally in environments not designed explicitly to test for such differences. Of course, single-gender groups may be interesting in their own right.

20. The original questionnaires were in English. Precautions were taken in the translation procedure to preserve consistency. It was first translated to German (by Person 1), and then back-translated to English (by Person 2), to be compared with the original English version. Together with the authors, discrepancies were discussed and eliminated until consistency was achieved. Persons 1 and 2 are native speakers of German. Person 1 was an Economics undergraduate with study abroad experience in the USA. Person 2 had studied English in school, and holds Bachelors and Masters degrees in Law from a UK university. Neither is an author of this paper.

21. We controlled for the possible non-independence from feedback and repeated interaction within sessions (since subjects were randomly re-matched for stage 2) by controlling for random effects at the levels of subjects nested in sessions.

22. Since payoffs map onto actions, and one's payoff is lower (higher) relative to the other's/group's average if one's contribution is higher (lower), a "more sensitive" subject is by definition one who makes a larger adjustment of contributions in the opposite direction of contributions relative to the other's/group's average contributions in the previous period.

23. Due to the large action set and the small number of repetitions, we restrict our attention to a model with lag of one period.

24. 31 women were classified as cooperators and 16 as individualists; 25 men were classified as cooperators and 18 as individualists.

25. Two-tailed Mann-Whitney tests find statistically significant differences in mean reported values in relation to Honesty ( $P < 0.05$ ), Emotionality ( $P < 0.001$ ), Extraversion ( $P < 0.06$ ) and Forgiveness ( $P < 0.05$ ): women report themselves as more honest, more emotional, less extroverted and less forgiving than men.

26. In congruence with the evidence for the superiority of the Big Six over the Big Five classification (see section 2.2), the Big Six Agreeableness works better than the Big Five Agreeableness in predicting behaviour. Spearman correlation coefficients with mean contribution are 0.123, 0.173 and  $-0.009$  for the whole sample, stage 1 and stage 2 respectively (in relation to stage 1,  $P < 0.1$ ). The Big Five Agreeableness score can be derived from the Big Six data as:  $2 \times (\text{Honesty-Humility}) + 2 \times (\text{Agreeableness}) + \text{Dependence} + \text{Sentimentality} + \text{Sociability} + \text{Liveliness}$  (see Lee and Ashton, 2005).

27. The personality variables have been centred prior to computing the interaction term, so as to eliminate undesirable multicollinearity effects (e.g., Marquardt, 1980).

28. Big Five Agreeableness performs less well on our data than Big Six Agreeableness: see footnote 25.



29. As already mentioned in footnote 25, our experimental dataset adds to a growing body of evidence showing that the Big Six outperforms the Big Five. An alternative or complementary explanation for the discrepancy between our results and Ben-Ner et al. is that their Big Five questionnaire (the NEO-FFI) contains only sixty items to measure all five factors. To put this in perspective, our Big Six questionnaire had 192 items and our measure of Agreeableness alone contained 32 items. This line of explanation of the discrepancy cannot, however, explain why data we have collected on another questionnaire (IPIP Agreeableness, as mentioned in footnote 19) also performed poorly, even though it had 78 items to measure Big Five Agreeableness.

30. For an important recent attempt, see Burlando and Guala (2005).

31. For instance, in our study the role played by personality seems restricted largely to the first half of the experiment, before the feedback and learning from the experiment acquires a greater weight in determining choices. Nonetheless, the interplay between personality traits and learning dynamics appears a very interesting avenue for future research.

32 For example, in relation to Agreeableness  $a = 0.87$  for men and 0.9 for women.

33. Copyright Kibeom Lee and Michael C. Ashton; printed with permission.

#### REFERENCES

Andreoni J and Vesterlund L (2001) 'Which is the fair sex? Gender differences in altruism', *Quarterly Journal of Economics*, 116, 293-312.

Antecol H (2001) 'Why is there interethnic variation in the gender wage gap? The role of cultural factors', *Journal of Human Resources*, 36, 119-143.

Ashton M C and Lee K (2001) 'A theoretical basis for the major dimensions of personality', *European Journal of Personality*, 15, 327-353.

Ashton M C, Lee K, De Vries R E, Perugini M, Gnisci A and Sergi I (2006) 'The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English', *Journal of Research in Personality*, 40, 851-875.

Ashton M C, Lee K, Perugini M, Szarota P, De Vries R E, Di Blas L, Boies K, and De Raad B (2004) 'A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages', *Journal of Personality and Social Psychology*, 86(2), 356-366.

Ashton M C, Paunonen S V, Helmes E and Jackson D H (1998) 'Kin altruism, reciprocal altruism, and the Big Five personality factors', *Evolution and Human Behavior*, 19, 243-255.

Ben-Ner A, Putternam L, Kong F and Magan D (2004a) 'Reciprocity in a two-part dictator game', *Journal of Economic Behavior and Organization*, 53, 333-352.

Ben-Ner A, Putternam L, and Kong F (2004b) 'Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving', *Journal of Economic Psychology*, 25, 581-589.

- Bono J E and Judge T (2004) 'Personality and transformational and transactional leadership: A meta-analysis', *Journal of Applied Psychology*, 89, 901-910.
- Boone C, De Brabander B and van Witteloostuijn A (1999) 'The impact of personality on behavior in five prisoner's dilemma games', *Journal of Economic Psychology*, 20, 343-377.
- Bowles S, Gintis H and Osborne M (2001) 'Incentive enhancing preferences', *American Economic Review*, 91, 155-158.
- Brandstätter H and Königstein M (2001) 'Personality influences on ultimatum bargaining decisions', *European Journal of Personality*, 15, S53-S70.
- Brown K M and Taylor L O (2000) 'Do as you say, say as you do: Evidence on gender differences in actual and stated contributions to public goods', *Journal of Economic Behavior and Organization*, 43, 127-139.
- Brown-Kruse J and Hummels D (1993) 'Gender effects in laboratory public good contribution. Do individuals put their money where their mouth is?', *Journal of Economic Behavior and Organization*, 22, 255-267.
- Büchner S, Coricelli G, and Greiner B (2007) 'Self centered and other regarding behavior in the solidarity game', *Journal of Economic Behavior and Organization*, 62, 293-303.
- Burlando R M and Guala F (2005) 'Overcontribution and decay in public good experiments: A test of the heterogeneous agents hypothesis', *Experimental Economics*, 8, 35-54.
- Buss D M (1991) 'Evolutionary personality psychology', *Annual Review of Psychology*, 42, 459-491.
- Cadsby C B and Maynes E (1998) 'Gender and free riding in a threshold public goods game: Experimental evidence', *Journal of Economic Behavior and Organization*, 34, 603-620.
- Caplan B (2003) Stigler-Becker versus Myers-Briggs: Why preference-based explanations are scientifically meaningful and empirically important', *Journal of Economic Behavior and Organization*, 50, 391-405.
- Clark K and Sefton M (2001) 'The sequential prisoner's dilemma: Evidence on reciprocation', *Economic Journal*, 111, 51-68.
- Cookson R (2000) 'Framing effects in public good experiments', *Experimental Economics*, 3, 55-79.
- Eckel C C and Grossman P J (2001) 'Chivalry and solidarity in ultimatum games', *Economic Inquiry*, 39, 171-188.
- Englestad H, Grossman P J and Lugovskyy O (2006) 'Predicting gamble choice using a domain-specific risk-attitude scale', Saint Cloud State University, unpublished paper.
- Gilligan C and Attanucci J (1988) 'Two moral orientations', in Carol G, Ward J V and Taylor J M (eds) *Mapping the Moral Domain*, Cambridge, MA, and London: Harvard University Press, 73-86.
- Gilligan C (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA, and London: Harvard University Press.

- Ginther D K and Hayes K J (2003) 'Gender differences in salary and promotion for faculty in the humanities', *Journal of Human Resources*, 38, 34-73.
- Goldberg L R (1993) 'The structure of phenotypic personality traits', *American Psychologist*, 48, 26-34.
- Graziano W G, Jensen-Campbell L A and Hair E C (1996) 'Perceiving interpersonal conflict and reacting to it: The case for agreeableness', *Journal of Personality & Social Psychology*, 70, 820-835.
- Hakim C (1998) 'Developing a sociology for the twenty-first century: Preference theory', *The British Journal of Sociology*, 49, 137-143.
- Hakim C (2000) *Work-lifestyle Changes in the 21<sup>st</sup> Century: Preference Theory*, Oxford: Oxford U P.
- Hogan J and Holland B (2003) 'Using theory to evaluate personality and job-performance relations: A socioanalytic perspective', *Journal of Applied Psychology*, 88, 100-112.
- International Personality Item Pool (2001) *A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences*, <http://ipip.ori.org>.
- Isaac M R and Walker J M (1988) 'Group size effects in public goods provision: The voluntary contributions mechanism', *Quarterly Journal of Economics*, 103, 179-200.
- Jaffee S and Hyde J S (2000) 'Gender differences in moral orientation: A meta-analysis', *Psychological Bulletin*, 126, 703-726.
- John O P and Srivastava S (1999) 'The Big Five trait taxonomy: History, measurement, and theoretical perspectives', in Pervin L A and John O P (eds) *Handbook of Personality: Theory and Research* (2nd ed.). New York: Guilford Press, 102-138.
- Judge T A and Ilies R (2002) 'Relationship of personality to performance motivation: A meta-analytic review', *Journal of Applied Psychology*, 87, 797-807.
- Judge T A, Heller D and Mount M K (2002) 'Five-factor model of personality and job satisfaction: A meta-analysis', *Journal of Applied Psychology*, 87, 530-541.
- Kurzban R and Houser D (2001) 'Individual differences in cooperation in a circular public goods game', *European Journal of Personality*, 15, S37-S52.
- Lee K and Ashton M C (2004) 'Psychometric properties of the HEXACO personality inventory', *Multivariate Behavioral Research*, 39, 327-356.
- Lee K and Ashton M C (2005) *HEXACO Personality Inventory Manual*, Unpublished manuscript.
- Marquardt D W (1980) 'You should standardize the predictor variables in your regression models', *Journal of the American Statistical Association*, 75, 87-91.
- Mason C F, Phillips O R and Redington D B (1991) 'The role of gender in a non-cooperative game', *Journal of Economic Behavior and Organization*, 15, 215-235.
- Messick D M and McClintock C G (1968) 'Motivational bases of choice in experimental games', *Journal of Experimental Social Psychology*, 4, 1-25.

Nowell C and Tinkler S (1994) 'The influence of gender on the provision of a public good', *Journal of Economic Behavior and Organization*, 25, 25-36.

Offerman T and Sonnemans J and Schram A (1996) 'Value orientations, expectations and voluntary contributions in public goods', *Economic Journal*, 106, 817-845.

Ortmann A and Tichy L K (1999) 'Gender differences in the laboratory: Evidence from prisoner's dilemma games', *Journal of Economic Behavior and Organization*, 39, 327-339.

Perugini M, Gallucci M, Presaghi F and Ercolani A P (2003) 'The personal norm of reciprocity', *European Journal of Personality*, 17, 251-283.

De Raad B (2000) *The Big Five Personality Factors*, Göttingen: Hogrefe & Huber.

Rapoport A and Chammah A M (1965) 'Sex differences in factors contributing to the level of cooperation in the prisoner's dilemma game', *Journal of Personality and Social Psychology*, 2, 831-838.

Seguino S, Steven T and Lutz M A (1996) 'Gender and cooperative behavior: Economic man rides alone', *Feminist Economics*, 2, 1-21.

Solow J L and Kirkwood N (2002) 'Group identity and gender in public good experiments', *Journal of Economic Behavior and Organization*, 48, 403-412.

Stockard J, Van de Kragt A J C and Dodge P J (1988) 'Gender roles and behavior in social dilemmas: Are there sex differences in cooperation and its justification?', *Social Psychology Quarterly*, 51, 154-163.

Van der Zee K and Perugini M (2006) 'Personality and solidary behavior', in Fetchenhauer D, Flache A, Buunk B and Lindenberg S (eds) *Solidarity and Prosocial Behavior. An Integration of Sociological and Psychological Perspectives*, New York: Springer, 77-92.

Van Lange P A M, De Bruin E M N, Otten W and Joireman J A (1997) 'Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence', *Journal of Personality and Social Psychology*, 73, 733-746.

Wiggins J S (1996) *The five-factor model of personality: theoretical perspectives*, New York: Guilford.