



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

An Inquiry into the Nature and Causes of
Individual Differences in Economics

Sean Brocklebank

Doctor of Philosophy
University of Edinburgh

2012

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own. No other person's work has been used without due acknowledgement. This thesis has not been submitted for any other degree or professional qualification.

Sean Brocklebank

Thesis abstract

The thesis contains four chapters on the structure and predictability of individual differences

Chapter 1. Re-analyses data from Holt and Laury's (2002) risk aversion experiments. Shows that big-stakes hypothetical payoffs are better than small-stakes real-money payoffs for predicting choices in big-stakes real-money gambles (in spite of the presence of hypothetical bias). Argues that hypothetical bias is a problem for calibration of mean preferences but not for prediction of the rank order of subjects' preferences.

Chapter 2. Describes an experiment: Participants were given personality tests and played a series of dictator and response games over a two week period. It was found that social preferences are one-dimensional, stable across a two-week interval and significantly related to the Big Five personality traits. Suggestions are given about ways to modify existing theories of social preference to accommodate these findings.

Chapter 3. Applies a novel statistical technique (spectral clustering) to a personality data set for the first time. Finds the HEXACO six-factor structure in an English-language five-factor questionnaire for the first time. Argues that the emphasis placed on weak relationships is critical to settling the dimensionality debate within personality theory, and that spectral clustering provides a more useful perspective

on personality data than does traditional factor analysis.

Chapter 4. Outlines the relevance of extraversion for economics, and sets up a model to argue that personality differences in extraversion may have evolved through something akin to a war of attrition. This model implies a positive relationship between extraversion and risk aversion, and a U-shaped relationship between extraversion and loss aversion.

Contents

1	Predicting Risky Behavior: a Re-analysis of Holt and Laury's data	18
1.1	Introduction	19
1.2	Method	22
1.2.1	Participants and demographics .	22
1.2.2	Measures	23

1.2.2.1	The Holt-Laury Task	23
1.2.2.2	Other Measures	26
1.3	Results	27
1.3.1	Descriptive statistics and hypothetical bias	27
1.3.2	Comparing predictive power	27
1.3.3	Restricting the sample	28

1.3.4	The (un)importance of chronology	29
1.3.5	Changing the variable from switchpoints to safe choices	30
1.4	Discussion	30
2	Personality accounts for stable preferences and expectations across a range of simple games	39
2.1	Introduction	41

2.1.1	Personality and Social Preferences	41
2.1.2	A Brief Review of Related Liter- ature	44
2.1.3	The Current Study	46
2.1.4	The Prosocial Ori- entation Scale .	48
2.2	Method	49
2.2.1	Participants . .	49
2.2.2	Measures	49
2.2.2.1	Person- ality .	49

	2.2.2.2	Games	49
2.3	Results	54
	2.3.1	Factor structure of social prefer- ences	54
	2.3.2	The PO Scale: re- liability	56
	2.3.3	Personality as a Predictor of Proso- cial Orientation	57
2.4	Implications for modelling: traits instead of types .		57

2.4.1	Charness and Ra-	
	bin (2002) . . .	59
2.4.2	Fehr and Schmidt	
	(1999)	64
2.5	Discussion	65
3	Contrasting the FFM and HEX-	
	ACO models of the structure of	
	personality: A spectral cluster-	
	ing approach	73
3.1	Introduction	75
3.2	Method	79
3.2.1	Measures	79

3.2.2	Participants and procedure . . .	79
3.3	Results	79
3.3.1	Network construc- tion and the im- portance of weak connections . . .	80
3.3.2	Spectral cluster- ing algorithm .	82
3.3.3	Parameter esti- mation	83

3.3.3.1	The num- ber of eigen- vectors	83
3.3.3.2	The scale param- eter . . .	84
3.3.3.3	The num- ber of clus- ters . . .	86
3.3.4	Comparison of do- mains from spec- tral clustering and factor analysis . . .	92

3.4	Discussion	94
3.4.1	Main findings .	94
3.4.2	Differences between the six- and five- cluster solutions: Psychological im- plications . . .	96
3.4.3	Differences between the six- and five- cluster solutions: Economic impli- cations	97

3.4.4	Strengths and Weak-	
	nesses	97
3.4.5	Summary	98
4	Personality as Strategy: The	
	Evolution of Extraversion	104
4.1	Introduction	105
4.1.1	The evolution of	
	individual differ-	
	ences	105
4.1.2	Strategic interac-	
	tion and person-	
	ality differences	106

4.1.3	The war of attrition as a metaphor for conflict . . .	108
4.1.4	Extraversion as a simple decision rule	109
4.2	The nature of extraversion	110
4.3	The model	112
4.3.1	Resources increase fertility	112
4.3.2	Standard simplifying assumptions	113

4.3.3	Mechanics of the war of attrition	114
4.3.4	The basic model is a simple deci- sion rule	117
4.3.5	Extraverts take on more risk . .	119
4.3.6	Asymmetry . .	121
4.3.7	Asymmetries and loss aversion . .	127
4.3.8	Adding more play- ers	131
4.4	Conclusion	133

Acknowledgements

First and foremost, I would like to thank my supervisors Ed Hopkins and Tatiana Kornienko for their guidance and support, and for putting up with my unconventional subject matter and (sometimes) unnecessary wanderings. I'd like to thank Tim Bates and Gary Lewis for teaching me so much about what it means to do good research and write a good paper, albeit in the “wrong” field. Scott Pauls, Greg Leibon and Dan Rockmore gave me a better way of thinking about statistics and structure in data and they taught me more math than I care to admit along the way. Kohei Kawamura and Miguel Costa-Gomes conducted my PhD defence in a more entertaining and humane way than I had any right to expect—I suspect that few PhD students enjoyed defending their thesis as much as I did.

Chapter 1

Predicting Risky Behavior: a Re-analysis of Holt and Laury's data

This paper argues that choices with large but hypothetical payoffs may be more useful for predicting subjects' important decisions than the small-stakes, real-money payoffs typically used in economics experiments. We use data from Holt and Laury (2002) and show that big-stakes, real-money choices are more accurately predicted with earlier big-stakes hypothetical choices than with earlier small-stakes, real-money choices. This result holds in spite of the presence of a hypothetical bias which causes subjects to report less risk aversion in hypothetical treatments. It is argued that hypothetical bias should be thought of as a mean level bias rather than a rank order bias. Experimenters who are interested in calibrating the mean level of subjects' risk preferences are recommended to avoid hypothetical payoffs, but experimenters who wish to control for individual variation are recommended to use hypothetical payoffs.

1.1 Introduction

The use of hypothetical payoffs within economic experiments is controversial. Hypothetical payoffs did receive early support from Kahneman and Tversky (1979):

Experimental studies typically involve contrived gambles for small stakes, and a large number of repetitions of very similar problems. These features of laboratory gambling complicate the interpretation of the results and restrict their generality. By default, the method of hypothetical choices emerges as the simplest procedure by which a large number of theoretical questions can be investigated. The use of the method relies on the assumption that people often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reason to disguise their true preferences.

But recent authors have been more skeptical. Both Laury and Holt (2008) and Harrison (2007) are somewhat hostile to the use of non-monetary payoffs, suggesting that they result in hypothetical bias. Hypothetical bias is the tendency for subjects to report themselves to be less risk averse with hypothetical payoffs than they are with equivalent real-valued payoffs. Both papers argue that because average reported preferences differ systematically in real vs. hypothetical treatments, researchers should use only real payoffs. In a similar vein, Holt and Laury (2002) write that their results (which exhibit hypothetical bias)

[Raise] questions about the validity of Kahneman and Tversky's suggested technique of using hypothetical questionnaires to address issues that involve very high stakes. In particular, it casts doubt on their assumption that "people often know how they would behave in actual situations of choice".

These arguments, however, are directed primarily at researchers who are attempting to estimate the mean level of risk preference (e.g. to calibrate a model for a representative agent). But measures of risk preferences are not only used to calibrate averages—they are sometimes used to control for individual variation. A researcher may want to control for individual variation in risk attitudes when subjects play a game that involves an element of risk, or when testing the relationship between risk preferences and other choices such as health behaviors or asset allocations. In these sorts of situations, researchers are interested in the rank order of subjects' preferences but not the mean level. Here, hypothetical bias is not a problem if it merely introduces a constant bias into all subjects' reported levels of risk aversion. As long as the rank order of subjects' risk preferences is preserved then risk attitudes elicited by hypothetical payoffs will still be strongly correlated with those elicited by real payoffs.

If hypothetical tasks can provide a valid method to control for risk attitudes, it is clear that they have several advantages over real-money tasks. Not only do hypothetical tasks save the financial costs associated with paying out money and the administrative costs of creating a system to pay out money, but they can also exist on a more appropriate scale than real payoffs. Economists are often interested in how people make decisions when risks are substantial: getting education, buying a house, buying insurance, saving for retirement, choosing a job, starting a business, et cetera; but economists are generally constrained to design risk-based experiments on a much smaller scale—generally only a few dollars rather than the thousands or even hundreds of thousands of dollars at stake in the real-life decisions. But hypothetical tasks are not similarly constrained: it is possible to ask people how they would behave with stakes as large as the researcher desires. This matters because subjects may exhibit different risk attitudes at different payoff scales, and

so small-stakes real-money payoffs may be tapping risk preferences at the “wrong scale” if the researcher is really interested in bigger choices as listed above. Some evidence from the importance of scale comes from Holt and Laury (2002), who show that their data are inconsistent with the constant relative risk aversion model and conclude that subjects’ preferences are not the same across scales. The analysis in section 1.3 will reinforce this conclusion.

So: can researchers use hypothetical tasks to control for risk attitudes? Do such tasks preserve the rank order of subjects’ true preferences, if not the average? To my knowledge, this is the first paper which addresses the question directly, though there are a small number of papers which deal with it indirectly. Andersen and Mellor (2009) show that hypothetical-payoff risk attitudes are only weakly related to real-payoff risk preferences and they also note that in previous literature it is the hypothetical-payoff preferences which have been better validated against real-world behaviors such as financial investment, insurance demand, risky health behaviors, education, marriage, and fertility. Similarly, Guiso and Paiella (2005) show that occupational choice and moving behavior are further real-world correlates of hypothetical risk preferences.

Thus there is some evidence to suggest that hypothetical payoffs are at least as good as real payoffs for sorting subjects by their risk preferences in order to predict real world behaviors. But to my knowledge there are no head-to-head tests in which big-stakes hypothetical payoffs compete against small-stakes real payoffs to predict choices with big-stakes real payoffs. That is the object of this study.

The data for the current study are those collected by Holt and Laury for their 2002 paper “Risk aversion and incentive effects”. The precise details of their procedure are given in section 1.2.2.1, but for now it will suffice to give a general outline of their procedure: the Holt-Laury experiment involved completing a series of ten binary

gambles in each of four separate tasks. Tasks were completed in the same sequence for all subjects, and most subjects completed all four tasks (some subjects only completed three tasks, but they are omitted from the present analysis). Each task has the same general structure, but the payoffs in the different tasks may be real or hypothetical, and they exist at different scales. Particularly, Task 1 is small-stakes with real money, Task 2 is big-stakes but hypothetical, Task 3 is big-stakes with real money, and Task 4 is just like Task 1. The exact numbers used in each task are shown in Tables 1.1 and 1.2. For the current analysis, the most important feature of Holt and Laury’s design is that the third task involved amounts which are large relative to undergraduate earnings (subjects could win up to \$346.50 on this task), and that it was preceded by one task with hypothetical money on the same scale and by another task with real money on a much smaller scale (maximum winnings on Task 1 were \$3.85). Thus we can directly test if choices in big-stakes real money gambles are better predicted by choices in big-stakes hypothetical gambles or by choices in small-stakes real-money gambles. This analysis is described in section 1.3, but to preview the result: the hypothetical payoffs have greater predictive power.

The rest of this paper is structured as follows: section 1.2 gives more detail on the experimental set up, section 1.3 describes the analysis, and section 1.4 concludes.

1.2 Method

1.2.1 Participants and demographics

Holt and Laury (2002) report data on 212 subjects. Of these, 130 were assigned to complete all of the experimental tasks. Only these subjects were used for the present analysis. The average year of birth of the subjects was 1975, with a standard deviation of six years. Sixty-nine of the subjects (53%) were men; further demographic

information can be found in the associated data file.

1.2.2 Measures

1.2.2.1 The Holt-Laury Task

Holt and Laury's (2002) procedure measures subjects' risk preferences through a series of four tasks, each of which contains ten binary choices between lotteries. Each binary choice is between a "safe" option (A) and a "risky" option (B). One example binary choice would be the following:

Option A (safe)

- a 60% chance of getting \$2, and a 40% chance of getting \$1.60

Option B (risky)

- a 60% chance of getting \$3.85, and a 40% chance of getting \$0.10

Notice that the probabilities are the same in the safe and risky options but that the variance of payoffs is larger for the risky option; this feature is true for each of the HL binary choices. For the example above, the expected value of Option A is \$1.84 and the expected value of Option B is \$2.35. Thus an expected-payoff maximizing subject would choose the risky option, but a subject who was more concerned with minimizing the variance of possible outcomes might make the safe choice.

Table 1.1 presents the ten binary choices presented to subjects in the first task. Subjects were not, however, shown the final "expected payoff difference" column of that table. The example above corresponds to the sixth row of Table 1.1. As one proceeds down the table, the expected value of both A & B increases as the likelihood of the better outcome within each pair rises, but the expected value of the risky option rises faster than the expected value of the safe option.

Table 1.1: The Ten Paired Lottery Choices with 1x Payoffs used in Tasks 1 & 4

Option A (safe)		Option B (risky)		Expected payoff difference
1/10 of \$2	9/10 of \$1.60	1/10 of \$3.85	9/10 of \$0.10	\$1.17
2/10 of \$2	8/10 of \$1.60	2/10 of \$3.85	8/10 of \$0.10	\$0.83
3/10 of \$2	7/10 of \$1.60	3/10 of \$3.85	7/10 of \$0.10	\$0.50
4/10 of \$2	6/10 of \$1.60	4/10 of \$3.85	6/10 of \$0.10	\$0.16
5/10 of \$2	5/10 of \$1.60	5/10 of \$3.85	5/10 of \$0.10	-\$0.18
6/10 of \$2	4/10 of \$1.60	6/10 of \$3.85	4/10 of \$0.10	-\$0.51
7/10 of \$2	3/10 of \$1.60	7/10 of \$3.85	3/10 of \$0.10	-\$0.85
8/10 of \$2	2/10 of \$1.60	8/10 of \$3.85	2/10 of \$0.10	-\$1.18
9/10 of \$2	1/10 of \$1.60	9/10 of \$3.85	1/10 of \$0.10	-\$1.52
10/10 of \$2	0/10 of \$1.60	10/10 of \$3.85	0/10 of \$0.10	-\$1.85

Note: for 20x, 50x and 90x treatments, the payoffs are scaled to:

20x: Option A is \$40 or \$32 Option B is \$77 or \$2

50x: Option A is \$100 or \$80 Option B is \$192.50 or \$5

90x: Option A is \$180 or \$144 Option B is \$346.50 or \$9

Thus, an expected-value-maximizing risk-neutral subject would pick the safe Option A for the first four rows and then switch over to the risky Option B for the final six rows. Risk averse subjects might switch from safe to risky at a later point, but the general pattern of choosing Option A initially (since the expected payoff is higher and the variance is lower) and Option B later (since, by the tenth row, there is no uncertainty and a higher payoff) should hold for all subjects.

Subjects were required to indicate a preference for Option A or Option B for each of the ten paired choices in Table 1.1. Subjects were informed ex ante that one of the choices would be randomly selected ex post and played out in order to determine the earnings for that task. Holt and Laury note that while we might expect incentives to be diluted by the use of only random selection of a single decision for each task, that in fact subjects did appear “to take even the low-payoff condition seriously, often beginning with the easier choices at the top and bottom of the table, with choices near their switch point more likely to be crossed out and changed” (Holt and Laury, 2002).

The four tasks of the HL experiments were all based around this framework. The only difference from task to task is in the scale of the payoffs and whether or not the payoffs were hypothetical. The scale of the payoffs was varied as a between-subjects treatment, with some subjects being offered potential payoffs 20 times those presented in Table 1.1, some being offered payoffs 50 times those presented in Table 1.1, and others being offered payoffs 90 times those in Table 1.1. The real vs. hypothetical payoff distinction was a within-subjects variable, so that subjects faced three real payoffs and one hypothetical payoff. The four tasks were as follows:

Task 1: real payoffs equal to 1x the values in Table 1.1

Task 2: hypothetical payoffs equal to either 20x, 50x, or 90x the values in Table 1.1

Task 3: real payoffs, matched to those offered in Task 2

Task 4: real payoffs equal to 1x the values in Table 1.1 (same as Task 1)

The outcome of each task was determined by rolling a ten-sided die in front of the subject before the next task began. Note that this could create an uncontrolled source of variation due to wealth effects from early tasks, but Holt and Laury controlled for this by requiring subjects to give up their winnings from Task 1 in order to be allowed to participate in Task 3. Since the scale of payoffs is so much higher in Task 3, all subjects agreed to this condition, and there was therefore no selection bias. There was of course no similar requirement for participation in Task 4, since subjects would not have been willing to forgo their large earnings from Task 3¹.

1.2.2.2 Other Measures

There is more than one way to measure subjects' behavior in these tasks. One method is to count the total number of safe choices made in each task. Another method is to record the number of safe choices made by subjects before they first switch over to choosing the risky option. Ideally, these two measures would be identical. Alas, experimental subjects are not all rational agents with and well-behaved preferences, so these measures are not the same and we must decide which to use. Most of the present study will refer to switchpoints, but the robustness of the results to using the average number of safe choices is considered in section 1.3.5.

¹One potential problem with the results from Holt & Laury (2002) is that at each of the three campuses where experiments were run, they took place over more than one day, so there is a risk that later subjects heard about the experimental details from the earlier subjects and were thus contaminated. If this were a problem, it could show up by adding noise to the responses given in Task 1 by the contaminated subjects (since they would know that the task doesn't count for anything). Unfortunately, data on which subjects participated on which days is not available, but I tested for possible contamination by splitting the subjects into "first half" and "second half" and checking to see if the results are stable, and I found that they are.

1.3 Results

1.3.1 Descriptive statistics and hypothetical bias

Table 1.3 shows the average switchpoint by task and by treatment. The most important feature of these data for present purposes is the appearance of hypothetical bias: subjects made an average of 4.8 safe choices before their switchpoint in Task 2, but they made 6.1 safe choices before their switchpoint on Task 3, even though Tasks 2 and 3 had payoffs of exactly the same nominal size. Furthermore, note that hypothetical bias increases with the payoffs: there were an extra 0.9 safe choices in Task 3 for the 20x treatment, but 2.0 extra safe choices in the 50x treatment and 2.2 extra safe choices in the 90x treatment. The fact that this gap trends upwards with the scale of the payoffs lends credence to the idea that it is indeed hypothetical bias.

But it's worth re-stating the argument made in section 1.1: hypothetical bias would clearly be a problem if we wanted to use Task 2 switchpoints to predict the average switchpoint in Task 3, but the bias is not necessarily a problem if we merely wish to predict the rank order of our subjects' switchpoints in Task 3.

1.3.2 Comparing predictive power

Here we will test which is the better predictor of the Task 3 (big, real) switchpoint—the Task 1 (small, real) switchpoint, or the Task 2 (big, hypothetical) switchpoint. Implicitly, we are testing whether it is more important for the predictor to be similar in the size dimension or similar in the real/hypothetical dimension.

Note that all estimates reported in Tables 1.4 through 1.11 are based on standardized coefficients (variables are normalized to have mean=zero, variance=one), so the coefficients can be interpreted as partial correlations.

Table 1.4 reports three models which predict the Task 3 switchpoint as a function

of earlier choices. Model 1 uses only Task 1 as a predictor (as well as dummies to control for treatment—these are also used in Models 2 and 3), Model 2 uses only Task 2 as a predictor, and Model 3 uses both Tasks 1 and 2. By comparing Models 1 and 2, we can see that—separately—both of the earlier switchpoints are significant predictors of the Task 3 switchpoint, and that the Task 2 has a higher partial correlation (.55 vs .42) and a higher R-squared (.39 vs .26). The difference in residual variances, however, is not significant (Bartlett’s test of equal variances: $\chi^2 = 12.6$, $p = 0.76$). Model 3 allows us to see how the Task 1 & 2 switchpoints perform when they are place head-to-head, and the result is again that Task 2 has a higher partial correlation (.47 vs .16) and greater statistical significance. In order to test if one of the coefficients is greater, we can conduct a test of the null hypothesis that Task 1 is a better predictor (i.e. that $\beta_1 > \beta_2$, which would be true if hypothetical bias were a problem for prediction) against the alternative that Task 2 is a better predictor (i.e. that $\beta_1 < \beta_2$). The resulting (one-tailed) p-value is 0.02 and the null hypothesis is thus rejected at the 5% level, lending support to the notion that the hypothetical task is a better predictor.

1.3.3 Restricting the sample

Table 1.5 repeats the structure of Table 1.4, but with the sample restricted to the 18 subjects who participated in the extremely high (90x) treatment. This comparison is interesting because subjects in the 90x treatment were those for whom the difference in the scale of the payoffs was largest and—as noted in section 1.3.1—they were the subjects with the largest hypothetical bias. And yet Table 1.5 shows results which are similar to but even stronger than those in Table 1.4. When the sample is restricted to subjects in the 90x treatment, the Task 1 switchpoint is no longer even a significant predictor either on its own or in combination with the Task 2

switchpoint. This result reinforces the core argument of this paper, since it implies that the hypothetical predictor is superior to the real predictor in spite of the clear presence of hypothetical bias in the data.

1.3.4 The (un)importance of chronology

One possible objection to the above result is that Task 2 was a better predictor than Task 1 merely for chronological reasons. Task 1 always came before Task 2, and perhaps subjects “got used to” the game and so the Task 2 switchpoints are more reliable and are a better predictor for that reason. One way to test this is to compare Task 1 and Task 2 switchpoints as predictors for the switchpoint on Task 4. If Task 2 was a superior predictor of Task 3 behavior only because Task 2 was later than Task 1, then Task 2 should also be a superior predictor of Task 4. If, however, the similarity of tasks is more important than chronology, then we would expect that the Task 1 switchpoint would be a better predictor of the Task 4 switchpoint.

Tables 1.6 and 1.7 replicate the structure of Tables 1.4 and 1.5 (respectively) but with the Task 4 switchpoint replacing the Task 3 switchpoint as the dependent variable. Contrary to the “mere chronology” hypothesis, the patterns in Tables 1.6 and 1.7 are exactly reversed from Tables 1.4 and 1.5. In Table 1.6 we see that the Task 1 switchpoint is a better predictor of the Task 4 switchpoint, and that this is true both singly (comparing Models 1 and 2, we see partial correlations of .61 vs .51 and R-squareds of .38 vs .27) and in combination (Model 3 shows partial correlations of .47 vs .25). Unlike the case above, the test of the null hypothesis that Task 1 is a better predictor is not rejected (the p-value is .93). Table 1.7 shows that the results in Table 1.6 are preserved when we restrict our sample to the subjects in the 90x treatment.

1.3.5 Changing the variable from switchpoints to safe choices

As a check on the robustness of the above results, tables 1.8-1.11 repeat (respectively) all of the tests in tables 1.4-1.7 but with the total number of safe choices for each subject replacing the switchpoint as the variable of interest (recall from above that these should be equal for well-behaved decision makers). Happily, the signs and relative magnitudes of the coefficients are identical under the both variations (this was probably inevitable given the high correlation between switchpoints and safe choices, but it is reassuring nonetheless).

1.4 Discussion

The main finding of this paper is that choices on Task 2 (big-stakes, hypothetical) were a better predictor of choices on Task 3 (big-stakes, real money) than were choices on Task 1 (small-stakes, real money), suggesting that it is more important for payoffs to be of the appropriate scale than for them to be backed by hard cash. This result holds in spite of the presence of hypothetical bias, and does not appear attributable to the sequencing of the experiment.

This result, of course, is subject to all of the standard external validity criticisms. The present data concern only one method for risk elicitation (albeit a very popular one), and the participants were mainly university students. One would like to see replications of the result before believing too strongly that hypothetical payoffs dominate real money payoffs in all predictive applications.

But one can be more confident believing that at least hypothetical payoffs are not significantly worse than real money payoffs for making predictions. Given the much lower costs of administering hypothetical questions, this suggests that they deserve wide use among researchers who are more interested in the rank order of

their subjects' risk preferences than the average value in the population.

Summing up, there are three closely related takeaway messages from this paper. First: when evaluating the usefulness of hypothetical payoffs, it is important to distinguish between calibration (of averages) and prediction (of rank order). Second: for predictive purposes, hypothetical payoffs of the appropriate scale appear somewhat better than real payoffs on a small scale, hypothetical bias notwithstanding. Third: given this evidence and the cheapness of administering them, hypothetical choices deserve wider use (i.e. Kahneman and Tversky had it right).

Table 1.2: The Four Holt & Laury tasks

	Hypothetical?	Scale of payoffs (relative to Table 1.1)
Task 1	No	1x
Task 2	Yes	20x, 50x, or 90x (depending on treatment)
Task 3	No	20x, 50x, or 90x (same as Task 2)
Task 4	No	1x

Note: there were 130 subjects; 93 in the 20x treatment, 19 in the 50x treatment, and 18 in the 90x treatment.

Table 1.3: Mean Safe Choices Before Switching to Risky, by Treatment

	All	20x	50x	90x
Task 1 (small, real)	4.9 (1.6)	4.8 (1.6)	4.9 (1.5)	5.1 (1.3)
Task 2 (big, hyp)	4.8 (1.8)	4.8 (1.9)	4.6 (1.8)	5.0 (1.7)
Task 3 (big, real)	6.1 (1.9)	5.7 (1.9)	6.6 (1.2)	7.2 (1.7)
Task 4 (small, real)	5.2 (1.4)	5.1 (1.4)	5.3 (1.2)	5.4 (1.3)
n	130	93	19	18

Note: Standard deviations in parentheses.

Table 1.4: Switchpoint on Task 3 (high, real)

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.42*** (.08)		.16* (.08)
Task 2 switchpoint		.55*** (.07)	.47*** (.08)
Dummies for treatment	yes	yes	yes
R-squared	.26	.39	.41

Note: n=130. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=4.42$, $p=0.02$.

Table 1.5: Switchpoint on Task 3 for 90x only

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.35 (.23)		.19 (.17)
Task 2 switchpoint		.74*** (.17)	.69*** (.17)
R-squared	.15	.54	.58

Note: n=18. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=3.41$, $p=0.04$.

Table 1.6: Switchpoint on Task 4 (small, real)

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.61*** (.08)		.47*** (.08)
Task 2 switchpoint		.51*** (.08)	.25*** (.08)
Dummies for treatment	yes	yes	yes
R-squared	.38	.27	.42

Note: n=130. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F = 2.25$, $p = 0.93$.

Table 1.7: Switchpoint on Task 4 for 90x only

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.64*** (.19)		.59*** (.19)
Task 2 switchpoint		.38 (.23)	.24 (.19)
R-squared	.41	.12	.47

Note: n=18. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F = 1.27$, $p = 0.86$.

Table 1.8: Number of safe choices on Task 3 (high, real)

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.38*** (.08)		.18** (.08)
Task 2 switchpoint		.49*** (.07)	.40*** (.08)
Dummies for treatment	yes	yes	yes
R-squared	.23	.33	.36

Note: n=130. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=2.4$, $p=0.06$.

Table 1.9: Number of safe choices on Task 3 for 90x only

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.21 (.24)		.16 (.21)
Task 2 switchpoint		.58** (.20)	.57** (.21)
R-squared	.04	.34	.37

Note: n=18. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=1.81$, $p=0.10$.

Table 1.10: Number of safe choices on Task 4 (small, real)

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.59*** (.07)		.46*** (.08)
Task 2 switchpoint		.50*** (.08)	.28*** (.08)
Dummies for treatment	yes	yes	yes
R-squared	.36	.26	.42

Note: n=130. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=1.7$, $p=0.90$.

Table 1.11: Number of safe choices on Task 4 for 90x only

Variable	Model 1 (Task 1 only)	Model 2 (Task 1 only)	Model 3 (Both tasks)
Task 1 switchpoint	.62*** (.20)		.60*** (.18)
Task 2 switchpoint		.40 (.23)	.35* (.18)
R-squared	.39	.16	.51

Note: n=18. Coefficients are standardized as partial correlations. Standard errors in parentheses.

* denotes $p < .10$, ** denotes $p < .05$, *** denotes $p < .01$.

The one-sided null that $1 > 2$ in Model 3 has $F=0.85$, $p=0.81$.

Bibliography

- [1] Anderson, L., Mellor, J., (2009). “Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure,” *Journal of Risk and Uncertainty* 39, 137-160
- [2] Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.
- [3] Harrison, G. W., & Rutström, E. E. (2008). Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. *Handbook of Experimental Economics Results* (Vol. 1, pp. 752-767). Elsevier.
- [4] Harrison, Glenn W. (2007), “Hypothetical Bias Over Uncertain Outcomes,” in *Using Experimental Methods in Environmental and Resource Economics*, John A. List, ed., Northampton, MA: Elgar.
- [5] Holt, C.A., Laury, S.K., (2002). “Risk aversion and incentive effects,” *The American Economic Review* 92(5), 1644–1655.
- [6] Guiso, L., Paiella, M., (2005). “The Role of Risk Aversion in Predicting Individual Behavior,” Bank of Italy, Working Paper No. 546.

- [7] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291. doi:10.2307/1914185
- [8] Laury, S.K., & Holt, C.A. (2008). "Payoff Scale Effects and Risk Preference Under Real and Hypothetical Conditions," in *Handbook of Experimental Economics Results*, Vol. 1, pp. 1047-1053. Elsevier.

Chapter 2

Personality accounts for stable preferences and expectations across a range of simple games¹

Behaviour in even simple experimental games shows considerable individual differences; but previous attempts to link these preferences to stable personality traits have had mixed results. Here we address three limitations of earlier empirical studies, namely: 1) uncertainties concerning the reliability of preferences over time; 2) use of limited bandwidth personality instruments; and 3) confounds where more than one psychological motive can lead to a particular choice. Sixty-seven participants completed 18 distinct real-money games twice over a two-week interval along with 6 measures concerning their expectations about other players' choices. Personality was measured using the full NEO-PI-R. Choices were highly stable across time ($r = .84$).

¹Note: this chapter expands upon a paper with the same title which was recently published in *Personality and Individual Differences*. The original paper was coauthored with Gary Lewis and Timothy Bates from Edinburgh University's psychology department. See: <http://www.sciencedirect.com/science/article/pii/S0191886911003321>

Moreover, choices on the 12 games and 6 expectations reflected a single underlying dimension of “prosocial orientation”, measuring concern for the payoffs received by other players. Scores on the prosocial orientation dimension were related to personality, with openness, (low) neuroticism, and (low) extraversion retained as significant predictors. Based on these results, we suggest slight modifications to existing theoretical models of social preferences to incorporate individual differences in a manner which is both parsimonious and empirically valid.

2.1 Introduction

When playing a dictator game, some people are more generous than others. Within experimental economics more generally, there is individual variation in behaviour in almost all games, particularly those involving social preferences (Camerer, 2003, is full of examples). Yet there has been comparatively little investigation into the empirical basis for, or the structure of, this individual variation.

Here we will describe an experiment which explores individual variation in social preferences. Sixty-seven subjects played a diverse range of simple games twice over a two-week interval, and we found that the majority of variation in subjects' choices was explained by a stable one-dimensional parameter, and that the social preference parameter was correlated with personality differences as measured by standard psychometric tests (as it should be—see next section).

2.1.1 Personality and Social Preferences

That psychologists' conceptions of personality *should* correlate with economists' conceptions of social preferences is best understood in the context of existing personality research. Personality psychologists have been trying to map personality space using more-or-less modern techniques since at least the 1930s (Barenbaum & Winter, 2008), and the consensus in the field is that the maximum number of reliably measured dimensions is five or six, and that this number does not rise for different measurement techniques or in different societies, cultures or time periods (at least for time periods as far back as the early 20th century; see Deary, 1996). The five- or six-dimensions result places tight restrictions on the sorts of claims about individual differences which economists are free to make. Unless an economist can show that their methods are tapping variation which was previously unmeasured by psycholo-

gists (which is unlikely, given the very large stock of available personality tests and the fact that all of them seem to reduce to the same five or six dimensions), then the variation is going to be correlated with an existing personality factor. Since social preferences are generally meant to tap motivations such as spite, envy or kindness, all of which are well represented in standard personality tests, there are good reasons to think that any stable variation in social preferences will be correlated with personality (we test this assertion, and find it to be supported).

But before addressing the issue of whether or not social preferences correlate with existing measures of personality, we need to settle the question of whether standard social preference games are tapping stable individual differences at all. Consider a simple dictator game in which a subject can either choose a bundle in which both he (the dictator) and another subject (the recipient) each get £6, or he can choose a bundle in which he gets £7, but the recipient only gets £2. Suppose that an experimenter gives this choice to a pool of subjects and finds that 50% of dictators choose the first bundle, and 50% choose the second bundle (this actually an example from Table 2.1). What are we to make of this apparent variation in subjects' social preferences? There are several possible explanations, including: noise (they chose at random), mood (they got out on the wrong side of bed in the morning), and personality (they really *do* have different social preferences). (These possibilities are not totally mutually exclusive, of course.) The *noise* theory is that we ended up with half of the subjects making each choice in our dictator game simply because their choices were random, perhaps because they did not understand or care about the game. If the noise explanation is right, then the experiment has essentially no external validity and we cannot infer much of anything useful from it. But perhaps the noise explanation is wrong and the actual reason our subjects differed was that their *moods* differed on the day that they played the game—subjects

who were in a good mood made the generous choice, and subjects who were in a bad mood made the selfish choice. If this mood explanation of the variation is correct, then the external validity of our experiment is small but nonzero: we can make inferences about how people in certain states of mind will respond to certain stimuli, but we cannot, for example, classify subjects themselves as being selfish or altruistic, since their behaviour is only due to a transient mood. But perhaps the mood explanation is also false; perhaps the variation in choices is due to persistent *personality* differences rather than just moods. Some subjects may be generally more altruistic than others, and the altruistic subjects may be more inclined to choose the bundle where both players get £6. This personality-based explanation is the one that experimentalists are rooting for (if only implicitly), since it is only when choices result from stable character traits that we can make interesting generalisations from the data. This issue matters, because it sets limits to the problems that can be addressed by experimental economics research.

So how can we tell which of the noise, mood, or personality explanations of choice variability is correct? We certainly cannot tell from looking at choices on a single game in isolation, since individual variation would be compatible with any of the three explanations. But if we gave subjects a group of related (but not identical) games, then we could at least decide whether or not the noise explanation was correct. If there is logical consistency in the way that subjects behave across similar games, then their choices are not pure noise. But we cannot distinguish between the mood and the personality explanations unless we take measurements at different points in time. If subjects' play is consistent within sessions and over time, then it is reasonable to conclude that it is explained by a stable personality trait, but if it is only consistent within sessions but not over time, then it is likely due to an unstable characteristic such as mood. Summing up, we can distinguish between the noise,

mood and personality explanations by taking repeated measurements with a closely related family of games.

The procedure described in the preceding paragraph is basic practice within the psychology of individual differences (for an overview, see Part I of Matthews, Deary and Whiteman, 2009), but it is comparatively uncommon in experimental economics. Here we attempt to redress this by examining the structure of individual differences in social preferences within a popular set of simple dictator games developed by Charness and Rabin (2002). Our subjects played twelve of Charness and Rabin's games in each of two sessions spaced two weeks apart, as well as completing a personality test. This allowed us to test whether social preferences exhibit the stability that we would hope for and, if so, whether the social preferences are correlated with other, well established measures of individual differences.

2.1.2 A Brief Review of Related Literature

It is worth noting at this point that even though there have been, to our knowledge, no previous tests of the stability of social preferences over time, there have been numerous papers on the relationship between choices in games and personality as measured by standard tests, and these papers have found a variety of (sometimes contradictory) relationships (discussed below). A reader might object here that a simple check of correlations between choices in games and the results of a personality test would be a simpler way than the procedure outlined above to test the stability of game choices: if choices correlate significantly with personality, then they must be due to personality traits, and if not, not. But there are several problems with this method. Firstly, there is the Type I error problem: checking correlations between a single dependent variable and five personality factors will yield at least one statistically significant relationship roughly 23% of the time even if there is no real

relationship at all ($0.23 = 1 - (.95)^5$; this is assuming a 5% size of the test). This is bad, but the number gets even bigger if we make further adjustments for the fact that the experimenter might be able to choose between several left-hand-side variables ex post, and the additional fact that the publication process is designed to report statistically significant results. Taking these facts together, it is clear that we would expect to find at least some published, statistically significant relationships between personality and choices in games even if there was no true relationship. Furthermore, if the null result were true, then we would expect to find *contradictory* results in the literature (because the coefficients would be random), and this is not totally different from what we do find (see below). So if we want to know whether a behaviour is meaningfully stable, it is not enough just to check for a correlation with the results of a personality test.

So what is the evidence? As noted above, the results are not always the same. Hirsh and Peterson (2009) found that the withdrawal aspect of neuroticism (tapping fear and insecurity) and the enthusiasm aspect of extraversion (tapping positive affect and sociability) from the Big Five aspect scale (DeYoung, Quilty, & Peterson, 2007) independently predicted a greater likelihood of cooperation in a prisoner's dilemma game (correlations of $-.14$, and $-.12$, respectively). The prisoners' dilemma arguably contains strategic as well as social preference aspects (depending on players' preferences, they may regard it as a coordination game), which further complicates interpretation of the results. By contrast, Lönnqvist, Verkasalo, and Walkowitz (2011) found that low neuroticism and high openness predicted more cooperation in a prisoners' dilemma. Using dictator games, Ben-Ner, Kong, and Putterman (2004) reported significant associations between agreeableness and (low) extraversion and the sum offered by the dictator in a dictator game. Finally, Kurzban and Houser (2001) reported non-significant associations between Big Five personality traits and

behaviour in a public goods game. Further studies have examined variation in social preferences using personality frameworks other than the five-factor model. For example, Boone, De Brabander, and van Witteloostuijn (1999) observed that the personality traits locus of control, self-monitoring, and sensation seeking had significant associations ($r = .28 - .44$) with levels of cooperative behaviour in a prisoners' dilemma game. Scheres and Sanfey (2006) observed significant associations between BAS-Drive and BAS-Reward and (low) offers in dictator games. And Swope, Cadigan, Schmitt, and Shupp (2008) reported no significant effects of the Myers-Briggs Type Indicator on social preferences (participants either played prisoners' dilemmas or dictator, ultimatum, or trust games).

These mixed results in studies using the five-factor framework, alongside results from studies using different personality measures, which are not directly comparable, suggest that, while social preferences may be correlated with personality, more research is required. In particular, research addressing limitations of earlier studies will be critical to understanding the role of personality on social preferences.

2.1.3 The Current Study

There are a number of possible explanations for the mixed results described above. Firstly and most straightforwardly, research has seldom addressed the reliability of social preferences. In fact, to our knowledge, no research has addressed stability in social preferences across experimental sessions. If reliability in choice behaviour is low (e.g. because the noise explanation is correct and participants choose randomly), this would explain both the high variability typically seen in games and the inconsistency of measured relationships with stable personality traits in previous research, as noted above.

Secondly, often the personality instruments used in studies associating social

preferences and personality have lacked comprehensive scope. For example, Swope, Cadigan, Schmitt, and Shupp (2008) used the Myers-Briggs Type Indicator, and Boone, De Brabander, and van Witteloostuijn (1999) used an assortment of scales: locus of control, self-monitoring, type-A behaviour, and sensation-seeking. While each of these measures may tap specific traits, the core five-factor model has demonstrated broader coverage of stable human behaviour than any other measurement instrument (e.g. Costa & McCrae, 1992; Goldberg, 1993), and so provides a more comprehensive tool by which to understand putative trait influences on social preferences. To address these limitations, in the present study we measured social preference twice over a two-week interval, and utilised the full-spectrum 240 question NEO-PI-R (Costa & McCrae, 1992) in order to gain a comprehensive assessment of personality.

Thirdly, at a more basic level, much research has focused on just one or two experimental games, such as the dictator and ultimatum games; however, important confounds have been identified in these games which render choices ambiguous as to underlying motivations or preferences (Charness & Rabin, 2002). For example, rejection of a low offer in the ultimatum game can reflect difference aversion or retaliation. These distinct motives, in turn, confound potential underlying personality traits, such as neuroticism and agreeableness, respectively. Likewise, in the prisoners' dilemma, a choice to defect can reflect aversion to differential outcomes, aversion to risk, or a self-regarding preference. These confounds can be mitigated by exploring a range of payoff pairings, varying absolute and relative payoff differences, as well as allowing multi-stage games (Charness & Rabin, 2002). Finally, and importantly, choices reflect expectations about the other player in addition to personal preferences. An example would be the expectation (or fear) that the other player will defect. Because of these confounds in single games, personality is likely to have

apparently divergent or null associations to preferences on different games because of the distinct ways in which each game might trigger personality-related preferences.

Here we have used a set of games which are well-established in the experimental economics literature (Charness & Rabin, 2002). By using a mixture of games, we can eliminate common confounds between Pareto-damaging behaviour, retaliation, and inequality reduction. These games also tap into the two primary factors which economic theorists have identified as critical for explaining social preferences: How much the other participants receive (comparison-based preferences; people will be less kind towards those who have more than themselves), and the perceived intentions of the other participants (intention-based preferences; people will be less kind towards those who have shown bad intentions). These factors have been separately identified by Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002; see also Daruvala, 2010, for a review), but have so far only been discussed in terms of their influence on average behaviour; the factor structure of these games has not yet (to our knowledge) been examined.

2.1.4 The Prosocial Orientation Scale

We find that choices across the full range of games are positively correlated—subjects who made the social surplus maximising choice in any one game were more likely to make the social surplus maximising choice in another game. We use this result to propose a new measure of social preferences which we deem the Prosocial Orientation Scale. An individual’s POS score is calculated by simply aggregating their choices and expectations from multiple games into a single scale, which we found to be significantly correlated with personality as measured on the NEO personality test. The measure is a simple tool which can be adapted by future researchers to control for social preferences in laboratory settings.

2.2 Method

2.2.1 Participants

Seventy-five participants were recruited from an undergraduate participation pool: Participants received partial course credit for attending as well as a financial remuneration based on choices made in the experimental tasks. Of the initial 75 participants, 71 returned for the second session. A further four participants' data were lost due to a data storage failure. Of the 67 remaining participants, 54 were female (mean age = 19; SD = 3.9 years).

2.2.2 Measures

2.2.2.1 Personality

Five-factor model (FFM) personality traits were measured using the 240-item NEO-PI-R (Costa & McCrae, 1992). Participants completed the inventory at individual computer terminals, selecting, for each statement contained in the test, one of the response options: “strongly disagree”, “disagree”, “neutral”, “agree” or “strongly agree”. These individual responses were coded as 1, 2, 3, 4 or 5 and aggregated into scores on the personality domains by summing up positively keyed items and subtracting negatively keyed ones (e.g. for Extraversion, responses to questions like “I laugh easily” would be added to the Extraversion total score, while responses to questions like “I usually prefer to do things alone” would be subtracted).

2.2.2.2 Games

A set of six dictator and six response games (described below) were taken from Charness and Rabin (2002) and are named according to their convention (with the

exception of Ed 128, which is derived from Berk 28 but was not in the original set of games). As an example of comparison-based preferences, in the game known as Berk 23 (see Figure 2.1), Player B chooses between an outcome in which Player A gets £8 and Player B gets £2, versus an outcome in which each receives £0. As an example of an intention-based preference, in the game known as Berk 22 (see Figure 2.2), Player A can choose £3.75 for themselves and £10 for Player B, or let Player B choose between £4 for each player or £2.50 for Player A and £3.50 for themselves. Here, if Player A ‘enters’ the game and allows Player B to make the choice, Player A deprives Player B of a guaranteed £10. Of course, Player B may now choose the lower payoff for themselves (£3.50 rather than £4) in order to punish Player A (Player A would then receive £2.50 rather than £4). Participants played all response games both as Player B and as Player A.

Games are listed in Table 2.1 corresponding to the (fixed) order that they were played by the participants. Games are presented so that the prosocial choice for Player B is always on the left (although the games were counter-balanced when presented to the participants), with the exception of Berk 26, in which the total payoff is identical for both choices available to Player B.

In order to explore the role of players’ expectations about the behaviour of others on their own choices, participants were asked to estimate the percentage of all other participants who would make the prosocial choice when acting as Player B in the relevant games. This was taken after they had made their choice in the role of Player A in the response games. Participants were informed that there would be a £10 prize for the participant with the most accurate estimates of other players’ behaviour.

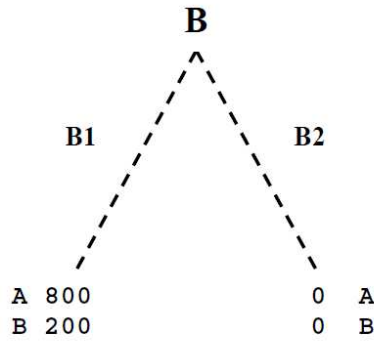
Participants were tested individually in separate experimental cubicles. Participants were informed both when they signed up for the experiment and again at the beginning of their first session that they would be required to return for the second

Figure 2.1: Screenshot of Berk 23

GAME 23

In this period, you are person B.

You may choose B1 or B2. Player A has no choice in this game. If you choose B1, you would receive 200 and player A would receive 800. If you choose B2, you would each receive 0.



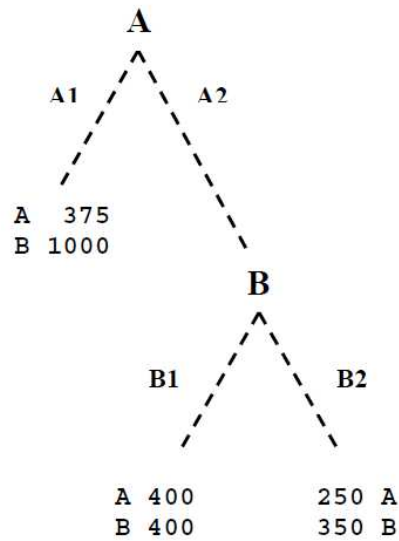
Note: This shows a screenshot of one of the dictator games as presented to subjects. Subjects had previously been told that 200 and 800 translate to £2 and £8, respectively, and that their responses would be treated anonymously.

Figure 2.2: Screenshot of Berk 22A

GAME 22

In this period, you are person A.

You may choose A1 or A2. If you choose A1, you would receive 375, and player B would receive 1000. If you choose A2, then player B's choice of B1 or B2 would determine the outcome. If you choose A2 and player B chooses B1, you would each receive 400. If you choose A2 and player B chooses B2, you would receive 250, and he or she would receive 350. Player B will make a choice without being informed of your decision. Player B knows that his or her choice only affects the outcome if you choose A2, so that he or she will choose B1 or B2 on the assumption that you have chosen A2 over A1.



Note: This shows a screenshot of one of the dictator games as presented to subjects. Subjects had previously been told that 375=£3.75, 1000=£10.00, etc., and that their responses would be treated anonymously.

Table 2.1: Game payoffs and the proportions of participants making each choice

		<i>Out</i>	<i>Enter</i>	<i>Left</i>	<i>Right</i>
<i>Dictator Games</i>					
Berk 17	B chooses (750,375) vs. (400,400)			.16	.84
Berk 23	B chooses (800,200) vs. (0,0)			.67	.33
Berk 29	B chooses (750,400) vs. (400,400)			.38	.62
Berk 15	B chooses (600,600) vs. (200,700)			.50	.50
Berk 26	B chooses (400,400) vs. (0,800)			.35	.65
Ed 128	B chooses (125,125) vs. (75,125)			.72	.28
<i>Response Games</i>					
Berk 13	A chooses (550,550) or lets	.86	.14		
	B choose (750,375) vs. (400,400)			.16	.84
	Expectations of B			.19	.81
Berk 30	A chooses (400,1200) or lets	.49	.51		
	B choose (400,200) vs. (0,0)			.70	.30
	Expectations of B			.65	.35
Berk 31	A chooses (750,750) or lets	.78	.22		
	B choose (800,200) vs. (0,0)			.57	.43
	Expectations of B			.57	.43
Berk 19	A chooses (700,200) or lets	.75	.25		
	B choose (600,600) vs. (200,700)			.46	.54
	Expectations of B			.29	.71
Berk 22	A chooses (375,1000) or lets	.31	.69		
	B choose (400,400) vs. (250,350)			.80	.20
	Expectations of B			.80	.20
Berk 28	A chooses (100,1000) or lets	.31	.69		
	B choose (125,125) vs. (75,125)			.65	.35
	Expectations of B			.41	.59

Note: Numbers in parentheses show (Player A, Player B) payoffs in British pence; Out = The proportion of Player Bs who opted to stay ‘out’ of the game by choosing the available payoffs and thus depriving Player A of making a choice; Enter = The proportion of Player Bs who ‘entered’ the game, thus allowing Player A to make a choice between the available payoffs; Left = The proportion of participants choosing the payoffs to the left; Right = The proportion of participants choosing the payoffs to the right; Expectations of B = Player A’s expectations of the average B choice.

half of the experiment in two weeks in order to obtain both course credit and monetary payment. Participants were paid at the end of the second session based on their rewards in one task from each of the two sessions chosen at random and this was common knowledge.

Participants first played all six dictator games before playing all six response games as the second player and finally playing the same six response games as the first player (see Table 2.1). Participants were told that payoffs would be based on converting winnings from one game at random in each session into British pence (i.e. $750 = \text{£}7.50$). Participants were not told of their partners' choices until the end of the second session, when they were paid. The NEO-PI-R was administered in two blocks: one at the end of the first session and one at the end of the second session.

2.3 Results

The proportions of choices made by participants for each game and the expectations of other players' behaviour are summarised in Table 2.1.

2.3.1 Factor structure of social preferences

It is unwise to attempt to correlate personality with decisions in individual games: single decisions contain too much noise. Instead we follow the logic of personality test construction: noise reduction through aggregation. Thus we initially examined our data to see if there was a case for summing behaviour into scales before assessing behavioural stability across sessions and the relationship with personality. To preview the conclusion: we found that there was a strong case for aggregation, and so we used our data to construct a single prosocial orientation scale (POS).

Within the dictator games, there are theoretical reasons for splitting the games

into categories, as described above. If participants have comparison-based preferences (i.e. if they care whether the other participants have more or less than them), then Ed128B, Berk15B, and Berk26B might elicit different behaviours than Berk29B, Berk17B and Berk23B, since the first three games represent situations where the deciding participant has either the same as or less than the recipient, whereas the final three games represent situations where the deciding participant has either the same as or more than the recipient. Similarly, if the participants have intention-based preferences, then we may observe yet different patterns of behaviour in games such as Berk30B or Berk31B, since the participant in role B may feel wronged by the participant in role A. Finally, we should allow for the possibility that participants' expectations about each other's choices will have a relationship with personality which is independent of the relationship with their own choices.

The surprising result is that choices all of these items turn out to be strongly intercorrelated. A correlation matrix constructed from all of the dictator games as well as expectations and responses for both A- and B-players in the response games shows that all items are positively intercorrelated. Indeed, a factor analysis of the data reveals that the first factor accounts for 58% of the variance, and a scree plot shows that the eigenvalues (10.9, 3.0, 2.2, 0.8, 0.7, ...) drop off very rapidly after the initial value. That first factor has positive loadings on all 24 variables. Thus we were persuaded to construct a single scale by averaging across all 18 roles in the 12 games as well as the 6 sets of expectations for use in subsequent analyses. To construct the scale, we coded all choices such that larger numbers correspond to greater total payoff (i.e. greater social surplus). For decision tasks, the social surplus maximising choice (e.g. choosing left in any of the response games) would count as 1, whereas the socially inefficient choice would count as 0. For expectations, the expected proportion of other subjects who make the surplus maximising choice

represents the score (e.g. expecting 78% of responders to go left in a response game corresponds to a value of 0.78). A subject’s score on the overall scale was a simple average of the 24 scores on the individual items. Thus the prosocial orientation scale (POS) represents the average proportion of social surplus maximising decisions that were either made or expected. Mathematically, the POS score for individual i is given by:

$$POS_i = \frac{1}{k} \sum_{m=1}^k x_m$$

Here, i denotes an individual and k is the number of games played plus the number of expectations elicited. If the m -th item was a game, then $x_m = 1$ if individual i made the prosocial choice on that game and $x_m = 0$ if the individual failed to make the prosocial choice. If x_m was an expectation, then x_m takes on a value between zero and one, where the value corresponds to the proportion of the population which player i expected to make the prosocial choice on the relevant game.

2.3.2 The PO Scale: reliability

The PO Scale demonstrates good reliability (Cronbach’s $\alpha = .94$), as do subscales composed of choices in the six dictator games (Cronbach’s $\alpha = .78$), the six “A” response games (Cronbach’s $\alpha = .74$), the six “B” response games (Cronbach’s $\alpha = .86$), and the six sets of expectations (Cronbach’s $\alpha = .94$). Furthermore, since we have data spanning a two week period, we can also test the session-to-session stability of the prosocial orientation scale by calculating each subject’s POS score in each of the weeks and checking the session-to-session correlation. When we do this, we obtain a correlation of 0.84 ($t = 12.6, p < .0001$), indicating good stability.

2.3.3 Personality as a Predictor of Prosocial Orientation

Having established that social preferences demonstrated high reliability and stability, we next examined the relationship of FFM traits to social preferences using linear modelling (multiple regression) with POS scores as the dependent variable and entering each of the FFM domains (as well as age and sex) as independent variables. This model accounted for 25.5% of variance in POS scores, with neuroticism ($\beta = -.33$, $p = .02$), extraversion ($\beta = -.32$, $p = .02$), and openness to experience ($\beta = .41$, $p = .003$) being significant predictors (see Table 2.2). Agreeableness, conscientiousness, sex, and age were not significant predictors and removing these variables did not significantly alter model fit ($F = 1.46$, $p = .23$) or the parameter estimates of the significant predictors. Pairwise interaction terms for all personality factors were non-significant.

As a check on the robustness of our results, we re-calculated the PO scale as the first (unrotated) principal component to emerge from a principal components analysis of the choice and expectation data. Doing so yielded results which were almost indistinguishable, both qualitatively and quantitatively, from the results presented above. This is encouraging, as it suggests that our results do not depend very strongly on the functional form we have chosen.

2.4 Implications for modelling: traits instead of types

The finding that a common factor underpinned choice behaviour across our games is of considerable theoretical interest because it suggests that variation in social preferences can be described by a single continuous parameter, even if the underlying model of social preferences has multiple parameters. This matters, because it suggests that social preferences can be reconciled with two of the central findings of personality

Table 2.2: Personality predictors of the Prosocial Orientation Scale (standardised coefficients are presented) for Model 1 and Model 2

	$\beta_{Model\ 1}$	$\beta_{Model\ 2}$
Neuroticism	-.33*	-.38**
Extraversion	-.32*	-.35*
Openness	.41**	.36**
Agreeableness	.17	
Conscientiousness	-.16	
Age	-.15	
Sex	-.17	
Multiple R^2	.26	.18

Note. $N = 67$; * = $p < .05$; ** = $p < .01$

psychology—that personality is continuous, and that it is five- or six-dimensional. The fact that personality is continuous implies that individual differences should be modeled with trait-based theories rather than type-based theories, and the fact that personality consists of only about five dimensions in total suggests the need for parsimony: theorists cannot use up too many dimensions modelling social preferences, or they won't have any dimensions left over to model risk attitudes, time preferences, loss aversion, work ethic, aggression, and so on. Below, I will show how these constraints can be jointly satisfied and I will relate the discussion to two popular models of social preference (Charness and Rabin 2002 and Fehr and Schmidt 1999), but first I want to briefly expand on the reasoning behind those models.

Economists are often in the habit of classifying subjects as different types according to the way that they act in laboratory games. Fehr and Schmidt (1999) classify subjects into fair types and selfish types based on their choices in a series of games, and Charness and Rabin (2002) classify subjects based on similar criteria as having one of four types of preferences: selfish preferences, competitive preferences, difference-averse preferences, or social welfare preferences (these terms will be defined in more detail in section 2.4.1 below). It is not hard to find other examples of

this sort of classification in experimental economics papers. But the more valid (and arguably more intuitive) way to think about individual differences is that each one is a trait measured along a continuum. Deary (2009) provides a brief history and exposition of the trait approach within personality psychology. In our case, the way to think about the issue is that subjects are not discretely selfish or fair-minded, but somewhere along a spectrum in between the two.

Economists are also, at least sometimes, in the habit of positing independent parameters along which individuals may vary without either testing to see if individuals actually do vary independently along those dimensions, or making reference to the total dimensionality of the parameter space in which they are theorising. Fehr and Schmidt (1999) and Charness and Rabin (2002) are examples of this, but again it is not difficult to find others. This approach is not a problem per se, as theorising generally precedes testing, but the theories must be tested eventually. Particularly, theory must face up to the "budget constraint" that there cannot be more than five or six major parameters describing individual differences. In this case, we have found evidence that social preferences in different situations co-vary strongly enough that they can be described as effectively one-dimensional. We denote that dimension by ψ (henceforth, the reader should think of ψ as a strictly increasing function of POS, but with the exact nature of the function dependent on the measures used to elicit POS). In the next section, we will briefly show how ψ can be parametrised in various other-regarding utility functions.

2.4.1 Charness and Rabin (2002)

Consider the linear model of social preferences used by Charness and Rabin (2002). In that model, players make decisions in order to maximize a utility function containing a weighted average of their own payoff (π_i) and the payoff of the other player

(π_j) . In addition to the payoff variables, the model includes three parameters: the weight that player i assigns to player j 's payoff when i has a relatively higher payoff (ρ), the weight that player i assigns to player j 's payoff when i has a relatively lower payoff (σ), and finally a parameter which allows for a change in i 's weights when j has misbehaved somehow (θ).

$$U_i(\pi_i, \pi_j) = (\rho r + \sigma s + \theta q)\pi_j + (1 - \rho r - \sigma s - \theta q)\pi_i$$

where

$$r = 1 \text{ if } \pi_i > \pi_j \text{ and } r = 0 \text{ otherwise}$$

$$s = 1 \text{ if } \pi_i < \pi_j \text{ and } s = 0 \text{ otherwise}$$

$$q = 1 \text{ if A has misbehaved, and } q = 0 \text{ otherwise}$$

Thus for a dictator game like Berk 15 (where $\pi_i \geq \pi_j$), this would simplify to

$$U_i(\pi_i, \pi_j) = \rho\pi_j + (1 - \rho)\pi_i$$

And for a dictator game like Berk 17 (where $\pi_i \leq \pi_j$), it would simplify to

$$U_i(\pi_i, \pi_j) = \sigma\pi_j + (1 - \sigma)\pi_i$$

If you wanted to model individual differences using this model, you might have thought that you would have to index all of the parameters by individual i

$$U_i(\pi_i, \pi_j) = (\rho_i r + \sigma_i s + \theta_i q)\pi_j + (1 - \rho_i r - \sigma_i s - \theta_i q)\pi_i$$

But, as noted above, the results of the present study suggest that this would be redundant—a version of this model which wished to account for individual differences would not need to allow for all parameters to vary across individuals; we would

instead only have to incorporate one additional parameter (ψ_i) representing POS which allowed the other weights to be scaled up or down (large values of ψ_i correspond to high scores on the PO scale)

$$U_i(\pi_i, \pi_j) = (\rho(\psi_i)r + \sigma(\psi_i)s + \theta(\psi_i)q) \pi_j + (1 - \rho(\psi_i)r - \sigma(\psi_i)s - \theta(\psi_i)q) \pi_i$$

Where the functions $\rho(\cdot)$, $\sigma(\cdot)$, and $\theta(\cdot)$ are the same for everyone, and, in addition to being weakly increasing, satisfy the following two conditions: (1) $\theta(\psi) \geq 0 \quad \forall \psi \in R$ (i.e. in response to an opponent's misbehaviour, an agent will not increase the weight placed on the opponent's payoff) and (2) $\rho(\psi) \geq \sigma(\psi) \quad \forall \psi \in R$ (i.e. agents give at least as much weight to the payoffs of relatively poor opponents as to relatively rich opponents). So the equation for Berk 15 now becomes::

$$U_i(\pi_i, \pi_j) = \rho(\psi_i)\pi_j + (1 - \rho(\psi_i)) \pi_i$$

And for Berk 17

$$U_i(\pi_i, \pi_j) = \sigma(\psi_i)\pi_j + (1 - \sigma(\psi_i)) \pi_i$$

This is a much better case from the perspective of modelling than the possible alternative case in which all weights varied independently. Because it is possible to model individual differences with a single parameter ψ , and because the use of this single parameter can account for as much as 58% of the variation in choice behavior (as suggested by our factor analysis above), experimentalists in the future will have much less of an excuse to ignore individual differences.

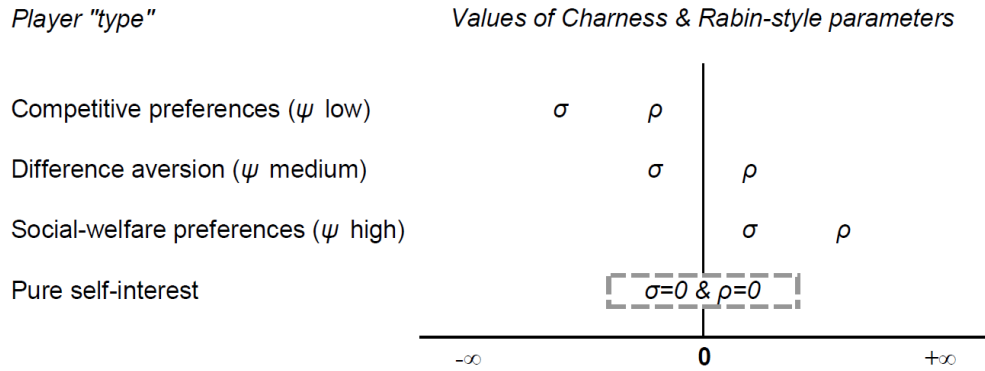
It is worth considering what this continuum-oriented perspective means for the

type-based taxonomy which Charness and Rabin outlined in their paper. Charness and Rabin posited four types of preferences:

- Competitive preferences ($\sigma < \rho < 0$). Players of this type are concerned about improving their relative position vis-a-vis everyone, and they place a negative weight on the payoffs of players who have either more or less than them (i.e. they are willing to make sacrifices to hurt both rich and poor alike, but they'll make bigger sacrifices in order to hurt the rich).
- Difference averse preferences ($\sigma < 0 < \rho$). These types place a positive weight on the payoff of players who have less than them, but a negative weight on the payoff of players who have more than them. The combination of these weights amounts to difference aversion of the classic Fehr and Schmidt (1999) type (i.e. they are willing to make sacrifices either to help the poor or to hurt the rich).
- Social-welfare preferences ($0 < \sigma < \rho$). These types place a positive weight on the payoff of the other player whether he has a higher or lower payoff than the agent (i.e. they are willing to make sacrifices to help both rich and poor, but they'll make bigger sacrifices to help the poor).
- Pure self interest ($\sigma = 0 = \rho$). These types place no weight at all on other players' payoffs when they make their decisions, acting purely to maximise their own payoff (i.e. they will make sacrifices to help no one).

But these types can (almost) be compressed to line up on a continuum. Recall that $\rho'(\psi) \geq 0$, $\sigma'(\psi) \geq 0$ and $\rho(\psi) - \sigma(\psi) \geq 0$ (from condition 2, above). In words, this just means that a person's generosity-to-the-poor parameter ρ_i is (weakly) greater than their generosity-to-the-rich parameter σ_i and that both parameters are increasing in ψ Figure 2.3 shows what this might look like. Individuals who are

Figure 2.3: Types and Continua



Note: The figure compares the four types identified by Charness and Rabin (2002) to the continuum-based scheme in this paper. The first three types can easily be fit into a continuum scheme, but the pure self-interest types do not fit very well.

characterised by small values of ψ_i have correspondingly small values of σ_i and ρ_i and thus competitive preferences; this is shown on the top line of Figure 2.3. Individuals whose values of ψ_i are somewhat higher end up with difference averse preferences, and those who are high enough on the ψ spectrum have positive values of ρ_i and σ_i and thus exhibit social welfare preferences (shown on the third line of Figure 2.3).

The problem with this schema is that it cannot accommodate pure self-interest (depicted on the fourth line of Figure 2.3). Self interested individuals, unlike typical members of the population, have no space between the (zero) weights they place on the payoffs of rich and poor: $\rho_i - \sigma_i = 0 - 0 = 0$. By constraining our model of social preferences as we did above, we remove the possibility of such preferences. This highlights the tradeoff of few parameters and fewer possible types versus more of both. Since theories of individual differences must obey the “budget constraint” of having only five or six separate dimensions, and since Charness and Rabin (2002) find that pure self-interest is relatively uncommon in their data anyhow, we think it is best to opt for the more parsimonious choice.

2.4.2 Fehr and Schmidt (1999)

We can also examine the implications of this result for other models of social preferences. Consider the model of Fehr and Schmidt (1999). They posit a model of social preferences where people gain utility from having their payoff increased, but lose utility from having large payoff gaps vis-a-vis the other players. Gaps are treated asymmetrically, however—it is worse to be behind than to be ahead. In the two-player case, their model has the following form

$$U_i(\pi_i, \pi_j) = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\}$$

Where π_i and π_j are defined as above and α_i represents the disutility which person i incurs for each unit by which player j 's payoff exceeds his own, and β_i , represents the disutility which person i incurs for each unit by which his own payoff exceeds player j 's. By assumption, $\beta_i \leq \alpha_i$ and $0 \leq \beta_i \leq 1$. Before considering the implications of the current result for the Fehr and Schmidt model, we can re-write it slightly to bring it in line with the Charness and Rabin's dummy variable notation:

$$U_i(\pi_i, \pi_j) = \pi_i - s \cdot \alpha_i(\pi_j - \pi_i) - r \cdot \beta_i(\pi_i - \pi_j)$$

where r and s are defined as above. In cases where $\pi_i \geq \pi_j$ (as in Berk 15, noted above), this simplifies to

$$U_i(\pi_i, \pi_j) = \pi_i - \beta_i(\pi_i - \pi_j)$$

$$U_i(\pi_i, \pi_j) = \beta_i\pi_j + (1 - \beta_i)\pi_i$$

And when $\pi_i \leq \pi_j$ (as in Berk 17, noted above)

$$U_i(\pi_i, \pi_j) = \pi_i - \alpha_i(\pi_j - \pi_i)$$

$$U_i(\pi_i, \pi_j) = -\alpha_i\pi_j + (1 + \alpha_i)\pi_i$$

Now we can see that if we set $\beta_i = \rho_i$ and $\alpha_i = -\sigma_i$, then the Fehr and Schmidt model is exactly equivalent to the Charness and Rabin model for the two-player case (with the exceptions that the Fehr and Schmidt model places tighter assumptions on the signs of α and β than Charness and Rabin, and Fehr and Schmidt implicitly constrain θ to equal zero). Since the Fehr and Schmidt model is just a constrained version of the Charness and Rabin model, we should be just as happy to reduce the scope of individual variation to a single dimension as we were above.

2.5 Discussion

The current study set out to determine whether, firstly, social preferences are stable over time, secondly, whether personality exerts an influence on social preferences in a series of simple games and, finally, how to model individual differences in social preferences in a way that is empirically valid.

On the first point, behavioural stability in choices across sessions was high, suggesting that both choice behaviour and expectations of choice behaviour in our menu of games is underpinned by a stable trait disposition.

Our data also provide strong evidence that personality traits exert significant influences upon social preferences. Moreover, these personality effects were seen to influence a single dimension of behaviour – termed here the Prosocial Orientation Scale (POS). Openness to experience was a positive predictor of the POS, such that

higher levels of openness predicted more benevolent behaviour, and expectations of more benevolent behaviour. Extraversion and neuroticism were negative predictors of prosocial preferences, such that more introverted and more emotionally stable individuals were more likely to make benevolent choices, and to expect others to do the same. These results confirm and extend the findings of Lönnqvist, Verkasalo and Walkowitz (2011), who found that high openness and low neuroticism predicted cooperation in a prisoners' dilemma game. These findings are in tension with some previous research reporting that high neuroticism relates to more benevolent social preferences (Hirsh & Peterson, 2009). However, Lönnqvist, Verkasalo and Walkowitz (2011) note that Hirsh and Peterson (2009) used hypothetical stakes, whereas their own work showed that the relationship between social preferences and personality breaks down in the absence of monetary stakes.

The present results run contrary to one of the core hypotheses of the study, namely that agreeableness would be a positive predictor of prosocial preferences. This null-result is striking in light of agreeableness being characterised as a trait indexing empathy and concern for social welfare (Jensen-Campbell & Graziano, 2001) and so having intuitive links to benevolent social preferences. While this result is puzzling, recent work has suggested that (at least) two distinct mechanisms motivate prosocial behaviours: a fairness-based system and a compassion-based system (Singer & Steinbeis, 2009). As such, it is plausible that the laboratory environment (participants completed the experiment alone in lab rooms) did not suitably invoke empathic concern for other participants thus not 'activating' the compassion-based system; a system which is likely to be analogous to agreeableness. Openness, however, with robust links to liberal political values (McCrae, 1996), may be more closely aligned to the fairness-based motivational system (Lewis and Bates, 2011), perhaps explaining the observed association between openness and benevolent social

preferences.

The finding that a common factor underpinned choice behaviour across our games is of considerable theoretical interest because it suggests that variation in social preferences can be described by a single parameter. It was noted earlier that social preference theories generally posit that the weight which agents place on each other's payoffs depends on whether those others have a higher or lower payoff than the agent, and also on whether the recipient seems to have good intentions. The model of behaviour posited in Charness & Rabin (2002), for example, has three distinct parameters to reflect these contingencies. Our analysis suggests that these parameters can be collapsed into a single dimension, because the participants who were most likely to sacrifice for those who had a lower payoff than themselves were also the most likely to sacrifice for those who had a higher payoff, and for those who had (or hadn't) shown bad intentions.

Specific limitations require mention. Firstly, females were overrepresented in our sample, as were students. While no sex or age effects were evident, it would be useful to extend these results to a sample with greater power to detect such effects. Secondly, our sample size would have had to be at least ten times larger to perform a full factor analysis at the item level and determine the factor structure with a greater degree of confidence.

In conclusion, the results suggest that differences in behaviour on simple games are stable, that they reflect a general preference for prosocial outcomes, that they have a significant link to personality traits of extraversion, neuroticism, and openness, and that they can be straightforwardly incorporated into several existing models of social preference. Future work seeking to identify trait associations with social preferences is recommended to place less emphasis on confounded games, such as the prisoners' dilemma and the ultimatum game, and instead make greater use of

games that avoid such drawbacks. Extensions to our preliminary investigation of the psychometric structure of social preferences will also be valuable.

Bibliography

- [1] Barenbaum, Nicole B. and Winter, David G. (2008). "History of Modern Personality Theory and Research." in Handbook of Personality: Theory and Research. O. P. John, R. W. Robins and L. A. Pervin, eds. New York, NY, Guilford Press: 3-26.
- [2] Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, 25, 581-589.
- [3] Bolton, G.E. & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity and competition. *American Economic Review*, 901, 166-193.
- [4] Boone, C., De Brabander, B., & van Witteloostuijn, A. (1999). The impact of personality on behavior in five Prisoner's Dilemma games. *Journal of Economic Psychology*, 20, 343-377.
- [5] Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*: Princeton University Press.
- [6] Charness, G. & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117, 817-869.

- [7] Costa, P. T. & McCrae, R. R. (1992). NEO-PI-R professional manual. Odessa, FL: Psychological Assessment Resources, Inc.
- [8] Daruvala, D. (2010). Would the right social preference model please stand up! *Journal of Economic Behavior & Organization*, 73, 199-208. doi: 10.1016/j.jebo.2009.10.003.
- [9] Deary, I.J. (1996). A (latent) big five personality model in 1915?: a re-analysis of Webb's data. *Journal of Personality and Social Psychology*, 71, 992-1005.
- [10] Deary, I. (2009). The Trait Approach to Personality. In P.J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 89-109). New York, NY: CUP.
- [11] DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: Ten aspects of the big five. *Journal of Personality and Social Psychology*, 93, 880-896.
- [12] Fehr, E. & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114, 817-868.
- [13] Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.
- [14] Hirsh, J. B. & Peterson, J. B. (2009). Extraversion, neuroticism, and the prisoner's dilemma. *Personality and Individual Differences*, 46, 254-256.
- [15] Jensen-Campbell, L. A. & Graziano, W. G. (2001). Agreeableness as a Moderator of Interpersonal Conflict. *Journal of Personality*, 69, 323-362.

- [16] Kurzban, R. & Houser, D. (2001). Individual differences in cooperation in a circular public goods game. *European Journal of Personality*, 15(S1), S37-S52.
- [17] Lewis, G. J. & Bates, T. C. (2011). From Left to Right: How the personality system allows basic traits to influence politics via characteristic moral adaptations. *British Journal of Psychology*, 102: no. doi: 10.1111/j.2044-8295.2011.02016.x.
- [18] Lönnqvist, J.-E., Verkasalo, M., & Walkowitz, G. (2011). It pays to pay - Big Five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Personality and Individual Differences*, 50, 300-304.
- [19] Maltby, J., Wood, A. M., Day, L., Kon, T. W. H., Colley, A., & Linley, P. A. (2008). Personality predictors of levels of forgiveness two and a half years after the transgression. *Journal of Research in Personality*, 42, 1088-1094.
- [20] Matthews, G, Deary, I. J. & Whiteman, M. C. (2009). *Personality Traits*. New York, NY: CUP.
- [21] McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, 120, 323-337.
- [22] Nettle, D. (2005). An evolutionary perspective on the extraversion continuum. *Evolution and Human Behavior*, 26, 363-373.
- [23] Pruitt, D. G. & Kimmel, M. J. (1977). Twenty Years of Experimental Gaming: Critique, Synthesis, and Suggestions for the Future. *Annual Review of Psychology*, 28, 363-392.

- [24] Scheres, A. & Sanfey, A. (2006) Individual differences in decision making: Drive and reward responsiveness affect strategic bargaining in economic games. *Behavior and Brain Functions*, 2, 35.
- [25] Singer, T. & Steinbeis, N. (2009). Differential Roles of Fairness- and Compassion-Based Motivations for Cooperation, Defection, and Punishment. *Annals of the New York Academy of Sciences*, 1167(Values, Empathy, and Fairness across Social Barriers), 41-50.
- [26] Swope, K., Cadigan, J., Schmitt, P., & Schupp, R. (2008) Personality preferences in laboratory economics experiments. *Journal of Socio-Economics*, 37, 998-1009.

Chapter 3

Contrasting the FFM and HEXACO models of the structure of personality: A spectral clustering approach¹

While dimensional models of personality are widespread, there is still not universal agreement on a structural framework. Much of the debate hinges on results from factor analysis. Here we use the methodology of spectral clustering to test the structure of personality in a large dataset ($n = 20,993$), with a broad bandwidth measure (300-item version of the IPIP NEO personality questionnaire), and compare our results to those obtained from factor analytic solutions. Support was found for five- and six-cluster solutions, depending on the weight given to weaker relation-

¹Note: this chapter expands upon a paper with the same title which was recently submitted to the *Journal of Personality*. The original paper was coauthored with Scott Pauls and Dan Rockmore (both from the mathematics department at Dartmouth College) and Timothy Bates (from Edinburgh University's psychology department).

ships between items. Both the five-cluster and five-factor solutions mapped onto the conventional Five Factor Model (FFM) dimensions. A six-factor solution yielded a sixth factor that was small and hard to interpret. However, the six-cluster analysis rendered a solution highly similar to the HEXACO model. We suggest that spectral clustering provides a useful alternative view of personality data, shedding light on the role of strong and weak item relationships for different theoretical models. For economists interested in modelling individual differences, the present results imply that the “budget constraint” on the number of dimensions that can be modelled has increased from five to six, and that models of preferences need not be linear.

3.1 Introduction

Taxonomy is basic to any science – often being viewed as the "facts" of a field, for which theories then compete to account (McCrae & John, 1992). In personality psychology, the end of the 20th century saw the emergence of a broad consensus as to the structure of personality: five orthogonal, broad bandwidth domains collectively known as the Five Factor Model (FFM) (McCrae & Costa, 1997; 2003). The five factors (and their meanings) are: Neuroticism (tendency to experience negative emotions like fear, anxiety, depression, etc.), Extraversion (tendency to be outgoing and experience positive emotions), Openness (preference for art and intellectual endeavours, tendency to be left-leaning politically), Agreeableness (tendency to empathise and be kind to others) and Conscientiousness (task orientation, ability to carry out plans). Variants such as the Big Five (emphasizing Intellect over Openness; Goldberg, 1990) have often been seen as opportunities to refine and redefine domains within this space, rather than as fundamental challenges. An actively researched alternative – the HEXACO model – appears to question the consensus more robustly, suggesting that personality consists of, not five, but six factors (Ashton & Lee, 2007). Roughly speaking, the HEXACO model splits FFM Agreeableness into two factors: Honesty-Humility and (residual) Agreeableness. – Whether HEXACO or the FFM provides a better fit is important for determining the dimensionality of personality, which is important to anyone interested in creating a good model of individual differences.

Factor analysis and more recently confirmatory factor analysis (Jöreskog, 1969) remain the primary techniques for exploring structure in questionnaire data and are the foundation for the FFM. Confirmatory factor analytic analyses of the FFM have met with mixed success (Marsh et al, 2010; Gignac, Bates & Jang, 2007), and attempts to use alternatives not based on factor analysis have only recently begun (Tiliopoulos, Pallier and Coxon, 2010). Here we present the first application (to our

knowledge) of spectral clustering (Von Luxburg, 2007) to personality data, analysing responses to a battery of 300 items selected to represent the dimensions of the Five Factor Model (Johnson, 2005). Results are computed for both factor analytic and spectral clustering solutions, and parameters within the spectral clustering optimisation are varied in order to contrast different theoretical models of personality structure: HEXACO and the FFM. We first briefly outline spectral clustering, then introduce its application to personality.

Spectral clustering is one of a family of clustering techniques that shares with factor analysis the basic objective of creating a low-dimensional representation of the data. The techniques differ in their specific optimisation targets: while factor analysis maximises the amount of variance accounted for in a given low-dimensional projection of the correlation matrix (Spearman 1927, Cattell 1978), spectral clustering, by contrast, minimises the amount of cutting necessary to divide a geometric representation of the data into separate clusters, where each cluster can effectively be viewed as a summary dimension of the underlying data (see e.g., Leibon, Pauls, Rockmore & Savell, 2008).

An advantage of the spectral clustering approach used here is the inclusion of a scale parameter, allowing the weight placed on the weakest versus the strongest correlations to be adjusted when performing the optimisation – this is done by varying the scale parameter. The effect of different scaling choices is highlighted in Table 3.1. Correlation matrices for two parcels of three items are depicted, with each item loading either exclusively the facets of one factor (left panel), or loading strongly on one domain, but with significant smaller cross-domain loadings (right-hand panel). This latter situation is common among items assessing the FFM, such as the NEO-PI-R (cf: Costa and McCrae, 1992). Graphically, the strongest correlations occur along the main diagonal blocks, with weaker correlations in the off-diagonal blocks.

Table 3.1: Schematic presentation of a correlation matrix of items sorted for their major domain, demonstrating the effect of sigma and of low (left matrix) and high (right matrix) between-domain correlations in determining the solution obtained in spectral clustering.

	A1	A2	A3	N1	N2	N3		A1	A2	A3	N1	N2	N3
A1	1.0	.7	.6	0	0	0	A1	1.0	.7	.6	.2	.3	.2
A2	.7	1.0	.7	0	0	0	A2	.7	1.0	.7	0	.2	.2
A3	.6	.7	1.0	0	0	0	A3	.6	.7	1.0	.2	.2	0
N1	0	0	0	1.0	.7	.6	N1	.2	0	.2	1.0	.7	.6
N2	0	0	0	.7	1.0	.7	N2	.3	.2	.2	.7	1.0	.7
N3	0	0	0	.6	.7	1.0	N3	.2	.2	0	.6	.7	1.0

Note: In a hypothetical test sampling two major domains (here “A” and “N”) with three items per domain, the largest correlations will occur on the blocks on the main diagonal, Correlations in the off-diagonal positions may either be negligible (left matrix) or small to moderate in size (right matrix). In spectral clustering, the scale parameter sigma determines the weight given to smaller correlations. As these are mainly to be found off the diagonal blocks in questionnaire data, in psychometric practice, sigma determines the influence of items with cross-domain loadings. In the case where these are negligible, no effect will occur. However when at least some off-diagonal block correlations are moderate in magnitude, emphasis on the on these may reveal a different structure than the one which emphasized only the strongest correlations on the block diagonal.

This relationship between the strength and location of correlations emphasises why the effect of the scale parameter available in spectral clustering (but not in factor analysis) is of value in interpreting differing theories of personality; If different scales give rise to different numbers of clusters this suggests the existence of theoretically meaningful relationships in the off-diagonal correlations with a structure that is not apparent when only the strongest correlations are considered. Equally, if these off-diagonal associations represent primarily sampling error or item-specific correlations, as would be predicted from the Five Factor Model, then emphasising weaker links to a greater or lesser degree will not affect the structure that is recovered. Thus the method affords the possibility, but not the necessity, of discovering additional meaningful clusters.

One of the primary advantages of spectral clustering over more traditional techniques is that spectral clustering can detect more general kinds of patterns in the data. Techniques such as principal components analysis and factor analysis search for linear relationships in the data, and k-means clustering will optimise only within convex structures, but spectral clustering can solve very general problems like intertwined spirals (Von Luxburg, 2007). Indeed, much of the early work in spectral clustering was in image analysis, where traditional techniques often failed to segment photographs into meaningful chunks (e.g. dog, boy, sky, grass) but where spectral clustering was able to create more intuitive groupings (Von Luxburg, 2007). A finding that structures discovered by spectral clustering provide a better fit to the data than those discovered by linear methods such as factor analysis or principal components suggests that the underlying structure may be nonlinear; this, in turn, has implications for the kinds of models which should be used to fit the data. These issues are considered further in the discussion.

Spectral clustering thus provides a method for revealing structure within personality data which is distinct from that afforded by factor analysis. It can provide both a robustness check when compared to factor analysis, and, because of the control it offers over sensitivity to item relationships, can explore the effect of weaker and stronger sources of variance on models of personality. While we did not predict that the method would favour of any particular alternative model, we were aware that Tiliopoulos et al.'s (2010) results using non-metric multidimensional scaling found support for Eysenck's three factor PEN model over Costa and McCrae's FFM. We thus focused on contrasting the major three-, five-, and six-domain classes of personality model in analyses, comparing results from factor analysis and spectral clustering.

3.2 Method

3.2.1 Measures

Subjects completed 300 items from the International Personality Item Pool's (IPIP) NEO questionnaire, which has been developed to measure the same constructs as the NEO-PI-R (Goldberg et al., 2006). The test includes five domain-level constructs: Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C). The IPIP NEO also measures six facets per domain. The mean correlation of facets on the IPIP proxy with corresponding facets in the NEO-PI-R is 0.94 after correcting for unreliability (Goldberg, 1999).

3.2.2 Participants and procedure

A total of 23,994 subjects participated between August 6, 1999 and March 18, 2000. Of these, 20,993 submissions were retained after excluding subjects for long strings of identical or missing responses or for duplicate submissions. Detailed information about the criteria for excluding responses can be found in Johnson (2005). The final sample was 63.1% female and had a mean age of 26.1 years ($SD=10.7$ years). Subjects were not actively recruited; they discovered the website either on their own or through word-of-mouth.

3.3 Results

The statistical procedure used here closely follows the method used in Leibon et al. (2008), so the reader looking for full details about implementing the procedure is referred there. Here, we wish to go into enough detail to make clear the role of the scale parameter, which plays a central role in the results that follow.

3.3.1 Network construction and the importance of weak connections

We begin by transforming the item correlation matrix into a matrix with a geometric interpretation (and thus no negative values). First, the correlations between items i and j (c_{ij}) are converted to distances on a unit sphere:

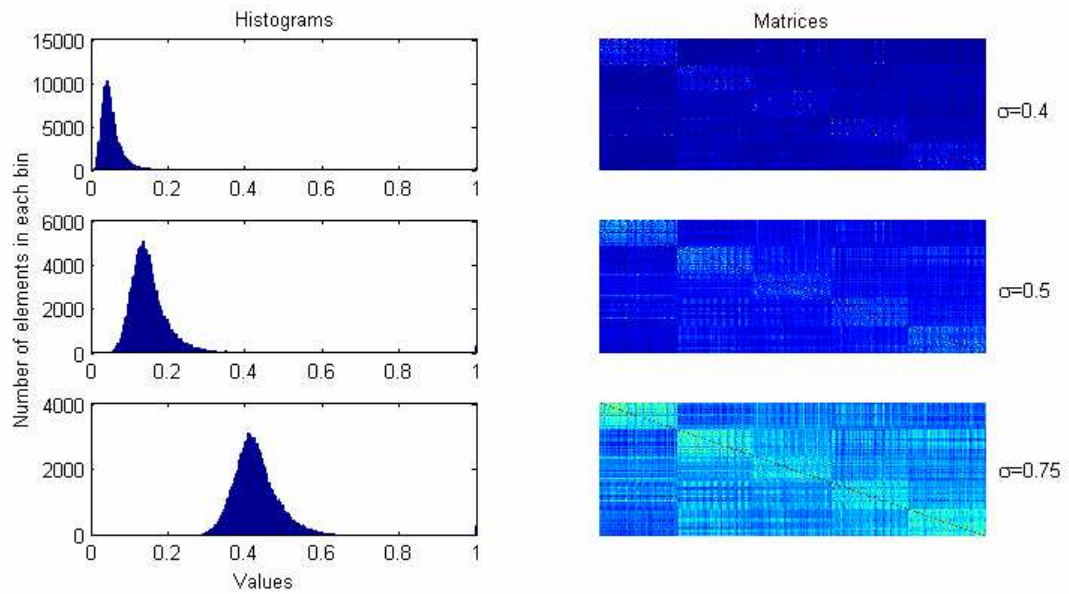
$$d_{ij} = \left(\frac{1 - c_{ij}}{2} \right)^{1/2}$$

This returns a dissimilarity matrix where distances are larger for dissimilar items (with respect to correlation). This is transformed to a similarity matrix, to use as our weighted adjacency matrix A , by applying a Gaussian:

$$a_{ij} = e^{-\frac{d_{ij}}{2\sigma^2}}$$

The parameter σ is the scale parameter – it determines how wide the Gaussian is. The effect of this is to specify how small the distances must be to create a strong edge in the resulting adjacency matrix. Edge strengths range between 1 (the strongest edge) and 0 (absence of an edge). When d_{ij} is much smaller than σ , then a_{ij} lies close to 1. Conversely if d_{ij} is much larger than σ , then a_{ij} is close to 0. Figure 3.1 shows, on the left, the histograms of entries of a_{ij} for three different choices of σ . On the right, we show images of the A matrices themselves – darker colours are close to 0 while lighter colours are closer to 1. Thus smaller σ creates smaller a_{ij} while larger σ creates larger a_{ij} . Intuitively, when small values of σ are used, the network is dominated by a small number of the strongest connections, but as larger values of σ are used, the weaker connections become increasingly important in determining the shape of the resulting network.

Figure 3.1: The effect of changing the scale parameter σ .



Note: The histograms at left show the histograms of the adjacencies a_{ij} for our data as σ increases from 0.4 to 0.5 to 0.75. Notice that the adjacencies spread as the scale parameter increases. The matrices at right are images of the 300×300 adjacency matrices of question responses; lighter colours indicate more similar items. Note that the questions have been sorted so that all 60 Neuroticism questions come before all 60 Extraversion questions, and so on in the order N, E, O, A, C. Thus the block diagonal appearance of the matrices indicates the tight within-factor relationships.

3.3.2 Spectral clustering algorithm

Spectral clustering relies on an analysis of the spectral data (eigenvalues and eigenvectors) of the Laplacian. We form the symmetrised graph Laplacian from the adjacency matrix A ,

$$L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$$

Here, I is the identity matrix and D is the diagonal matrix of degrees of the nodes defined by,

$$d_i = \sum_{j=1}^{300} a_{ij}$$

The optimisation problem for the weighted network encoded in A is to find a decomposition of the nodes of the network into disjoint subsets A_1, \dots, A_k (for given k) that minimises the sum

$$\sum_{t=1}^k \frac{Cut(A_t)}{Vol(A_t)}$$

where $Cut(A_t)$ is the total weight of edges from nodes in A_t out to the rest of the network and $Vol(A_t)$ is the sum of the degrees in the nodes of A_t . As shown by Shih and Malik (2000) a relaxation of this combinatorial problem turns this into an eigenvector problem for this symmetrised Laplacian. To accomplish this we proceed as follows (see Ng, Jordan & Weiss, 2002):

1. Find the eigenvalues and eigenvectors of the symmetrised Laplacian. It is possible to show that the matrix is positive semidefinite – i.e., has only nonnegative eigenvalues. Discard the zero eigenvalues and associated eigenvectors
2. Determine the number of eigenvectors, l

3. Determine the number of clusters, k
4. Use k -means clustering on the embedding of the nodes into \mathbb{R}^l given by the first l (undiscarded) eigenvectors

3.3.3 Parameter estimation

As detailed above, the spectral clustering algorithm uses three parameters: the number of eigenvectors to feed into the clustering algorithm (l), the scale parameter (σ), and the number of clusters (k). The success of the method depends on picking reasonable values for these parameters. While we approach picking these parameters sequentially, it is important to realise that all three are intertwined. For example, at different scales, we may find different good choices for l and k . In a sense, we wish to pick values for these three parameters in conjunction with one another.

3.3.3.1 The number of eigenvectors

In contrast to other spectral methods (such as principal components analysis), the smaller the eigenvalue of the symmetrised Laplacian, the more important it is. The zero eigenvalues correspond to the connected components (two nodes are connected if there is a sequence of nonzero weight links between the nodes, and a connected component is a maximally connected subnetwork) of the networks, naturally thought of as the most obvious form of clustering. As the eigenvalues increase (i.e., move away from zero), they encode a diminishing amount of clustering information. If the spectrum naturally separates into a clump of eigenvalues near zero and a clump separate from that, the “lower” eigenvalues and corresponding eigenvectors give a natural dimensionality reduction in the data.

In picking the number of significant eigenvalue/eigenvector pairs, we use an ad

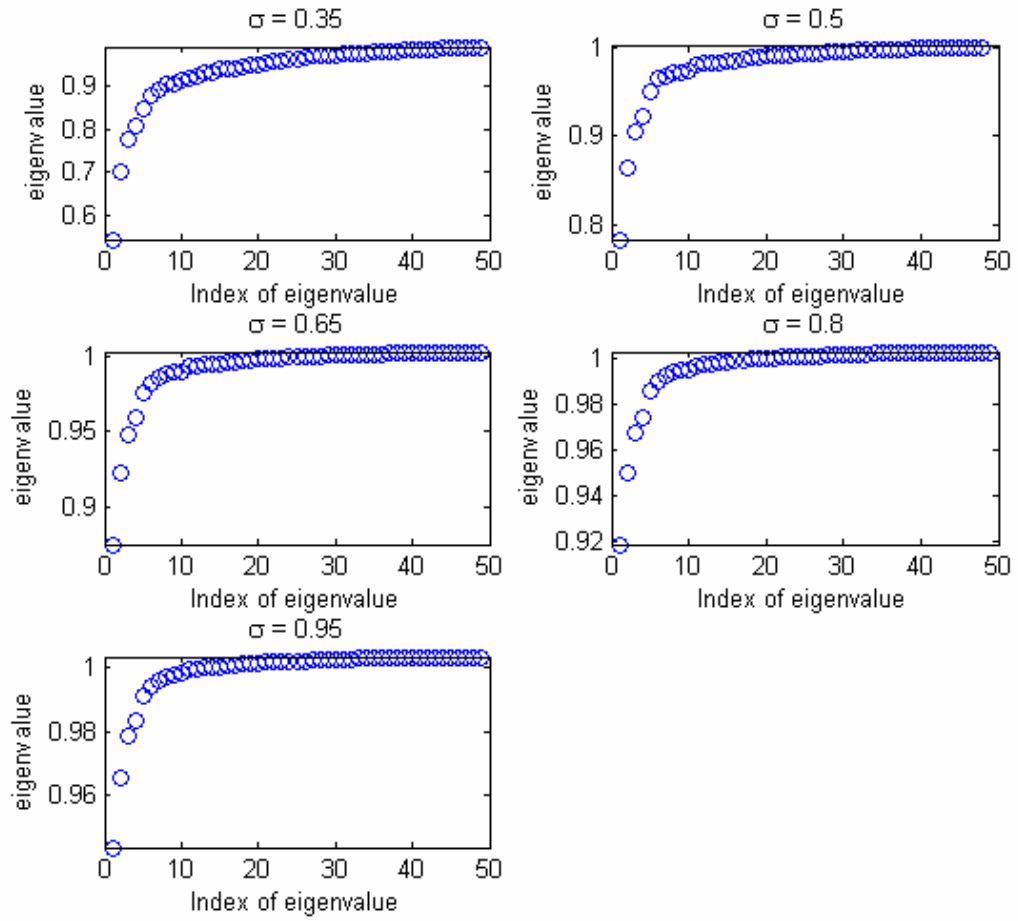
hoc method similar in many respects to the “scree” method used to pick the number of factors in factor analysis. We first note that this will depend on the parameter σ which we vary between 0 and 1. The goal is to look for domains of stability in the spectrum of the symmetrised Laplacian, as measured by the number of small nonzero eigenvalues that are well separated from the “bulk” of the spectrum which generally tend to cluster around 1. Figure 3.2 shows a plot of the nonzero eigenvalues for a range of scale parameters σ from 0.35 to 1. We see that for a significant range of σ , the first four eigenvalues are visually separated from the bulk. We note that for σ less than 0.35, the graph associated with A has more than one connected component, which is implausible. Thus, we will pick $l = 4$ and constrain σ to fall in the range we have considered.

3.3.3.2 The scale parameter

To further estimate σ , we consider the effect of the choice of σ on the optimal number of clusters. To assess this for a given σ , we use the method of “cluster consistency” described as follows. For a fixed l and σ , apply k -means repeatedly with $k_i \in \{2, 3, \dots, k_{max}\}$ (where k_{max} is a reasonable choice for the maximum number of clusters - for this data, we used $k_{max} = 10$) to the l -dimensional spectral coordinates. For a given k , we then use the following procedure to test the consistency of the clustering with k clusters we have computed above.

1. For each k_i , pick 150 questions at random from the total and perform spectral clustering with l fixed and $k = k_i$.
2. Compute the percentage of the 150 questions whose cluster classification differs from the original clustering.

Figure 3.2: Eigenvalue spectra for different values of the scale parameter σ



Note: Eigenvalues of the symmetrized Laplacians are on the vertical axis, and are indexed on the horizontal axis. The first (zero-value) eigenvalue has been omitted in all cases.

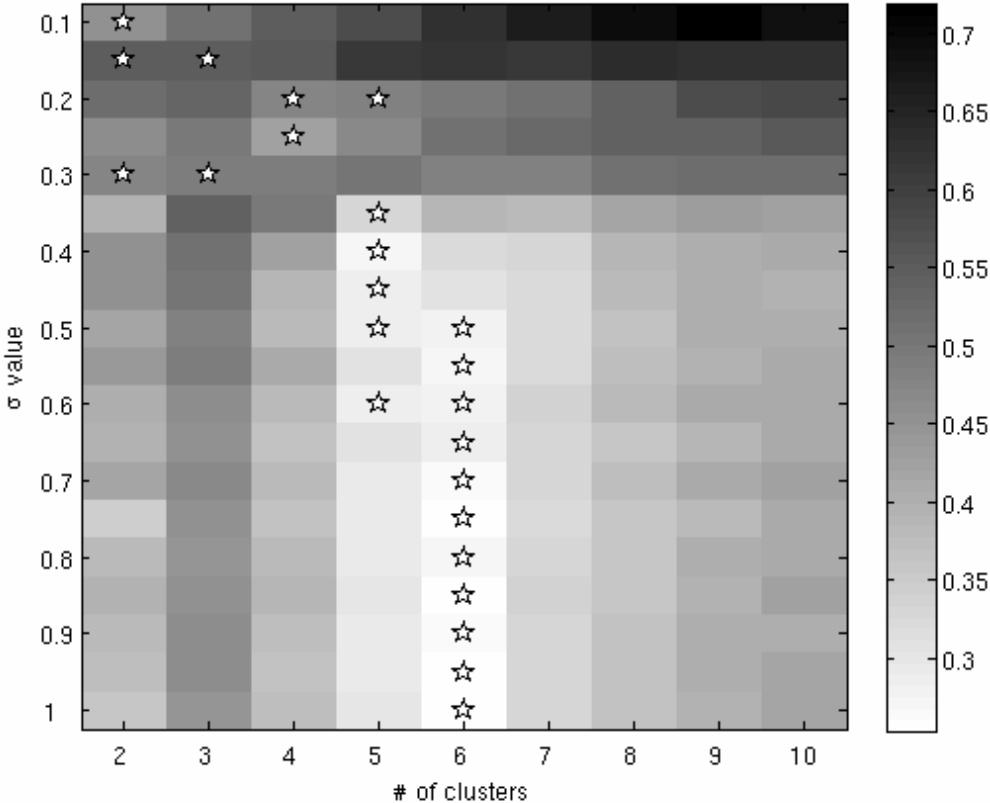
In general we repeat this many times and compare the distributions of reclassifications for all k to select the most consistent clustering (i.e. the k which has the lowest proportion of reclassified items). We denote by n the number of times we repeat the procedure.

In our case, for each $\sigma \in 0.1, 0.15, 0.2, 0.25 \dots, 1$ we use this procedure with $l = 4$ and computed consistency for $k = 2, \dots, 10$ with 100 repetitions of the procedure above for each k . Figure 3.3 shows the mean of the minimum classification error percentages for all the trials for a given (σ, k) pair. In the figure, a cell has a star if it achieves the minimum misclassification error in that row (two stars appear in a row if the minima are not significantly different from one another). Figure 3.4 is closely related—it plots the minimum misclassification error as a function of σ (i.e. it connects the stars). Figure 3.4 shows that classifications become more stable as σ increases up to a value of about 0.4, at which point it levels off with about 30% of items being misclassified. This suggests that we should choose a value of σ in the range 0.4 to 1. Referring back to Figure 3.3, we see that this range appears to be most consistent for a number of clusters— k —equal to five or six. The Discussion will consider the optimal value of k in more depth.

3.3.3.3 The number of clusters

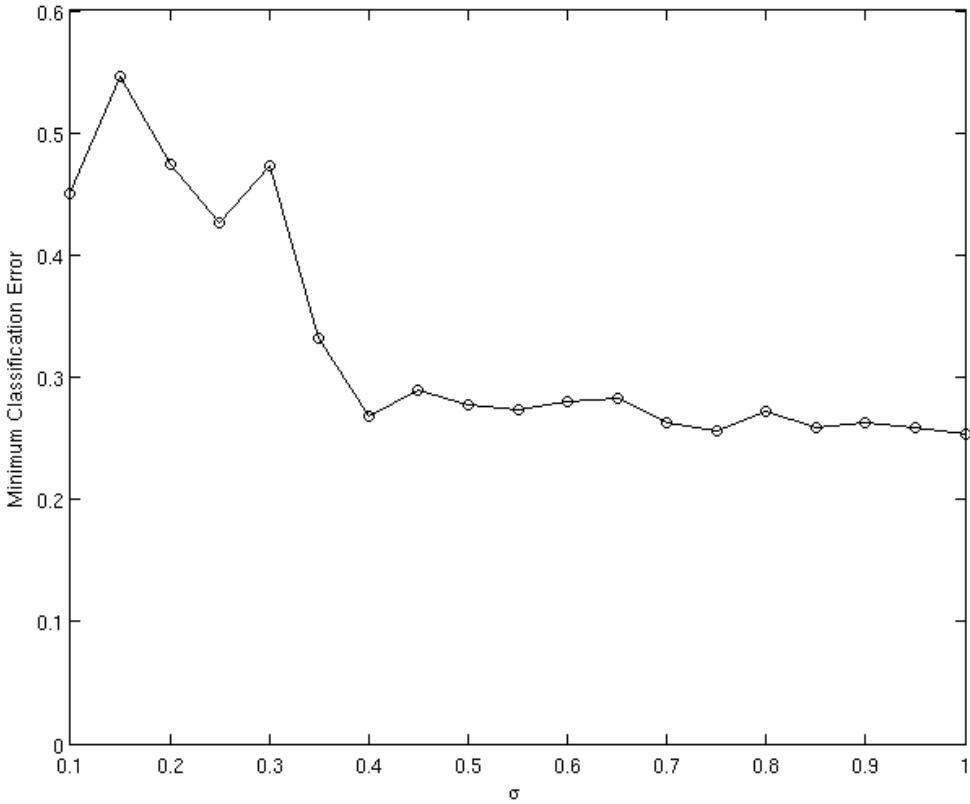
Last, we wish to fix k . This is in some sense the key point of the paper as the number of clusters is the analogue for the number of factors. As noted above, the optimal k depends on the value of the scale parameter σ . For reasonable values of σ (above 0.4 for our data), $k = 5$ or 6 has the lowest classification error for all values of k between 2 and 10, and this holds across a range of values of the scale parameter σ . To check that these are global and not merely local optima, we repeat the method of cluster consistency for larger values of k (we test k up to 40) and with more repetitions per

Figure 3.3: Mean of the minimum classification error percentages for all the trials for a given pair of k (the number of clusters) and σ (the scale parameter).



Note: Stars indicate the lowest value in a given row (multiple stars are used when the values are not significantly different from one another). Comparing the columns for 5 and 6 clusters shows that they are optimal at different levels of the scale parameter

Figure 3.4: Minimum classification error percentages as a function of the scale parameter σ



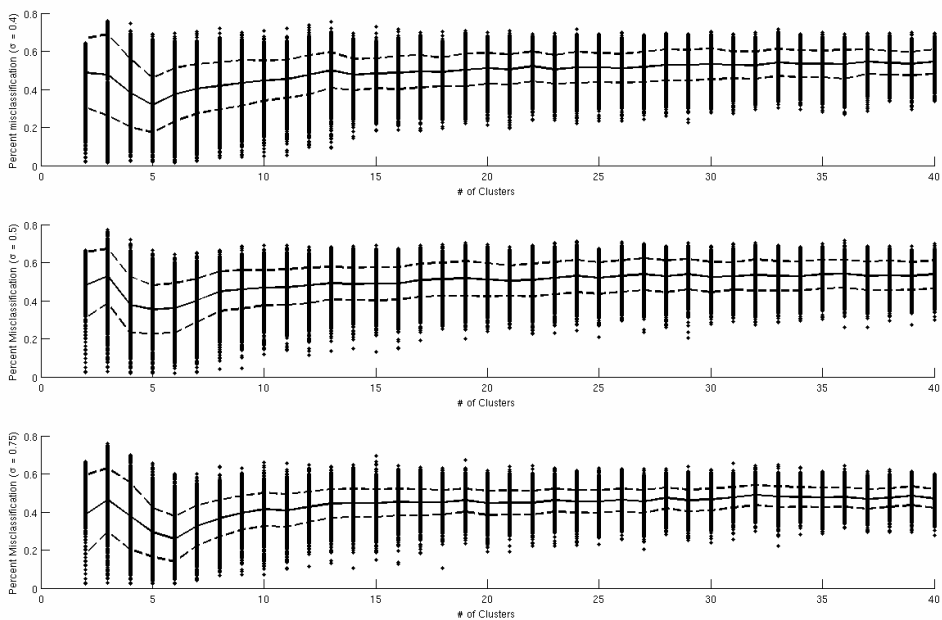
Note: This graph essentially just connects all of the stars in Figure 3.3. Notice that the minimum classification error is stable for values of sigma between 0.4 and 1. Comparing with Figure 3.3, we see that the low end of this stable range suggests five clusters, while the high end of the stable range favours six clusters. This suggests that the number of personality domains is a function of the scale at which we look at the data.

k (200, instead of the previous 100). We set $l = 4$, $k_{max} = 40$, $n = 200$, and σ to each of 0.4, 0.5 and 0.75. We set k_{max} to ensure that the largest value of k included in our tests substantially exceeded the number of postulated facets in the FFM (30). Figure 3.5 shows the results. In each pane, the horizontal axis shows the values of k while the vertical shows the percentage of misclassifications. Each circle represents one trial. The solid line is the mean while the dotted lines are the mean plus or minus one standard deviation. The top, middle and bottom panes show the results for $\sigma = 0.4$, 0.5 and 0.75 respectively. The fact that the misclassification rate rises and then flatlines above $k = 6$ confirms that our optima are indeed global. Table 3.2 gives the numerical values of misclassification for various k and σ .

While we do not focus here on the facet-level of personality, it is noteworthy that while Figure 3.5 shows evidence for structure at the domain level (i.e., five or six major clusters), there is no indication of facet-level structure within the data. Particularly, there is no obvious drop in the proportion of reclassified items around 30 clusters, which is what we would have expected if there had been measurable structure at the facet level.

Two optimal values for k – six and five– emerged at different values of σ (above and below about 0.5, respectively). It should be stressed that both values of k are valid and both are reflected in the data. When more emphasis is put on the strongest correlations (by choosing a low value of σ) then the five-cluster solution is preferred, but when more weight is put on some of the weaker connections (by using a higher value of σ), then the six-cluster solution is preferred. The issue of which value of k is preferred overall is considered further in the discussion.

Figure 3.5: Proportion of questions misclassified



Note: The vertical axis shows the proportion of items which were differentially classified by clustering on various randomly chosen 150-item subsets of the original questionnaire versus clustering based on all 300 questions. The horizontal axis shows the number of clusters. Each circle is a different trial. The solid line represents the mean misclassification while the dotted lines are plus or minus one standard deviation. The value of the scale parameter used was 0.4 in the top pane, 0.5 in the middle pane, and 0.75 at the bottom. The cluster consistency test was run 200 times for each value of k .

Table 3.2: Classification error as a function of the scale parameter σ .

Classification error when sigma = 0.4

	Number of clusters				
	k = 3	k = 4	k = 5	k = 6	k = 7
mean misclassification rate	0.48	0.38	0.32	0.37	0.4
standard error	0.0035	0.0028	0.0024	0.0022	0.0021

Classification error when sigma = 0.5

	Number of clusters				
	k = 3	k = 4	k = 5	k = 6	k = 7
mean misclassification rate	0.56	0.38	0.35	0.36	0.4
standard error	0.0022	0.0024	0.0021	0.0021	0.0017

Classification error when sigma = 0.75

	Number of clusters				
	k = 3	k = 4	k = 5	k = 6	k = 7
mean misclassification rate	0.46	0.38	0.29	0.26	0.33
standard error	0.0027	0.0028	0.0021	0.0019	0.0016

Note : This table shows the mean and standard error of the proportion of items which were misclassified for different values of σ and k . Each entry is based on 4000 simulations, and so the standard errors are small enough to ensure that all the differences within each pane are statistically significant at all conventional levels.

3.3.4 Comparison of domains from spectral clustering and factor analysis

We next compared five- and six-cluster solutions to five- and six-factor solutions obtained using varimax-rotated factor analysis. In each case, the factor solution was calculated by assigning each item to the factor on which the item has its highest loading in a varimax rotation. The cluster solution is calculated as the configuration of item assignments with the lowest sum of squares out of 10,000 runs of the spectral clustering algorithm described above.

The five-cluster and five-factor models were very similar: of the 300 items on the questionnaire, 286 were assigned identically by the factor and cluster solutions. Both solutions closely resembled the conventional Five Factor Model domains of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness.

The six-factor and six-cluster analyses gave sharply contrasting solutions. The factor analytic sixth factor contained just seven items (4 from facet N4: Self-Consciousness, and 3 from other facets of Agreeableness and Neuroticism). By contrast, the six-cluster solution yielded a large - 34 item - sixth cluster. This cluster comprised 18 items originally assigned to Conscientiousness, 15 items originally assigned to Agreeableness and a single item originally from Neuroticism. Thus while the sixth factor is a small nuisance factor with a majority of members from a single facet, the sixth cluster was large and meaningful.

The sixth cluster is highly similar to the HEXACO Honesty-Humility factor, perhaps surprisingly so given that we worked from an item bank chosen to model the NEO-PI-R, which may not adequately sample the four facets of Honesty-Humility (cf: Ashton & Lee, 2005). That the sixth cluster consisted almost entirely of Agreeableness and Conscientiousness items supports Lee and Ashton's observation that

correlations of their Honesty-Humility factor with FFM domains are primarily with Agreeableness and Conscientiousness (Lee & Ashton, 2004). 13 out of the 15 Agreeableness items re-assigned to the Honesty-Humility cluster were drawn from the A2 (Morality) and A5 (Modesty) facets of Agreeableness: Individual items re-assigned to the sixth cluster included “*Cheat to get ahead*” (A2; reversed), “*Take advantage of others*” (A2; reversed), “*Dislike being the centre of attention*” (A5), and “*Seldom toot my own horn*” (A5). Of the 18 Conscientiousness items that were reassigned to the sixth cluster, 9 came from C3 (Dutifulness), which contains items such as “*Keep my promises*”, and “*Tell the truth*”, further reinforcing the impression that the sixth cluster is Honesty-Humility. The lone Neuroticism item reassigned to the sixth cluster was from N4 (Self-consciousness), “*Am afraid to draw attention to myself.*”

The other five clusters remained similar to the original FFM domains, with notable exceptions. Openness lost most of O3 (Emotionality); 4 items went to the Neuroticism/Emotionality cluster, namely “*Experience my emotions intensely,*” “*Seldom get emotional,*” “*Am not easily affected by my emotions*” and “*Experience very few emotional highs and lows.*” In addition, 2 of the O3 items went to the Agreeableness cluster: “*Feel others emotions*” and “*Don’t understand people who get emotional.*” O3 items retained by Openness included “*Enjoy examining myself and life*” and “*Try to understand myself.*” Another change from the FFM was that Extraversion lost all 10 items from the facet E4 (Activity Level) to Conscientiousness. E4 includes items such as “*Am always busy*” and “*Like to take it easy.*” Extraversion also lost 4 items from E3 (Assertiveness) to Conscientiousness, including “*Take charge*” and “*Take control of things.*”

3.4 Discussion

3.4.1 Main findings

The chief results of the present study were fourfold. First, despite its very different approach when compared to the algorithm of factor analysis, spectral clustering yielded a five-cluster solution highly similar the five-factor solution from factor analysis (with both reflecting the FFM). Second, spectral clustering yielded a six-cluster solution corresponding to that of the HEXACO model, not only in having a sixth cluster, but in the nature of items comprising the sixth cluster, and in the effects on the remaining 5 clusters representing Extraversion, Neuroticism/Emotionality, Openness, Agreeableness and Conscientiousness. By contrast, factor analysis yielded a six-factor solution containing a small and hard to interpret sixth factor. Third, the facets underlying the FFM domains altered in notable ways, refocusing these traits. Finally, the results highlight the importance of differentially weighting larger or smaller correlations amongst items, respectively, in determining whether five- or six-domain solutions fit best. Spectral clustering, then, may provide useful traction in resolving questions of the structure of personality.

The finding that whether five- or six-cluster solutions were preferred depended on the weighting given to the strongest connections among the data (more weight favouring a five-cluster FFM solution, less weight favouring a six-cluster HEXACO solution) casts light on the distinction between these two competing models for the structure of personality. There is no necessary reason to prefer one set of weights (and hence a particular value of the scale parameter σ) over another, so we do not claim that the use of spectral clustering can end the dimensionality debate (though most values of σ resulted in the six-cluster HEXACO solution). However, the contribution of this paper is to take this debate in a new direction by showing that

the dimensionality issue is tightly linked to the issue of the importance of weaker connections amongst items.

That the five-cluster model reflects the domain structure postulated by the classic FFM is significant given the fact that these two statistical techniques – factor analysis and spectral clustering – are based on different transformations of the raw data, and have different optimisation targets. The present results, then, represent a robustness check on the FFM, which the FFM passes. This finding is in contrast with Tiliopoulos et al. (2010), who found support for lower dimensional two- or three-cluster solutions. These authors used non-metric multidimensional scaling (NMDS) on facet-level data from 384 subjects' responses on the NEO-PI-R. The domains they supported were similar to the three factors of Eysenck's PEN model (Eysenck 1991, 1992). Two potential origins of this difference in outcome between the present result and that of Tiliopoulos et al. (2010) may be relevant. Tiliopoulos et al., constrained by a smaller number of subjects, analysed their data at the facet level. Here, with 20,993 subjects, we found evidence at the item level for facets loading on multiple domains: O3 for instance contained items that were clustered with Emotionality/Neuroticism, Agreeableness and Openness. These distinct patterns among items within a facet are confounded in analyses at the facet level. Second, and perhaps more importantly, in algorithms such as spectral clustering and NMDS negative and positive correlations are treated asymmetrically, such that a correlation of minus one is more similar to zero than is plus one. Because Neuroticism items correlate slightly negatively with items from the other four domains, if Neuroticism items are not reverse scored, they have an artificially high distance from the other four domains. To illustrate the effect of this transformation, as well as to place our results in the context of those of Tiliopoulos et al., we performed NMDS on both the original data set and the reverse-coded set; both NMDS and spectral clustering of

the raw data gave evidence for two clusters (“Neuroticism” and “not-Neuroticism”) and somewhat weaker evidence for five. Taken together, the relatively lower n , consequent need to use facet-level analysis, and, most importantly, reversal of Neuroticism prior to analysis seem likely to account for the differences between the present result and that of Tiliopoulos et al. (2010).

3.4.2 Differences between the six- and five-cluster solutions: Psychological implications

As might be expected for competing models of the same data, the FFM and six-cluster HEXACO solutions extracted here are similar in many important ways. The five shared domains are similar in both solutions. Neuroticism remained almost unchanged, and near identical to the corresponding FFM dimension. Extraversion retained most of its items, but lost four items from E3 (Assertiveness) and the entire “Activity Level” facet (E4) to Conscientiousness. Extraversion was thus focused on warm sociability and excitement seeking.

Openness retained most items, including the more intellectual items of facet O3 (Openness to Emotion) such as “*Try to understand myself*”, but lost the remaining content of O3 in meaningful ways to Neuroticism/Emotionality and Agreeableness (see results section). It thus lost connection to the direct experience of emotion, including empathy, and was refocused on understanding emotion.

The Agreeableness cluster changed in ways highly consistent with the theoretical predictions of the HEXACO model (Ashton and Lee, 2005), losing A2 (Morality) and A5 (Modesty), and thus, together with gaining empathy-related content from O3, focused more clearly on compliance and pro-sociality. The movement of hostility items from Neuroticism to Agreeableness as predicted by Ashton and Lee (2007) did not occur.

Conscientiousness gained the E3 (Assertiveness) and E4 (Activity Level) items from Extraversion as noted above, and lost most of C3 (Dutifulness) and C6 (Cautiousness) to Honesty-Humility, thus becoming refocused towards task engagement and away from integrity, as predicted by Ashton and Lee (2007).

3.4.3 Differences between the six- and five-cluster solutions: Economic implications

What do the results of the present analysis imply for economic modelling? At the outset, it's worth reiterating a point stressed in chapters one and two of this thesis: the central results of personality psychology have yet to be incorporated into mainstream economic models. Thus any changes to those psychological results will not have much effect on economics as it has been practised heretofore. But in a world where the dimensionality of personality is taken as a serious constraint on economic theory, the present results would allow additional freedom of movement in two respects. Firstly, moving to the HEXACO from the FFM would allow the "budget constraint" on the number of independent parameters which theorists can propose to be increased from five to six. Secondly, moving away from linear factor models and towards (potentially) nonlinear models from spectral clustering would allow economists greater scope to model both preferences and the relations between them in nonlinear ways.

3.4.4 Strengths and Weaknesses

Among the strengths of the analyses were the very large sample size, a comprehensive item battery, the fact that the item pool was chosen to represent the five- rather than six-factor domains, the use of multiple methods, and ability to contrast different

parameterised solutions within the spectral clustering solutions. Weaknesses include reliance on one dataset: While these data are very large, and we were able to recover and validate solutions using resampling techniques within this parcel of subjects, it would be valuable to examine the results of spectral clustering in data derived from other cultures and subject pools, test formats, and, particularly, on the items of the NEO-PI-R itself. External validation in terms of differential validity of the five and six cluster solutions in genetic or experimental studies would also be valuable.

3.4.5 Summary

To summarise, the results from applying spectral clustering to a personality dataset were that we were able to recover the five-factor model despite a very different analytic approach and optimisation target. Large elements of consistency were found among the five canonical domains of Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness whether extracted using the traditional factor analytic approach or spectral clustering. Giving weight to reliable but moderate effect-size covariance amongst items supported a six-factor HEXACO structure of personality. The effect of the six-cluster model was to sequester dutifulness, cautiousness, morality and modesty items from Conscientiousness and Agreeableness to a new Honesty/Humility cluster. Agreeableness, without its elements of immorality was refocused onto a dimension of compliance versus what Sell, Tooby, and Cosmides (2009) termed “formidability” - a willingness to force others to re-calibrate upwards the weight they assign to your welfare. Assertiveness and activity were shifted from Extraversion to Conscientiousness, refocusing Conscientiousness on competent, active striving, and focussing Extraversion on warm, gregarious excitement seeking and positive emotion. Additional effects included removing emotion-experiencing items from O3 to the domain most relevant to the emotion; either Neuroticism or

Agreeableness. Whether five- or six-cluster solutions were favoured depended on the emphasis given to weak relationships; but most solutions favoured the HEXACO model.

Acknowledgements

I would of course like to thank my coauthors Scott Pauls, Tim Bates and Dan Rockmore. I would also like to thank the Santa Fe Insitute for providing the fertile environment which got this work started. Finally, I would especially like to thank Professor John Johnson of Pennsylvania State University for making his data available to perform this analysis.

Bibliography

- [1] Ashton, M. C., & Lee, K. (2005). Honesty-humility, the Big Five, and the Five-Factor Model. *Journal of Personality*, 73(5), 1321-1354. doi: 10.1111/j.1467-6494.2005.00351.x.
- [2] Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150-166. doi: 10.1177/1088868306294907.
- [3] Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, 87(5), 707-721. doi:10.1037/0022-3514.87.5.707
- [4] Cattell, R. B. (1978). *The Scientific Use of Factor analysis in Behavioral and Life Sciences*. New York: Plenum Press.
- [5] Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory- Professional Manual*. Florida: Psychological Assessment Resources Inc.

- [6] Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences*, *12*(8), 773-790. doi:10.1016/0191-8869(91)90144-Z
- [7] Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*(6), 667-673. doi:10.1016/0191-8869(92)90237-J
- [8] Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit, reliability, and the five-factor model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*(5), 1051-1062. doi: 10.1016/j.paid.2007.02.024.
- [9] Goldberg, L. R. (1990). An alternative “description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229. doi:10.1037/0022-3514.59.6.1216
- [10] Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- [11] Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84-96. doi:10.1016/j.jrp.2005.08.007

- [12] Johnson, J. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103-129. doi: 10.1016/j.jrp.2004.09.009.
- [13] Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202. doi:10.1007/BF02289343.
- [14] Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO – personality inventory. *Multivariate Behavioral Research*, *39*(2), 329. doi:10.1207/s15327906mbr3902_8
- [15] Leibon, G., Pauls, S., Rockmore, D., & Savell, R. (2008) Topological structures in the equities market network, *Proceedings of the National Academy of Sciences*, *105*, 20589—20594.
- [16] Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., et al. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471-491. doi: 10.1037/a0019227.
- [17] McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509-516.
- [18] McCrae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood* (2nd ed.). New York, NY, US: Guilford Press.
- [19] McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175-215.

- [20] Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On Spectral Clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, *14*, 849-856.
- [21] Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106(35)*, 15073-15078. doi: 10.1073/pnas.0904312106
- [22] Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22(8)*, 888–905. doi:10.1109/34.868688
- [23] Spearman, C. (1927). *The Abilities of Man, Their Nature and Measurement*. Macmillan New York.
- [24] Tiliopoulos, N., Pallier, G., & Coxon, A. P. (2010). A circle of traits: A perceptual mapping of the NEO-PI-R. *Personality and Individual Differences*, *48(1)*, 34-39. doi: 10.1016/j.paid.2009.08.010.
- [25] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17(4)*, 395—416. doi:10.1007/s11222-007-9033-z

Chapter 4

Personality as Strategy: The Evolution of Extraversion

Extraversion, like all personality traits, shows considerable variance within the population. This paper develops an evolutionary model to account for that diversity. I suggest that in a general contest framework, when there are costs to randomisation, then something akin to the extraversion continuum is the inevitable. The basic model in this paper is a two-player symmetric war of attrition, but it is extended to account for asymmetries and for multiple players. The model implies that extraverts take more risks and that there is a U-shaped relationship between extraversion and loss aversion.

"Men are born equal but they are also born different"

–Erich Fromm

4.1 Introduction

4.1.1 The evolution of individual differences

People differ, and their different personalities are important because these differences, as measured by self report, are correlated with a range of important life outcomes including educational attainment (Goldberg, Sweeney, Merenda et al. 1998; van Eijck and de Graaf, 2004), college grade point average (Almlund, Duckworth, Heckman, and Kautz, 2011), the probability of remaining self-employed (Caliendo, Fossen and Kritikos, 2008), occupational choice (Barrick and Mount 1991; Ham, Junankar and Wells, 2009), health choices (Heckman, Stixrud and Urzua, 2006), longevity (Roberts, Kuncel, Shiner et al. 2007) and criminal activity (John, Caspi, Robins et al., 1994). But where do personality differences come from? While personality differences have been a major research area on-and-off for nearly a hundred years, it is only recently that researchers have begun to place the study of personality within an evolutionary framework. Daniel Nettle (2006) argues that for each of the Big Five personality factors¹ recognized by psychologists, there is some sort of evolutionary tradeoff which maintains variation in the population (see Kanazawa 2011 for a recent review of this literature).

This paper will outline (to my knowledge) the first formal model of these tradeoffs, generalising from Nettle's (2005) reasoning about the evolution of extraversion.

¹The five are: conscientiousness, agreeableness, neuroticism, openness, and extraversion (CANOE).

Broadly, Nettle suggests that increasing extraversion results in greater short-term mating success, but at greater cost in terms of risk of injury and neglected offspring. The same tradeoff motivates this paper. The model involves players competing for resources in a war of attrition. Fertility, as always, drives the game's replicator dynamics. I discuss three cases: the two-player symmetric, two-player asymmetric, and n -player symmetric games. The solution to each of these games is an evolutionarily stable state (ESS) involving a continuum of actions. These games were first solved in the 1970s and 1980s (see references below) and so their solutions are not original to this paper. Instead, the contributions of this paper are: to apply formal evolutionary game theory to explain the origins of a personality trait, to show that the different versions of the war of attrition can be linked by a single distribution of types and a single elementary bidding rule, and to derive the implications of this for individual differences in both risk aversion and loss aversion.

4.1.2 Strategic interaction and personality differences

Before getting into details, it is worth briefly considering why we should see variation at all. Why hasn't the population evolved to a common optimum? Why do personalities differ?

Some characteristics, such as thirst, are conspicuous in their uniformity. We all seem to feel thirsty in the same way, and for the same sorts of reasons². It is easy to imagine why this is the case, since those of our would-be ancestors who were not thirsty enough died of dehydration, and those who thirsted too much either died of hyperhydration, or were preyed upon when they exposed themselves too often at the

²McKinley and Johnson (2004) note that "small increases of 1-2% in the effective osmotic pressure of plasma result in stimulation of thirst in mammals." This seems to be a universal trait, and indeed the relevant brain area, the *lamina terminalis*, which influences thirst, is notably monomorphic.

water's edge. It seems only natural that the whole population would evolve to crave roughly the correct amount of water, and so there should be little diversity among (healthy) people in this regard.

Yet there are other characteristics, such as ambition, which are remarkably variable across people. Shouldn't there be an ideal amount of ambition, to which the whole population will evolve? If not, why not? To answer that question, we must consider the strategic nature of the two situations. Thirst arises in a non-strategic setting—my need to restore homeostasis is independent of my neighbours' behavior (as long as water isn't scarce). So in this case, one genotype is pushed to fixation; this is the inevitable result in non-strategic population genetics (see, e.g., Nowak 2006, Ch6). Ambition, however, usually plays itself out in a strategic game—how hard you are willing to fight for dominance may well affect my own willingness to fight. If I think you're a pushover, then I may challenge you, but if I think you're a hyper-competitive daredevil, then I won't. There need not be one optimum behaviour, since pushovers may do well in a world full of daredevils, and daredevils may do well in a world full of pushovers. What we are likely to see in this daredevil-pushover game is a population containing some daredevils and some pushovers. This is the insight from John Maynard Smith's famous hawk-dove game (Maynard Smith, 1974) in which the unique evolutionarily stable state (ESS) involves both strategies, so that the population contains some hawks and some doves³. If we use the hawk-dove game

³In the Maynard Smith's basic model, the symmetric payoff matrix, depicted here for the row player, is:

$$\begin{array}{cc} & \begin{array}{c} h \\ d \end{array} \\ \begin{array}{c} h \\ d \end{array} & \begin{array}{cc} \frac{V-C}{2} & \frac{V}{2} \\ 0 & \frac{V}{2} \end{array} \end{array}$$

(it is assumed that $\frac{V-C}{2} < 0$). The mixed strategy ESS occurs when the frequency of hawks is $\frac{V}{C}$.

There are two interpretations of the equilibrium: in the *polymorphic* case, a fraction $\frac{V}{C}$ of the population plays *hawk* all the time, and a fraction $1 - \frac{V}{C}$ plays *dove* all the time. In the *monomorphic* interpretation, all members of the population are the same, and choose their strategies randomly before each fight, selecting *hawk* with probability $\frac{V}{C}$. As Gintis (2009) notes, the polymorphic

as our metaphor for conflict, we should *not* expect to see the population converge to a single, ideal level of ambition—we expect diversity. And the hawk-dove game is a good metaphor for conflict, especially once the action space is generalised from a binary choice to a continuum, at which point it becomes the war of attrition⁴.

4.1.3 The war of attrition as a metaphor for conflict

The war of attrition (described in detail in section 4.3.3 below), supports multiple interpretations, but here it is taken to represent a competition that our ancestors got into where there was some resource at stake, and the players had to compete to see who would get it in such a way that the losers still had to bear some cost from competing⁵. The interesting feature of games like this for the purpose of this paper is that there is no pure strategy equilibrium; the only stable outcome is when players randomise. This is interesting because it suggests that a fairly general situation which must have occurred frequently in the environment of evolutionary adaptedness (i.e. ancestors competing in games with non-refundable bids) should have given rise to behavioural diversity which is a candidate for personality variation.

There is more than one way to get a mixed strategy, though: monomorphism and polymorphism. The monomorphic equilibrium is symmetric: all players use a mixed strategy in which all supported actions are played at the equilibrium frequencies. The polymorphic equilibrium is asymmetric: each player always plays the same action, and players' types are distributed according to the equilibrium density function.

equilibrium is more likely in the case where there is a cost to randomizing. This paper assumes the polymorphic outcome throughout.

Note also that hawk-dove is isomorphic to the game of chicken.

⁴There are of course other possible sources of individual variation. Penke et al. (2007) notes that selective neutrality (i.e. evolutionary irrelevance) and mutation-selection balance could also maintain variation, but that these are not likely to be the relevant forces in this case.

⁵The war of attrition is like a second-price all-pay auction. As discussed in section 4.3.3 below, the qualitative insights of this paper apply equally to the first-price all-pay auction, but the parameters and distributions end up being slightly different.

Throughout this paper I will focus on the polymorphic equilibrium for both empirical and theoretical reasons. Empirically, the polymorphic interpretation is more consistent with the evidence of behavioural stability cited above, and theoretically, the polymorphic equilibrium is more likely when there are costs to randomising⁶.

So when players compete in a contest with non-refundable bids, the result is a polymorphic equilibrium where the different strategic choices can be interpreted as personality differences similar to the extraversion continuum. That, in a nutshell, is the core argument of this paper.

4.1.4 Extraversion as a simple decision rule

After laying out the basic model, it extended to show how the expected variance of outcomes relates to personality, how asymmetries between contestants can be dealt with and how the incorporation of multiple players affects the outcome.

Looking at the variance of outcomes allows the derivation of a well-supported prediction about extraversion (and particularly about the sensation-seeking component at the core of extraversion, which is discussed in more detail in the section 4.2): that it is correlated with risk aversion⁷ (Aluja, García and García, 2003). This relationship is predicted based on the fact that the players who persistently submit the highest bids in the contest (i.e. the sensation-seeking extraverts) also end up with the greatest variance of outcomes.

Introducing asymmetries into the model produces a novel prediction: that there is a *U*-shaped relationship between extraversion and loss aversion. Further details

⁶I am thinking of thinking costs. If all of your opponents are using mental energy to produce serially-independent randomisations from the appropriate exponential distribution but you just do the same thing every time, then you will get a better (net) payoff on average. If everyone does this, then the equilibrium is polymorphic.

⁷As well as the relationship noted in the text, risk aversion has also been found to be positively related to other personality traits. See Almlund, Duckworth, Heckman and Kautz (2011, table 6) for an overview.

are given in sections 4.3.6 and 4.3.7.

Introducing multiple players changes the outcome very little. After each of the extensions mentioned above, we end up with an equilibrium where players are characterized by a type θ from the exponential probability density function $e^{-\theta}$. A player i 's idiosyncratic value θ_i represents his personality—high values of θ correspond to extraverts, and low values of θ correspond to introverts. Given their diverse personality types, all agents use the same basic decision rule to choose their level of effort: $\theta \times (\text{prize value}) \times (\text{number of opponents})$. The bidding rule is identical across games, as well as across types.

Before getting to the model, I will review some of the evidence for the perspective taken in this paper, beginning with a brief summary of what is known about the personality trait of extraversion.

4.2 The nature of extraversion

Extraversion is the most widely accepted of the higher order personality traits, and its formal definition is similar to the popular one: an extraverted person has an outward focus—they are generally more social, lively, assertive, dominant, and oriented towards sensation than their more introverted fellows (see, e.g., Matthews, Deary, & Whiteman, 2009). The word itself comes from the Latin *extr* (outwards) *versin-em* (from the verb "to turn"); introversion has a correspondingly inward-turning etymology.

Nettle (2005) notes that recent evidence supports the view that "extraversion is a consequence of the strength of response to naturally rewarding stimuli, such as sex, food, or physical exhilaration. For the extravert, the salience of these rewards is greater than for the introvert, with the result that extraverts invest more time

and energy on them". Indeed, he finds that high extraversion scores are positively correlated with ambition, competitiveness, interest in sex, and yearning for fame, as well as risk of hospitalization and susceptibility to addiction. Furthermore he notes that

Increasing extraversion is associated with increasing desire to take risks, explore new environments, and compete for status. In tandem, it is associated with seeking varied mating opportunities, including extra-pair copulations (especially for men) and serial monogamy (for women). These strategies will respectively increase the number and genetic quality of offspring. Increasing extraversion thus increases fitness by promoting social dominance and mating success. However, high extraversion levels entail the risk of physical harm and, possibly, reduce investment in the protection of existing offspring. The dimension can therefore be conceived as a continuum along which different fitness costs and benefits are traded off.

Lending further support to the view that the defining feature of extraversion is reward sensitivity, Lucas et al. (2000) tested subjects from 39 nations using popular scales for extraversion, and found that "sensitivity to rewards, rather than sociability, forms the core of extraversion". Indeed, they find that "extraverted participants...reported more pleasant affect even when alone". This suggests that all of the other facets associated with extraversion, such as dominance, self-confidence, warmth, and sociability are simply by-products of the fact that extraverts are more sensitive to rewards.

Following the Lucas et al. result, I will measure both extraversion and reward sensitivity with a single parameter θ . High values of θ are associated with increased sensitivity and greater extraversion, low values of θ correspond to low sensitivity

to rewards and, consequently, introversion. In equilibrium, the values of θ are generated by the probability density function $e^{-\theta}$ on $[0, \infty)$ ⁸. Once the players are competing, θ is used in the bidding rule for player i $bid_i = \theta_i \times (\text{prize value}) \times (\text{number of opponents})$. In the asymmetric case, there is also an intermediate step where the players gauge their roles, but the decision rule is the same.

It is straightforward to see that the bid function reproduces some of the most salient features of extraversion. Most basically, different members of the population exert more or less effort in social situations depending on their value of θ . Also, the extent to which the person exhibits extraverted behaviour depends on what's at stake. When potential gains are larger, all players exert more effort. What makes one person more extraverted than another, then, is that the more extraverted person occupies a relatively higher place in the distribution of bids for any given prize value. But this will not be true in all situations. As you will see in section 4.3.6, asymmetries can cause even extreme extraverts to behave in ways that look more characteristic of introverts (and vice versa).

4.3 The model

4.3.1 Resources increase fertility

In any evolutionary game, the prize is fitness. The present model uses the simplest possible fertility function

⁸It is perhaps worth mentioning that personality traits such as extraversion, when measured, are usually found to be roughly normally distributed, rather than exponentially distributed as in this paper. The exponential distribution, however, is merely an artifact of the assumption that cost increases linearly with time. If the cost function were convex (e.g. $(\text{time})^2$), then the distribution of types would be more hump-shaped, and hence more normal. See Norman, Taylor, & Robertson (1977).

$$\phi(V) = V$$

Where ϕ represents fertility and V represents resources. The function $\phi(V)$ should not be thought of as representing total fertility, but merely incremental fertility gains from the resources which are being competed for. Note that while the linear function is clearly an oversimplification, Maynard Smith (1982) showed that the function can, without loss of generality, be replaced with any monotonically increasing function.

4.3.2 Standard simplifying assumptions

Before moving on to describe the nature of contests in this model, it is worth dwelling on a few of the implicit assumptions. This model makes three simplifying assumptions which are common in evolutionary models: that contestants are randomly matched, that they come from an unlimited population, and that they reproduce asexually.

Contestants are unrelated because personality traits are, in the words of McAdam (1995) the "psychology of the stranger" and they are not meant to explain the particulars of intra-familial relations and their effects on inclusive fitness.

The assumption of an unlimited population is for expediency only; in this case, as in many others, the essential conclusions can be reproduced, with minor caveats, for finite populations. For more about the war of attrition in a finite population, see Just and Zhu (2004).

The assumption of asexual replication has a two-part justification. Part one is that it is simpler. Part two of the justification is that it may actually be a virtue of such a model to consider the sexes separately. In nature, risk attitudes and personality factors can be differentially selected for between the sexes; sometimes

selection pressures can even point in opposite directions, softening the expression of a trait in one sex while making the same trait more prominent in the opposite sex⁹. In humans, for instance, it is widely reported that risk attitudes differ between men and women (e.g. Powell & Ansic, 1997); such an outcome may well have been produced if men and women played similar but separate games.

4.3.3 Mechanics of the war of attrition

The contest in this paper is an all-pay auction, which in an evolutionary context is more realistic than the alternative of refundable bids. All-pay auctions can be first-price or second-price; both of these are qualitatively similar in that they have only mixed strategy equilibria, and so either one would yield a polymorphic population. In this paper, I focus on the second-price all-pay auction (aka the war of attrition) but all of the central results of this paper could be replicated qualitatively with a first-price auction, however the equilibrium density functions and the range of bids would be different¹⁰.

In the symmetric two-player war of attrition, the setup is as follows: players a and b compete for a prize with a known common value of V , where V represents the incremental gain to fitness from winning the contest. Both players select non-negative bids x , and the prize is awarded to the player who chooses the higher bid. Both players pay an amount equal to the loser's bid. From player a 's perspective, the expected value of making his bid x_a against player b 's bid of x_b is given by:

⁹Nettle (2006) gives the example of the great tit. "Individuals of this species differ on a behavioral dimension called exploration, with high scorers being aggressive and bold in exploring the environment...in years of abundance, there is a strongly negative linear relationship between female survival and exploration score...For males, the pattern is diametrically opposite."

¹⁰There is also one specific advantage of the war of attrition over the first-price auction—only the war of attrition can give rise to Pyrrhic victories, since it is only in the war of attrition that players sometimes submit bids which exceed the value of the resource (we will see this shortly). When two players making such bids face one another, we end up with a contest where even the winners lose. Pyrrhic victories occur, so it seems better to have a model which can account for them.

$$E(x_a, x_b) = \begin{cases} V - x_b & \text{if } x_a > x_b \\ -x_a & \text{if } x_b > x_a \\ \frac{V}{2} - x_a & \text{if } x_a = x_b \end{cases}$$

The rationale behind this setup is that the contest escalates through the values from 0 upwards to the losing bid x_{low} (where $x_{low} \leq x_i \forall i$). The contest ends when the loser quits, so both players pay x_{low} —thus payoffs are related to the losing rather than the winning bid.

Originally, the currency for bids in the war of attrition was conceived to be time spent waiting in a contest of display or of patience, such as when male dung flies wait for females on a fresh cow pat (see Parker 1970*a,b*). The original assumption was of a linear relationship between time and cost. Later work on the war of attrition has revealed the game to be much more general, however. Maynard Smith (1982) shows that this game can be directly applied to any cost function which increases continuously, whether linearly or otherwise. The costs can be of many kinds, and are usually described as expenditure of energy.

The symmetric war of attrition has no pure strategy equilibrium. To see why not, notice that if there was an equilibrium bid value chosen by both players $x_a = x_b = \hat{x}$, then each player would earn $E(\hat{x}, \hat{x}) = \frac{V}{2} - \hat{x}$ from playing the game. This amount is positive so long as $\hat{x} < \frac{V}{2}$. But this cannot be an equilibrium, because each player has an incentive to deviate and increase his bid. For example, if player a were to bid $x_a = \hat{x} + \varepsilon$ (where ε may be small), then he would earn $E(x_a, \hat{x}) = V - \hat{x}$ which is larger than $\frac{V}{2} - \hat{x}$, and so he would deviate. But this cannot be an equilibrium either, because b would have an incentive to outbid a by a tiny amount $x_b = \hat{x} + 2\varepsilon$ to earn $E(x_b, x_a) = V - \hat{x} - \varepsilon$ (instead of getting $E(\hat{x}, x_a) = -\hat{x}$). And so on. It turns out that the game has a unique mixed strategy equilibrium where players choose their

bids x from a density function $p(x)$ given by a negative exponential distribution with mean V

$$p(x) = \frac{1}{V}e^{-\frac{x}{V}}$$

As stated previously, there are two possible interpretations for this mixed strategy equilibrium: monomorphic and polymorphic. In the monomorphic interpretation, all agents are the same and choose their bids randomly from the density function $p(x)$ each time they play the game. The polymorphic interpretation is that the players have different strategies which average strategy in the population is equal to $p(x)$. There are an infinity of ways to achieve this average, but the simplest is for an individual agent of type i to always makes the same bid x_i , and for the function $p(x)$ to give the relative frequency of each type in the population. This polymorphic result is more likely when there is a cost (e.g. a mental cost) to randomizing, and is consistent with the evidence of consistent behaviour over time which was discussed in the introduction.

Since participation times are exponentially distributed, a randomly chosen contestant has a constant probability $\frac{\delta t}{V}$ of dropping out of the contest during any interval of time $(t, t + \delta t)$, where δt is small. Since both participants' dropout probabilities are independently distributed, the chance that either one of them will drop out and end the game during any interval of time δt is just $\frac{2\delta t}{V}$. Thus the value of the lower bid ($x_{low} < x_{high}$), and thus the contest length, is also exponentially distributed, but with mean $\frac{V}{2}$

$$c(x) = \frac{2}{V}e^{-\frac{2x}{V}}$$

This shows that this game is not profitable on average. The expected cost of par-

participating in the contest is $\frac{V}{2}$ for each player, so the surplus is completely exhausted—a typical winner will win $+\frac{V}{2}$ and the typical loser will lose $-\frac{V}{2}$. The expected value of the game is zero.

To be more precise, the expected value of the game is equal to the lower of the two available prizes. Until now the prize was V for the winner and 0 for the loser. But if instead the winner received V_1 and the loser received V_2 , where $0 < V_2 < V_1$, then the incremental gains from participating in the contest would be $\Delta V = V_1 - V_2$, and the players' bids would be given by an exponential distribution with mean ΔV .

4.3.4 The basic model is a simple decision rule

We can now examine the simplest case of the complete model: the two player symmetric version. Recall that we are taking the simple polymorphic interpretation of this density function derived above, which means that, for a fixed prize value, individuals persistently make the same bid (and their children inherit these values), and the proportion of the population playing a particular bid x_i is given by the density function as $p(x_i)$. Notice, however, that the density of x depends on V , the fertility value of the contest at hand. Since V will vary over different contests, the polymorphic equilibrium cannot be the simple one in which a player i always chooses a bid x_i , because the equilibrium density at that bid must be a function of V . Players' behaviour has to be sensitive to the size of the prize at hand.

The equilibrium of this game will be one in which each player's behaviour is fixed in relation to the rest of the population, but flexible with respect to the available payoff. We can characterise the sense in which agents' relative behaviour is fixed by assigning each player i a characteristic private value θ_i from the simple exponential distribution $e^{-\theta}$

PROPOSITION A: *if players' types are distributed according to the pdf $e^{-\theta}$ and if*

each player i bids $\theta_i V$ in a contest for a resource of value V , then the population will be in equilibrium.

PROOF: It is straightforward to see that the conditions in the proposition hold, then the actual bids will correspond to the equilibrium pdf $p(x) = \frac{1}{V} e^{-\frac{x}{V}}$ given above. ■

Notice that this is a departure from standard evolutionary game theory; players usually have a fixed action but here they have a fixed decision rule. As noted, the reason for this departure is to allow the agents sufficient flexibility to respond to the changing prize values which they encounter in various contests while keeping their decision process as simple as possible. The model could of course be simplified further by assuming that V is constant and assigning agents a fixed action, but then the model would no longer be one of personality.

Naturally, this requires that θ is unobservable—if you knew your opponent's θ , (call it θ_j) then you would have an incentive to bid either $\theta_j V + \varepsilon$ (if $\theta_j < 1$) or 0 (if $\theta_j \geq 1$). The equilibrium holds only in the case where each opponent knows only that their enemy's type is taken from the density function $\psi(\theta)$. It is reasonable to assume that this information is private, though, because there are no incentives for truthful signalling—everybody would want to announce that they had $\theta > 1$, and so no one would be believed.

This basic model demonstrates the central argument of this paper. In social competitions of an all-pay nature, there is only a mixed strategy equilibrium. For a polymorphic population, this means that agents who are otherwise identical may persistently choose different effort levels to obtain an object of common value. Since "persistent variations in behaviour over time" is the definition of personality, the variation in strategies may be interpreted as a variation in personality. In this case, variations in θ seem very much like variations in extraversion, since both represent willingness to expend energy in social situations for the purpose of obtaining goods

of reproductive value.

4.3.5 Extraverts take on more risk

One of the central features of extraversion mentioned above is that extraverts exhibit greater taste for risk than introverts (Nettle 2005, 2006), and this is also a feature of the current model.

PROPOSITION B: *players with greater idiosyncratic values of θ (extraverts) take on more risk, and are not compensated for doing so*

PROOF: Since extraverts are characterized by higher values of θ , it is necessary to show that the variance of fertility is increasing in θ . Since θ maps linearly onto the bids x , we only need to show that the variance of fertility is increasing in x . Thus suppose we have an agent i who always chooses a bid $x_i = z$ when competing for a resource of value V . The expected payoff from playing z against a randomly chosen individual from a population which (in aggregate) follows the mixed strategy I defined by the density function $p(x)$ is

$$\xi(z, I) = \int_0^z (V - x)p(x)dx + \int_z^\infty (-z)p(x)dx$$

In equilibrium, when $p(x) = \frac{1}{V}e^{-\frac{x}{V}}$, the expected value of $\xi(z, I)$ is zero for all values of z . The variance of outcomes, however, is not equal for all values of z . The variance of outcomes is given by

$$\sigma^2(z, I) = E([\xi^2]) - E([\xi])^2 = E([\xi^2])$$

$$\sigma^2(z, I) = \int_0^z (V - x)^2 p(x)dx + \int_z^\infty z^2 p(x)dx$$

After some manipulation (see appendix one), it is possible to show that

$$\sigma^2(z, I) = V^2 P(z)$$

Where $P(z)$ is the cumulative distribution function given by

$$P(z) = \int_0^z p(x)dx = 1 - e^{-\frac{z}{V}}$$

Since $P(z)$ is strictly increasing in z , the function $\sigma^2(z, I)$ is also strictly increasing in z . Another way to show this is to take the derivative of $\sigma^2(z, I)$ with respect to z , and note that it is always positive

$$\frac{\partial \sigma^2(z, I)}{\partial z} = V^2 p(z) > 0$$

Thus extraverts who choose higher bids are taking on extra risk compared to introverts, and they are doing so without compensation, since all payoffs are equal in equilibrium. ■

It is worth noting that players in this game are not risk loving or risk averse in the traditional sense of having convex or concave utility functions. In any evolutionary model in an infinite population, all players must be risk neutral with regard to fertility (as long as they do not face the risk of total extinction, which they do not in an infinite population). Thus all utility functions are linear in fertility. Furthermore, since the relationship between resources and fertility is (in this model) linear, utility is linear in resources as well. Thus, formally, all players have the same attitude towards risk. But if some strategies deliver more variance than others, this means that some agents (extraverts) must take uncompensated fertility risks. These people will be observationally risk loving (or less risk averse) and perhaps strongly so if the relationship between resources and fertility is concave rather than linear as I

have assumed (risk aversion in the real world is of course generally measured against resources rather than fertility).

4.3.6 Asymmetry

So far we have seen that certain kinds of games will generate variation in strategies (aka personalities) even when there is neither variation in ability nor variation in the players' valuations. Clearly this is an oversimplification, but here we will see that asymmetries can make the case even stronger.

There are several versions of the asymmetric war of attrition. They all have some things in common—they all deal with the two player case where it is assumed that only two possible types A & B exist. A is considered to be the favoured type and usually wins the game. Beyond this basic setup, though, the various assumptions differ notably. We will follow the treatment of Hammerstein and Parker (1982). Hammerstein and Parker's model has a unique mixed strategy equilibrium with a continuous strategy space, unlike alternative models such as Nalebuff and Riley's (1985) model of private information with a continuum of asymmetric equilibria, or Kim's (1993) version of the game which has only pure strategy equilibria and which requires making the war of attrition discreet rather than continuous and which also introduces trembles in the implementation of strategies.

Hammerstein and Parker's model—henceforth simply called the asymmetric model—assumes the existence of relevant asymmetries, so that A -types have either "more to gain or less to pay for persistence" than B -types. The asymmetry *must* be related to the payoffs. The authors note that "the continuous war of attrition with a payoff irrelevant (uncorrelated) asymmetry is a pathological structurally unstable model, since arbitrarily small changes in the parameters would lead to fundamentally different solutions".

With two players and two possible types, there are only four contest situations: (A, A) , (A, B) , (B, A) & (B, B) . The probability that any randomly chosen contest will find player 1 in role A and player 2 in role B is denoted w_{AB} . The other three probabilities w_{AA} , w_{BA} & w_{BB} are defined similarly. Because the "player 1" and "player 2" labels are assigned by chance, it must be that $w_{AB} = w_{BA}$. The model assumes that there is a nonzero probability that both contestants have the same role, i.e. $w_{AA} \neq 0$ and $w_{BB} \neq 0$. These cases of unexpected symmetries arise due to occasional errors in perception, meaning that the players have incomplete information about their opponent's type. Although such errors may be infrequent, they must occur with positive probability or else the game has no ESS.

This model allows for real asymmetries either in the cost of participation or in the value of the prize. Hammerstein and Parker note that both types of asymmetry give similar results. Thus, rather than exploring the closely related outcomes from varying the value of the resource V_{AA} , V_{AB} , V_{BA} , and V_{BB} in the four different cases, or the cost of participation C_{AA} , C_{AB} , C_{BA} , and C_{BB} , I will introduce only the smallest of asymmetries and assume the costs of participation are still the same for both types, but that the resource is worth $V_A = V + \varepsilon$ to the favoured A -types (where ε is small and positive) and is worth $V_B = V - \varepsilon$ to the disfavoured B -types. Furthermore, the value of the resource is independent of who the game is played against (i.e. $V_A = V_{AA} = V_{AB}$, and $V_B = V_{BA} = V_{BB}$). It is worth noting that the following results will hold qualitatively even when ε is large, but I will examine the equilibrium in the limit as $\varepsilon \rightarrow 0$ because it provides the neatest solution.

A straightforward interpretation of roles A and B is that of possessor and interloper (or "rightful owner" and "usurper"). A -types represent those who already possess the disputed territory, mate, or whatever is at stake, while the B -type is challenging his possession (and this is common knowledge but with some probability

of mistaken identity as noted above). But this point is not central to the analysis. All that we require is that the value of the prize is slightly higher for A than for B . This might happen because of greater familiarity (the current owner does not have to waste time learning the properties of the resource) or for any other reason. In the simpler hawk-dove-bourgeois game, the owner-interloper split allows the relatively efficient "bourgeois" solution to be an ESS of the game (Maynard Smith 1979)¹¹. The bourgeois strategy was to choose hawk when playing as the owner of the disputed resource, and dove otherwise. Thus when all agents played bourgeois, the wasteful combat was eliminated. The asymmetric game outlined here can be likened to a continuous version of the hawk-dove-bourgeois game.

The general solution to the asymmetric model derived by Hammerstein and Parker is one where the A -types choose their bids from an exponential distribution with a lower limit of s (i.e. $x_A \geq s$), while the B -types choose their bids from an exponential distribution with an upper limit of s (i.e. $x_B < s$). This means that in the (A, B) and (B, A) contest situations, the A -type will always submit a higher bid and claim the prize, while in the (A, A) and (B, B) contest situations, each player has a chance of winning. Of course, victory will be more costly in the (A, A) case, and is quite likely to be Pyrrhic. The general solutions for $p_A(x)$, $p_B(x)$, and s are in the footnote¹², but by assuming more symmetry than the original model requires, we can produce something tidier.

If we define

¹¹Bourgeois was the only ESS in the simple version of that game, but more complicated versions have other ESSs. See Mesterton-Gibbons (1992) for some variants on hawk-dove-bourgeois.

¹²The boundary s is given by: $s = -\frac{V_{BB}}{C_{BB}} \ln \left(\frac{w_{BA}C_{BA}}{w_{BA}C_{BA} + w_{BB}C_{BB}} \right)$.

The pdf for A -types' bids is $p_A(x) = \frac{C_{AA}}{V_{AA}} e^{-\frac{C_{AA}}{V_{AA}}(x-s)}$ if $x \geq s$ and 0 otherwise. B 's pdf is given by $p_B(x) = \frac{1}{V_{BB}} \left(\frac{w_{BA}C_{BA}}{w_{BB}} + C_{BB} \right) e^{-\frac{C_{BB}}{V_{BB}}x}$ if $0 \leq x < s$ and 0 otherwise.

$$W = \frac{w_{BA}}{w_{BA} + w_{BB}}$$

So that W is the probability that a randomly chosen B -type is *actually* competing against an A -type (remember that there are occasional mistakes in perception).

And if we take the value of q in the limit as $\varepsilon \rightarrow 0$, we can write

$$s = V \ln \frac{1}{W}$$

(Or, equivalently, $s = -V \ln W$). This allows the equilibrium density function for bids in the asymmetric game to be written as

$$p_A(x, W) = \begin{cases} 0 & \text{if } 0 \leq x < V \ln \frac{1}{W} \\ \left(\frac{1}{W}\right) \frac{1}{V} e^{-\frac{x}{V}} & \text{if } V \ln \frac{1}{W} \leq x \end{cases}$$

$$p_B(x, W) = \begin{cases} \left(\frac{1}{1-W}\right) \frac{1}{V} e^{-\frac{x}{V}} & \text{if } 0 \leq x < V \ln \frac{1}{W} \\ 0 & \text{if } V \ln \frac{1}{W} \leq x \end{cases}$$

So far, the asymmetric equilibrium I have presented is just a simplified version of the one derived by Hammerstein and Parker. It is worth briefly examining these equations to get some intuition from them. Consider what happens as W varies. As $W \rightarrow 1$ (i.e. as the probability w_{BB} of a B -type facing another B -type goes to zero), the boundary $s \rightarrow 0$, which means that B -types place increasing weight on values of x close to zero, and A -types behave more and more like players in the symmetric 2 player war of attrition, producing bids which correspond in the limit to the density function $p_A(x) = \frac{1}{V} e^{-\frac{x}{V}}$ with bounds $(0, \infty)$. The other extreme, where $W \rightarrow 0$ (i.e. as $w_{BB} \rightarrow 1$), means that the boundary $s \rightarrow \infty$, and it is the B -types who come to behave increasingly like players in the symmetric 2 player war of attrition (which

makes sense, because most of their fights are against other B s).

Now we will re-express this equilibrium in terms of the personality function and bidding rule described in the previous section. As promised in the introduction, the asymmetric case can be described with the same functions as the symmetric case, but there is an intermediate step as well. The intermediate step involves mapping the person's idiosyncratic type θ_i onto a pair of values θ_i^A and θ_i^B each of which can then become inputs into the bidding rule, depending on which role player i finds himself in.

We begin with the familiar density function for personality types

$$\psi(\theta) = e^{-\theta} \quad \theta \in [0, \infty)$$

What is needed is a rule for translating θ_i into θ_i^A and θ_i^B such that these latter variables are characterized by the density functions

$$d_A(\theta^A, W) = \begin{cases} 0 & \text{if } 0 \leq \ln \frac{1}{W} < \theta^A \\ \frac{1}{W} e^{-\theta^A} & \text{if } \theta^A \geq \ln \frac{1}{W} \end{cases}$$

and

$$d_B(\theta^B, W) = \begin{cases} \frac{1}{1-W} e^{-\theta^B} & \text{if } 0 \leq \ln \frac{1}{W} < \theta^B \\ 0 & \text{if } \theta^B \geq \ln \frac{1}{W} \end{cases}$$

It is important to end up with these particular functions because they give densities for θ^A and θ^B such that agents can use the same bidding rule given in the previous section

$$x_i^J(\theta_i, \Delta V) = \theta_i^J \Delta V \quad J \in \{A, B\}$$

and they will produce bids at the equilibrium densities given by $p_A(x, W)$ and $p_B(x, W)$ above. As for the actual mapping from θ to θ^J , it can be shown (see appendix) that the function

$$F_A(\theta_i, W) = \theta_i + \ln \frac{1}{W} = \theta_i^A$$

performs the mapping for the A -types, while the function

$$F_B(\theta_i, W) = -\ln [W + (1 - W)e^{-\theta_i}] = \theta_i^B$$

does the requisite plotting for B -types.

Both F_A and F_B as well as their resulting distributions d_A and d_B preserve player i 's position relative to the population. This means that players' behaviors across a range of very different situations will show very similar patterns. Logically, this is not necessary. The system just described would also be stable if each player were characterized by two distinct θ 's – one for use in the favoured position, and one for use in the disfavoured position. As long as each of the two values of θ was taken independently from the density function $e^{-\theta}$, then the system would be in equilibrium. But a system with two different values of θ is redundant and arguably less realistic, since people's behaviours are in fact correlated across a range of situations. Furthermore, there is neurological evidence that the same brain regions are involved in the different situations (discussed in the next section), which suggests that the more parsimonious version is the better one.

So what are the implications of all this? The basic story from the previous section still holds: people must play different values of x in order for ESS conditions to prevail, and the value of x chosen by a player depends on the value of θ assigned to him by Nature. To review: agents playing the asymmetric war of attrition are

characterized by an idiosyncratic value θ_i from the distribution $e^{-\theta}$ —this is the same value which was used to represent their type in the symmetric game. The value θ_i is either amplified to θ_i^A if player i is acting as the favoured type, or else diminished to θ_i^B if the player is acting as the disfavoured type. Finally, the appropriate value of θ_i^J is used as an input to the same rule as before, so θ_i^J times ΔV gives the bid. Things are slightly more complicated in the asymmetric case, of course, because of the different values of θ_i^A and θ_i^B . But this additional complexity allows the model to address other issues—namely the existence of loss aversion and the observed variation in intensities of loss aversion.

4.3.7 Asymmetries and loss aversion

PROPOSITION C: *there is a U-shaped relationship between extraversion and loss aversion*

Consider the 'owner & interloper' interpretation of the roles A & B . In equilibrium, owners will always submit higher bids than interlopers. I will argue that this is best understood as a *strategic endowment effect*. Recall that the endowment effect is the phenomenon whereby people value an object more highly once their property right to it has been established. Experiments have repeatedly shown that the minimum price that a typical owner is willing to accept for an object will generally be much higher than the price that a typical buyer is willing to pay for the same object, even when the roles of "buyer" and "owner" are assigned by chance and the ownership rights are only a few minutes old (see Kahneman, Knetsch, & Thaler 1990 for a summary of several such experiments). The endowment effect is usually assumed to arise from *loss aversion*, which causes losses to loom larger than equivalent gains. The effect was first discussed in the context of prospect theory (Kahneman & Tversky 1979).

While loss aversion has been extensively documented empirically, there has been little theoretical work to explain how the trait evolved. The asymmetric model presented in this section provides one explanation for the origin of loss aversion—that it may have arisen from strategic interactions where ownership cues were used to settle the contest (hence *strategic* loss aversion). The basic idea is that owners feel attached to their possessions because this feeling causes them to try harder to keep possession than interlopers will try to take possession from them. When owners expend more effort than interlopers, we have the essence of the asymmetric equilibrium described above. We also have a textbook case of loss aversion.

Gintis (2007) also argues that loss aversion evolved because our ancestors played a game similar to hawk-dove-bourgeois. He notes that bourgeois-esque strategies in which asymmetries of ownership are used to settle contests can be an efficient way to avoid wasteful combat in the absence of enforced property rights, and hence groups who use such cues are likely to be favoured compared to those engaging in endless conflict. Indeed, as Gintis notes, this interpretation accounts for the fact that some form of property rules characterise not only human societies but many animal groups as well.

The model which Gintis uses to make his point is quite different from the one presented in this paper, however, even though both models share a family resemblance to hawk-dove-bourgeois. Gintis' model, for instance, is a probabilistic model unrelated to Hammerstein and Parker's asymmetric war of attrition. In the Gintis model, there is a possibility of getting the counter intuitive "anti-property" equilibria where the interlopers are favoured over the owners, even if the owners are slightly favoured with respect to payoffs. This kind of result cannot arise in the current model since *A*-types are defined by the fact that they are favoured with respect to payoffs or participation costs, and any equilibrium will always result in *As* submitting higher

bids than B s (Hammerstein & Parker 1982)¹³.

There is also another feature of loss aversion which was not addressed by Gintis, and which has, as far as I know, never been explained—the variation in the extent to which people are loss averse. The degree of loss aversion is typically quantified by a parameter λ , which measures the relative slopes of the value function for losses and for gains. For a large sample, the average value of λ is typically somewhere around 2 (Camerer, 2003), implying that the experience of losing £1 is approximately twice as salient as the experience of gaining £1. However most empirical studies find that the subjects exhibit a wide range of λ s—Tom et al. (2007), for example, report values ranging from 0.99 to 6.75. This kind of variation arises naturally from the current model.

To see how, recall that in the asymmetric case a player i is characterized by two values θ_i^A and θ_i^B which then become inputs to the bidding rule. But the bidding rule can be interpreted as a utility function—we bid more when we crave more intensely. Since the index of loss aversion λ is generally defined (e.g. in Tversky & Kahneman 1992) as the *ratio* of the value (i.e. utility) of gaining \$1 to the value of losing \$1, $\lambda_i = \frac{-U_i(-\$1)}{U_i(\$1)}$ (see Köbberling & Wakker 2005 for variants on this definition) then it is clear that in our case, we can express λ_i as the ratio $\frac{\theta_i^B}{\theta_i^A}$ since the ΔV term will cancel out of the utility function. Those players for whom the ratio $\frac{\theta_i^A}{\theta_i^B}$ is very high will be highly loss averse, and vice versa. The point to note is that the ratio is not the same for all types. Mathematically, λ_i is given by

$$\lambda_i = \frac{\theta_i^A}{\theta_i^B} = \frac{\theta_i - \ln W}{-\ln[W + (1 - W)e^{-\theta_i}]}$$

¹³Note that both Gintis' and my own explanations are meant to account only for the relative *slopes* of the value function (i.e. steeper below the horizontal axis than above it). The *shape* of the value function (i.e. concave in the positive domain and convex in the negative) is likely to be accounted for by other considerations, like diminishing marginal productivity.

To see how λ varies with θ , we need to check the sign on the first derivative

$$\frac{d\lambda_i}{d\theta_i} = \frac{1}{(\ln[W + (1 - W)e^{-\theta_i}])} \left(\frac{\ln W - \theta_i}{(\ln[W + (1 - W)e^{-\theta_i}])} \cdot \frac{(1 - W)e^{-\theta_i}}{W + (1 - W)e^{-\theta_i}} - 1 \right)$$

This expression is negative for small values of θ and then becomes positive again for larger values of θ , which means that the relationship between λ and θ is roughly U -shaped in θ .

The shape of this function follows naturally from the definition of $\lambda = \frac{\theta^A}{\theta^B}$. Those players with θ near zero will have θ^B near zero as well, while θ^A will be near s (remember that s is the lower bound of θ^A and the upper bound of θ^B) and so the limit of λ as θ goes to zero is ∞ , implying that extreme introverts are extremely loss averse as well. The ratio λ will then fall in θ , implying that increasing extraversion is associated with decreasing loss aversion. At some point, however, the U bottoms out and the relationship reverses itself. When θ tends to ∞ , θ^B rises with a limit of q while θ^A rises without bound, so the ratio $\frac{\theta^A}{\theta^B}$ tends to ∞ , implying that extreme extraverts are also extremely loss averse. The exact point at which the U inflects depends on W , and would also depend on V_A , V_B , C_A , and C_B if I had included those variables. Thus it is not possible to make a clear prediction about how λ and θ are related without knowing more about the types of asymmetries faced by our ancestors. It is possible, however, to derive two implications about λ : that it will not be constant across the population, and that the most extreme extraverts and introverts in the population will have the highest values of λ and be the most loss averse. The first of these predictions has been demonstrated, as I already noted. The second has not, to my knowledge, been tested.

The idea that this kind of a game lies at the root of loss aversion is consistent

with the neurological evidence as discussed by Tom et al. (2007), who subjected patients to functional magnetic resonance imaging (fMRI) while exposing them to 50/50 win/loss gambles of varying size. The authors find "...increasing activity for gains and decreasing activity for losses [which] demonstrated joint sensitivity to both gains and losses in a set of regions, including the dorsal and ventral striatum and VMPFC¹⁴". This study's finding that losses and gains are coded for by the same regions is in contrast to earlier suggestions that the gain and loss aspects were governed separately (e.g. Camerer, 2005). Also notable among Tom et al.'s findings is that areas of activation were "observed throughout, though not strictly limited to, the targets of the mesolimbic and mesocortical dopamine (DA) systems". This is notable because the dopamine systems exhibit considerable genetic polymorphism across humans (Cravchik & Goldman 2000) and they are thought to be linked to extraversion (Depue & Collins 1999). However, it is obviously difficult or impossible to map from theoretical models such as this one to make predictions about neurological structure—I offer this evidence for consideration merely because it provides an interesting fit with the theory.

4.3.8 Adding more players

The results for the two player symmetric war of attrition carry over with almost no alteration to n players¹⁵. The n player war of attrition was first described by Haigh & Cannings (1989)¹⁶, and the basic setup is similar to the two player case: there are n players competing for n prizes ($V_n < V_{n-1} < \dots < V_1$) which become available in

¹⁴Ventromedial prefrontal cortex.

¹⁵As far as I know, there is no asymmetric war of attrition for more than two players. Maynard Smith (1982) makes the case that these are unlikely to arise in nature.

¹⁶Bulow and Klemperer (1999) generalized the n -player war of attrition to a "standards" situation in which players must continue to pay after having dropped out of the game (as in the case where an industry is fighting over a technical standard).

ascending order of value. At the beginning of the game, all players select some cost that they are willing to bear; the player with the lowest willingness to pay drops out first, collects the prize V_n and leaves the game. The remaining players then repeat this process through a further $n - 2$ rounds until, eventually, the second last player drops out with reward V_2 , leaving the winner to immediately collect V_1 ¹⁷.

The game has a unique mixed strategy ESS where the players choose their bids for each round from the exponential function with mean $(\hat{n})(\Delta V_{\hat{n}})$, where \hat{n} corresponds to the number of *opponents remaining*¹⁸ in the game during any given round, and $\Delta V_{\hat{n}} = V_{\hat{n}} - V_{\hat{n}+1}$ (so $\Delta \hat{V}$ need not be the same in each round)

$$p_{\hat{n}}(x) = \frac{1}{(\hat{n})(\Delta V_{\hat{n}})} e^{-\frac{x}{(\hat{n})(\Delta V_{\hat{n}})}}$$

Notice that when there is only one opponent left, the function $p_{\hat{n}}(x)$ simplifies to the equilibrium function $p(x)$ described in section 4.3.4 for the symmetric two player war of attrition.

Once again, however, we would like to see a bidding rule given personality type. The fixed bidding rule for the whole population will be

$$x_i(\hat{n}, \theta_i, \Delta V) = \hat{n} \cdot \theta_i \cdot \Delta V$$

Recall that the rule was formerly written as

$$x_i(\theta_i, \Delta V) = \theta_i \cdot \Delta V$$

¹⁷The setup I describe corresponds to model *C* in Haigh & Cannings' paper. They also consider n player games in which there is only one prize and consequently no ESS (this is model *A*) and n player games in which the n prizes which become available in pre-determined but non-ascending order (this is model *B*, and it does produce an ESS). I have restricted myself to model *C* because it is simpler, but qualitatively similar results can be obtained from model *B*.

¹⁸The number of opponents remaining will always be equal to the number of rounds remaining just before the current round begins. Thus \hat{n} begins with a value of $(n - 1)$, and falls by one per round so that at the beginning of the final round, $\hat{n} = 1$.

Clearly, the former rule is just a special case when the number of remaining opponents is one.

Given this bidding rule, the distribution of personality types within the population will be given by the pdf

$$\psi(\theta) = e^{-\theta} \quad \theta \in [0, \infty)$$

Once again, this is the same distribution as we had in the two player case, which means that the inferences drawn from the basic model will continue to hold.

4.4 Conclusion

In summary, I have argued that something like a war of attrition must have characterised at least some of the conflicts which our ancestors entered into and that these conflicts would induce the evolution of something like the extraversion continuum. It was shown that the simple decision rule which motivates behaviour in the symmetric two-player case is robust to the introduction of asymmetries of ability in two-player contests as well as to n-player contests. Furthermore, I have argued that this explanation accounts for the fact that extraverts seem to take more risk than introverts, and that it predicts a U-shaped relationship between extraversion and loss aversion.

The interpretation throughout this paper is that the behavioural variation necessary to obtain equilibrium in the war of attrition is coded into the population via differences in subjective reward sensitivity. As outlined above, this fits with the Lucas et al. (2000) finding that reward sensitivity lies at the core of differences in extraversion. But it is still very much a conjecture at this point to say that wars of attrition led to the variance in reward sensitivity; it would be very interesting to see data (e.g. from fMRI) which related reward sensitivity to behaviour in all-pay

contests, but I am not aware of any such study.

The finding that extraversion should be related to risk aversion warrants a cautious interpretation. As mentioned above, Nettle (2005) found that extraverts do exhibit a greater taste for physically risky activities (like skydiving) and that they have a correspondingly higher rate of hospitalisation. And while many economists like to imagine that people can be well characterised by a single risk aversion parameter, in real world risk attitudes appear to be more complicated: Almlund et al. (in press) review evidence that risk aversion has been linked to sensation seeking (a facet of extraversion), openness to experience, neuroticism, agreeableness, and conscientiousness (i.e. all five of the Big Five), and Anderson and Mellor (2009) showed that risk attitudes are not stable across elicitation methods. If risk attitudes are indeed multi-faceted then it will prove difficult or impossible to produce a single theory of their evolution, and the current paper must be interpreted as an attempt to explain that part of risk aversion which relates primarily to behaviour in status and resource contests.

The finding that extraversion should be related to loss aversion is, as noted above, untested (to my knowledge). But given the rapid proliferation of personality and economics experiments documented in Almlund et al. (in press) and in several of James Heckman's other writings, it is likely that such a test will be conducted shortly, if indeed it is not already in press.

Appendix

A1. Calculating the variance of payoffs

These calculations show how the variance of outcomes changes with the choice of bid in a polymorphic population playing the two player war of attrition. First, note that the equilibrium pdf for the war of attrition is given by the probability density function $p(\cdot)$. The integrand of $p(\cdot)$ is also defined below.

$$p(z) = \frac{1}{V} e^{-\frac{z}{V}}$$

$$P(z) = \int_0^z p(x) dx = 1 - e^{-\frac{z}{V}}$$

The expected payoff from playing any bid z against an opponent who plays the strategy I defined by $p(\cdot)$ is given by

$$\xi(z, I) = \int_0^z (V - x)p(x) dx + \int_z^\infty (-z)p(x) dx$$

Given $p(x)$ above, $\xi(z, I)$ takes a value of 0 for all z . Thus the variance of outcomes is

$$\sigma^2(z, I) = E([\xi^2]) - E([\xi])^2 = E([\xi^2])$$

$$\sigma^2(z, I) = \int_0^z (V - x)^2 p(x) dx + \int_z^\infty z^2 p(x) dx$$

$$\sigma^2(z, I) = \int_0^z (V^2 - 2Vx + x^2)p(x) dx + z^2 \int_z^\infty p(x) dx$$

$$\sigma^2(z, I) = V^2 \int_0^z p(x)dx - 2V \int_0^z xp(x)dx + \int_0^z x^2p(x)dx + z^2 \int_z^\infty p(x)dx$$

$$\sigma^2(z, I) = V^2P(z) - 2V \int_0^z xp(x)dx + \int_0^z x^2p(x)dx + z^2(1 - P(z))$$

(See below for IBP#1)

$$\sigma^2(z, I) = V^2P(z) - 2V(VP(z) - z[1 - P(z)]) + \int_0^z x^2p(x)dx + z^2(1 - P(z))$$

$$\sigma^2(z, I) = V^2P(z) - 2V^2P(z) + 2Vz(1 - P(z)) + \int_0^z x^2p(x)dx + z^2(1 - P(z))$$

$$\sigma^2(z, I) = -V^2P(z) + 2Vz(1 - P(z)) + \int_0^z x^2p(x)dx + z^2(1 - P(z))$$

(See below for IBP#2)

$$\sigma^2(z, I) = -V^2P(z) + 2Vz(1 - P(z)) + [-z^2(1 - P(z)) + 2VzP(z) + 2V^2P(z) - 2Vz] + z^2(1 - P(z))$$

$$\sigma^2(z, I) = -V^2P(z) + 2Vz + 2V^2P(z) - 2Vz$$

$$\sigma^2(z, I) = V^2P(z)$$

Since $P(z)$ is strictly increasing in z , the function $\sigma^2(z, I)$ is also increasing in z . This can also be shown by taking the derivative of $\sigma^2(z, I)$ with respect to z :

$$\frac{\delta\sigma^2(z, I)}{\delta z} = V^2 p(z) > 0$$

Integration by parts #1

$$\int_0^z x \cdot p(x) dx = [xP(x) - \int P(x)]_0^z$$

$$\int_0^z x \cdot p(x) dx = zP(z) - [z + Ve^{-\frac{z}{V}} - V]$$

$$\int_0^z x \cdot p(x) dx = VP(z) - z(1 - P(z))$$

Integration by parts #2

$$\int_0^z x^2 p(x) dx = \left[x^2 P(x) - 2 \int x P(x) dx \right]_0^z$$

$$\int_0^z x^2 p(x) dx = \left[x^2 P(x) - 2 \left(\frac{x^2}{2} - xVP(x) - V^2 P(x) + xV \right) \right]_0^z$$

$$\int_0^z x^2 p(x) dx = [x^2 P(x) - x^2 + 2VzP(x) + 2V^2 P(x) - 2xV]_0^z$$

$$\int_0^z x^2 p(x) dx = z^2 P(z) - z^2 + 2VzP(z) + 2V^2 P(z) - 2Vz$$

$$\int_0^z x^2 p(x) dx = -z^2(1 - P(z)) + 2VzP(z) + 2V^2 P(z) - 2Vz$$

A2. The functions F_A and F_B

Given the density function

$$\psi(\theta) = e^{-\theta} \quad \theta \in [0, \infty)$$

the corresponding cumulative distribution function is

$$D_\psi(z) = \int_0^z \psi(\theta) d\theta = 1 - e^{-z}$$

Similarly, for

$$d_A(\theta^A) = \frac{1}{W} e^{-\theta^A} \quad \theta^A \in [q, \infty)$$

the cdf is

$$D_A(a) = \int_q^a d_A(\theta^A) d\theta^A = 1 - \frac{1}{W} e^{-a}$$

And finally, for

$$d_B(\theta^B) = \frac{1}{1-W} e^{-\theta^B} \quad \theta^B \in [0, q)$$

the cumulative distribution is given by

$$D_B(b) = \int_0^b d_B(\theta^B) d\theta^B = \frac{1}{1-W} (1 - e^{-b})$$

What we want is a function $F_A(z) = a$ such that $D_\psi(z) = D_A(a)$, and another function $F_B(z) = b$ such that $D_\psi(z) = D_B(b)$. These functions F_A and F_B will permit an agent with $\theta_i = z$ to maintain his relative place in the distribution when he sets $\theta_i^A = a$ and $\theta_i^B = b$. Beginning with the A -types

$$1 - e^{-z} = 1 - \frac{1}{W}e^{-a}$$

$$e^{-z} = \frac{1}{W}e^{-a}$$

$$-z = \ln \frac{1}{W} - a$$

$$a = z + \ln \frac{1}{W}$$

$$F_A(\theta_i, W) = \theta_i + \ln \frac{1}{W} = \theta_i^A$$

And likewise for the B -types

$$1 - e^{-z} = \frac{1}{1 - W}(1 - e^{-b})$$

$$(1 - W)(1 - e^{-z}) = (1 - e^{-b})$$

$$e^{-b} = 1 - (1 - W)(1 - e^{-z})$$

$$e^{-b} = 1 - (1 - W - e^{-z} + We^{-z})$$

$$e^{-b} = W + (1 - W)e^{-z}$$

$$-b = \ln [W + (1 - W)e^{-z}]$$

$$b = -\ln [W + (1 - W)e^{-z}]$$

$$F_B(\theta_i, W) = -\ln [W + (1 - W)e^{-\theta_i}] = \theta_i^B$$

Bibliography

- [1] Almlund, M., Duckworth, A., Heckman, J., & Kautz, T. (in press). Personality psychology and economics. In E.A. Hanushek, S. Machin & L. Wößmann (Eds.), *Handbook of the economics of education*. Amsterdam: Elsevier.
- [2] Aluja, A., García, Ó., & García, L. F. (2003). Relationships among extraversion, openness to experience, and sensation seeking. *Personality and Individual Differences*, 35(3), 671-680. doi:16/S0191-8869(02)00244-1
- [3] Anderson, L. R., & Mellor, J. M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, 39(2), 137-160. doi:10.1007/s11166-009-9075-z
- [4] Barrick, M. R. and Mount, M. K. (1991). "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis." *Personnel Psychology* 44(1): 1-26.
- [5] Bulow, J., & Klemperer, P. (1999). "The Generalized War of Attrition." *The American Economic Review*, 89(1), 175-189.

- [6] Caliendo, Marco, Fossen, Frank and Kritikos, Alexander (2008). "The Impact of Risk Attitudes on Entrepreneurial Survival." *IZA Discussion Paper* No. 3525.
- [7] Camerer, C. (2003). Behavioral Game Theory: experiments in strategic interaction. Princeton University Press, Princeton, NJ.
- [8] Camerer, C. (2005). "Three cheers—psychological, theoretical, empirical—for loss aversion". *Journal of Marketing Research*, 42, 129-133.
- [9] Cravchik, A. & Goldman, D. (2000). "Neurochemical individuality—Genetic diversity among human dopamine and serotonin receptors and transporters". *Archives of General Psychiatry*, 57, 1105-1114.
- [10] Depue, R. A. & Collins, P. F. (1999). "Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion". *Behavioral and Brain Sciences*, 22, 491-533.
- [11] Gintis, H. (2007). "The evolution of private property". *Journal of Economic Behavior & Organization*, 64, 1-16.
- [12] Goldberg, L.R., Sweeney, D., Merenda, P. F. and Hughes, J. E., Jr. (1998). "Demographic Variables and Personality: The Effects of Gender, Age, Education, and Ethnic/Racial Status on Self-Descriptions of Personality Attributes." *Personality and Individual Differences* 24(3): 393-403.
- [13] Haigh, J., & Cannings, C. (1989). "The n-person war-of-attrition". *Acta Applicandae Mathematicae* 14, 59–74.

- [14] Ham, R., Junankar, P.N. and Wells, R. (2009). "Occupational Choice: Personality Matters." *IZA Discussion Paper* No. 4105.
- [15] Hammerstein, P. & Parker, G. A. (1982). "The asymmetric war of attrition". *Journal of Theoretical Biology*, 96, 647–682.
- [16] Heckman, James J., Stixrud, Jora and Urzua, Sergio (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24(3): 411-482.
- [17] John, O. P., Caspi, A., Robins, R. W. and Moffitt, T. E. (1994). "The "Little Five": Exploring the Nomological Network of the Five-Factor Model of Personality in Adolescent Boys." *Child Development* 65(1): 160-178.
- [18] Just, W. & Zhu, F. (2004). "Individual-based simulations of the War of Attrition". *Behavioural Processes*, 66, 53-62.
- [19] Kahneman, D., Knetsch, J., & Thaler, R. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem". *Journal of Political Economy*, 98, 1325-1348.
- [20] Kahneman, D., & Tversky, A. (1979). "Prospect Theory". *Econometrica*, 47, 263-292.
- [21] Kanazawa, S. (2011). Evolutionary Psychology and Individual Differences. *The Wiley-Blackwell Handbook of Individual Differences*, 353 - 378.

- [22] Kim, Y. (1993). "Evolutionary stability in the asymmetric war of attrition". *Journal of Theoretical Biology*, 161, 13-21.
- [23] Köbberling, V. & Wakker, P. (2005). "An index of loss aversion". *Journal of Economic Theory*, 122, 119 – 131.
- [24] Lucas, R. E., Diener, E., Grob, A., Suh, M. E., & Shao, L. (2000). "Cross-cultural evidence for the fundamental features of extraversion". *Journal of Personality and Social Psychology*, 79, 1039–1056.
- [25] Matthews, G., Deary, I. J., & Whiteman, C. (2009). *Personality traits* (3rd ed). Cambridge: Cambridge University Press.
- [26] Maynard Smith, J. (1974). "The theory of games and the evolution of animal conflicts". *Journal of Theoretical Biology*, 47, 209-21.
- [27] Maynard Smith, J. (1979). "Game theory and the evolution of behaviour". *Proceedings of the Royal Society of London (B)*, 205, 475-488.
- [28] Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- [29] McAdam, D. (1995). "What do we know when we know a person?" *Journal of Personality*, 63, 365-396.
- [30] Nalebuff, B., & Riley, J. (1985). "Asymmetric equilibria in the war of attrition". *Journal of Theoretical Biology*, 113, 517-527.
- [31] Nettle, D. (2005). "An evolutionary approach to the extraversion continuum". *Evolution and Human Behavior*, 26, 363-373.
- [32] Nettle, D. (2006). "The evolution of personality variation in humans and other animals". *American Psychologist*, 61, 622-631.

- [33] Nowak, M. A. (2006). *Evolutionary Dynamics*. Cambridge/London: Belknap/Harvard.
- [34] Parker, G. (1970a). "The reproductive behaviour and the nature of sexual selection in *Scatophaga stercoraria* L. II. The fertilization rate and the spatial and temporal relationships of each sex around the site of mating and oviposition". *Journal of Animal Ecology*, 39, 205-228.
- [35] Parker, G. (1970b). "The reproductive behaviour and the nature of sexual selection in *Scatophaga stercoraria* L. IV. Epigamic recognition and competition between males for the possession of females". *Behaviour*, 37, 113-139.
- [36] Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality (target article). *European Journal of Personality*, 21, 549-587.
- [37] Powell, M. & Ansic, D. (1997). "Gender differences in risk behaviour in financial decision-making: An experimental analysis". *Journal of Economic Psychology*, 18, 605-628.
- [38] Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A. and Goldberg, L. R. (2007). "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes." *Perspectives on Psychological Science* 2(4): 313-345.
- [39] Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, 315(5811), 515 -518.

- [40] Tversky, A. & Kahneman, D. (1992). "Advances in prospect theory: cumulative representation of uncertainty". *Journal of Risk and Uncertainty*, 5, 297–323.
- [41] van Eijck, Koen and de Graaf, Paul M. (2004). "The Big Five at School: The Impact of Personality on Educational Attainment." *Netherlands' Journal of Social Sciences* 40(1): 24-40.

General Appendix

A brief guide to personality theory

Table 4.1: Short guide to the Five Factor Model

Domain	Brief description
Neuroticism	suceptibility to fear, anxiety and other negative emotions
Extraversion	energy, positive emotions and sociability
Openness	appreciation for art, emotion and left-leaning politics
Agreeableness	tendency to be compassionate, cooperative and trusting
Conscientiousness	self-discipline, sense of duty, achievement striving

The Five Factor Model is credited largely to the work of Paul Costa and Robert McCrae, who wrote prolifically about it in the 1990s (see references in Chapter 2). They showed that in nearly all personality tests, there are five orthogonal dimensions along which people vary, similar to those given above. Considerable consensus has emerged around this finding within the past 10 years, so that it can now be considered the central model within personality psychology.

One point worth stressing about the FFM is that it is a descriptive rather than a theoretical model of personality. While there are theoretical interpretations which have been created to explain the existence of the five factors, there is little consensus about the theory.

Table 4.2: IPIP's NEO 300

Note: This table shows the 300 items of the International Personality Item Pool's NEO questionnaire referred to in Chapter 3. The first column gives the question number (questions cycle through domains in the order Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness). The second column indicates whether the question correlates positively or negatively with its domain. The third/fourth columns indicate which of the 30 facets each item belongs to. The final column gives the text of the item itself.

<i>#</i>	<i>+/-</i>	<i>Facet</i>	<i>Item</i>
1	+	N1 Anxiety	Worry about things.
2	+	E1 Friendliness	Make friends easily.
3	+	O1 Imagination	Have a vivid imagination.
4	+	A1 Trust	Trust others.
5	+	C1 Self-Efficacy	Complete tasks successfully.
6	+	N2 Anger	Get angry easily.
7	+	E2 Gregariousness	Love large parties.
8	+	O2 Artistic Interests	Believe in the importance of art.
9	+	A2 Morality	Would never cheat on my taxes.
10	+	C2 Orderliness	Like order.
11	+	N3 Depression	Often feel blue.
12	+	E3 Assertiveness	Take charge.
13	+	O3 Emotionality	Experience my emotions intensely.
14	+	A3 Altruism	Make people feel welcome.
15	+	C3 Dutifulness	Try to follow the rules.
16	+	N4 Self-Consciousness	Am easily intimidated.
17	+	E4 Activity Level	Am always busy.
18	+	O4 Adventurousness	Prefer variety to routine.
19	+	A4 Cooperation	Am easy to satisfy.
20	+	C4 Achievement-Striving	Go straight for the goal.
21	+	N5 Immoderation	Often eat too much.

22	+	E5	Excitement-Seeking	Love excitement.
23	+	O5	Intellect	Like to solve complex problems.
24	+	A5	Modesty	Dislike being the center of attention.
25	+	C5	Self-Discipline	Get chores done right away.
26	+	N6	Vulnerability	Panic easily.
27	+	E6	Cheerfulness	Radiate joy.
28	+	O6	Liberalism	Tend to vote for liberal political candidates.
29	+	A6	Sympathy	Sympathize with the homeless.
30	+	C6	Cautiousness	Avoid mistakes.
31	+	N1	Anxiety	Fear for the worst.
32	+	E1	Friendliness	Warm up quickly to others.
33	+	O1	Imagination	Enjoy wild flights of fantasy.
34	+	A1	Trust	Believe that others have good intentions.
35	+	C1	Self-Efficacy	Excel in what I do.
36	+	N2	Anger	Get irritated easily.
37	+	E2	Gregariousness	Talk to a lot of different people at parties.
38	+	O2	Artistic Interests	Like music.
39	+	A2	Morality	Stick to the rules.
40	+	C2	Orderliness	Like to tidy up.

41	+	N3	Depression	Dislike myself.
42	+	E3	Assertiveness	Try to lead others.
43	+	O3	Emotionality	Feel others' emotions.
44	+	A3	Altruism	Anticipate the needs of others.
45	+	C3	Dutifulness	Keep my promises.
46	+	N4	Self-Consciousness	Am afraid that I will do the wrong thing.
47	+	E4	Activity Level	Am always on the go.
48	+	O4	Adventurousness	Like to visit new places.
49	+	A4	Cooperation	Can't stand confrontations.
50	+	C4	Achievement-Striving	Work hard.
51	+	N5	Immoderation	Don't know why I do some of the things I do.
52	+	E5	Excitement-Seeking	Seek adventure.
53	+	O5	Intellect	Love to read challenging material.
54	+	A5	Modesty	Dislike talking about myself.
55	+	C5	Self-Discipline	Am always prepared.
56	+	N6	Vulnerability	Become overwhelmed by events.
57	+	E6	Cheerfulness	Have a lot of fun.
58	+	O6	Liberalism	Believe that there is no absolute right or wrong.

59	+	A6	Sympathy	Feel sympathy for those who are worse off than myself.
60	+	C6	Cautiousness	Choose my words with care.
61	+	N1	Anxiety	Am afraid of many things.
62	+	E1	Friendliness	Feel comfortable around people.
63	+	O1	Imagination	Love to daydream.
64	+	A1	Trust	Trust what people say.
65	+	C1	Self-Efficacy	Handle tasks smoothly.
66	+	N2	Anger	Get upset easily.
67	+	E2	Gregariousness	Enjoy being part of a group.
68	+	O2	Artistic Interests	See beauty in things that others might not notice.
69	-	A2	Morality	Use flattery to get ahead.
70	+	C2	Orderliness	Want everything to be "just right."
71	+	N3	Depression	Am often down in the dumps.
72	+	E3	Assertiveness	Can talk others into doing things.
73	+	O3	Emotionality	Am passionate about causes.
74	+	A3	Altruism	Love to help others.

75	+	C3	Dutifulness	Pay my bills on time.
76	+	N4	Self-Consciousness	Find it difficult to approach others.
77	+	E4	Activity Level	Do a lot in my spare time.
78	+	O4	Adventurousness	Interested in many things.
79	+	A4	Cooperation	Hate to seem pushy.
80	+	C4	Achievement-Striving	Turn plans into actions.
81	+	N5	Immoderation	Do things I later regret.
82	+	E5	Excitement-Seeking	Love action.
83	+	O5	Intellect	Have a rich vocabulary.
84	+	A5	Modesty	Consider myself an average person.
85	+	C5	Self-Discipline	Start tasks right away.
86	+	N6	Vulnerability	Feel that I'm unable to deal with things.
87	+	E6	Cheerfulness	Express childlike joy.
88	+	O6	Liberalism	Believe that criminals should receive help rather than punishment.
89	+	A6	Sympathy	Value cooperation over competition.
90	+	C6	Cautiousness	Stick to my chosen path.
91	+	N1	Anxiety	Get stressed out easily.
92	+	E1	Friendliness	Act comfortably with others.
93	+	O1	Imagination	Like to get lost in thought.

94	+	A1	Trust	Believe that people are basically moral.
95	+	C1	Self-Efficacy	Am sure of my ground.
96	+	N2	Anger	Am often in a bad mood.
97	+	E2	Gregariousness	Involve others in what I am doing.
98	+	O2	Artistic Interests	Love flowers.
99	-	A2	Morality	Use others for my own ends.
100	+	C2	Orderliness	Love order and regularity.
101	+	N3	Depression	Have a low opinion of myself.
102	+	E3	Assertiveness	Seek to influence others.
103	+	O3	Emotionality	Enjoy examining myself and my life.
104	+	A3	Altruism	Am concerned about others.
105	+	C3	Dutifulness	Tell the truth.
106	+	N4	Self-Consciousness	Am afraid to draw attention to myself.
107	+	E4	Activity Level	Can manage many things at the same time.
108	+	O4	Adventurousness	Like to begin new things.
109	-	A4	Cooperation	Have a sharp tongue.
110	+	C4	Achievement-Striving	Plunge into tasks with all my heart.

111	+	N5	Immoderation	Go on binges.
112	+	E5	Excitement-Seeking	Enjoy being part of a loud crowd.
113	+	O5	Intellect	Can handle a lot of information.
114	+	A5	Modesty	Seldom toot my own horn.
115	+	C5	Self-Discipline	Get to work at once.
116	+	N6	Vulnerability	Can't make up my mind.
117	+	E6	Cheerfulness	Laugh my way through life.
118	-	O6	Liberalism	Believe in one true religion.
119	+	A6	Sympathy	Suffer from others' sorrows.
120	-	C6	Cautiousness	Jump into things without thinking.
121	+	N1	Anxiety	Get caught up in my problems.
122	+	E1	Friendliness	Cheer people up.
123	+	O1	Imagination	Indulge in my fantasies.
124	+	A1	Trust	Believe in human goodness.
125	+	C1	Self-Efficacy	Come up with good solutions.
126	+	N2	Anger	Lose my temper.
127	+	E2	Gregariousness	Love surprise parties.
128	+	O2	Artistic Interests	Enjoy the beauty of nature.
129	-	A2	Morality	Know how to get around the rules.

130	+	C2	Orderliness	Do things according to a plan.
131	+	N3	Depression	Have frequent mood swings.
132	+	E3	Assertiveness	Take control of things.
133	+	O3	Emotionality	Try to understand myself.
134	+	A3	Altruism	Have a good word for everyone.
135	+	C3	Dutifulness	Listen to my conscience.
136	+	N4	Self-Consciousness	Only feel comfortable with friends.
137	+	E4	Activity Level	React quickly.
138	-	O4	Adventurousness	Prefer to stick with things that I know.
139	-	A4	Cooperation	Contradict others.
140	+	C4	Achievement-Striving	Do more than what's expected of me.
141	+	N5	Immoderation	Love to eat.
142	+	E5	Excitement-Seeking	Enjoy being reckless.
143	+	O5	Intellect	Enjoy thinking about things.
144	-	A5	Modesty	Believe that I am better than others.
145	+	C5	Self-Discipline	Carry out my plans.
146	+	N6	Vulnerability	Get overwhelmed by emotions.

147	+	E6	Cheerfulness	Love life.
148	-	O6	Liberalism	Tend to vote for conservative political candidates.
149	-	A6	Sympathy	Am not interested in other people's problems.
150	-	C6	Cautiousness	Make rash decisions.
151	-	N1	Anxiety	Am not easily bothered by things.
152	-	E1	Friendliness	Am hard to get to know.
153	+	O1	Imagination	Spend time reflecting on things.
154	+	A1	Trust	Think that all will be well.
155	+	C1	Self-Efficacy	Know how to get things done.
156	-	N2	Anger	Rarely get irritated.
157	-	E2	Gregariousness	Prefer to be alone.
158	-	O2	Artistic Interests	Do not like art.
159	-	A2	Morality	Cheat to get ahead.
160	-	C2	Orderliness	Often forget to put things back in their proper place.
161	+	N3	Depression	Feel desperate.
162	-	E3	Assertiveness	Wait for others to lead the way.
163	-	O3	Emotionality	Seldom get emotional.
164	-	A3	Altruism	Look down on others.

165	-	C3	Dutifulness	Break rules.
166	+	N4	Self-Consciousness	Stumble over my words.
167	-	E4	Activity Level	Like to take it easy.
168	-	O4	Adventurousness	Dislike changes.
169	-	A4	Cooperation	Love a good fight.
170	+	C4	Achievement-Striving	Set high standards for myself and others.
171	-	N5	Immoderation	Rarely overindulge.
172	+	E5	Excitement-Seeking	Act wild and crazy.
173	-	O5	Intellect	Am not interested in abstract ideas.
174	-	A5	Modesty	Think highly of myself.
175	-	C5	Self-Discipline	Find it difficult to get down to work.
176	-	N6	Vulnerability	Remain calm under pressure.
177	+	E6	Cheerfulness	Look at the bright side of life.
178	-	O6	Liberalism	Believe that too much tax money goes to support artists.
179	-	A6	Sympathy	Tend to dislike soft-hearted people.
180	-	C6	Cautiousness	Like to act on a whim.
181	-	N1	Anxiety	Am relaxed most of the time.

182	-	E1	Friendliness	Often feel uncomfortable around others.
183	-	O1	Imagination	Seldom daydream.
184	-	A1	Trust	Distrust people.
185	-	C1	Self-Efficacy	Misjudge situations.
186	-	N2	Anger	Seldom get mad.
187	-	E2	Gregariousness	Want to be left alone.
188	-	O2	Artistic Interests	Do not like poetry.
189	-	A2	Morality	Put people under pressure.
190	-	C2	Orderliness	Leave a mess in my room.
191	+	N3	Depression	Feel that my life lacks direction.
192	-	E3	Assertiveness	Keep in the background.
193	-	O3	Emotionality	Am not easily affected by my emotions.
194	-	A3	Altruism	Am indifferent to the feelings of others.
195	-	C3	Dutifulness	Break my promises.
196	-	N4	Self-Consciousness	Am not embarrassed easily.
197	-	E4	Activity Level	Like to take my time.
198	-	O4	Adventurousness	Don't like the idea of change.
199	-	A4	Cooperation	Yell at people.
200	+	C4	Achievement-Striving	Demand quality.
201	-	N5	Immoderation	Easily resist temptations.

202	+	E5	Excitement-Seeking	Willing to try anything once.
203	-	O5	Intellect	Avoid philosophical discussions.
204	-	A5	Modesty	Have a high opinion of myself.
205	-	C5	Self-Discipline	Waste my time.
206	-	N6	Vulnerability	Can handle complex problems.
207	+	E6	Cheerfulness	Laugh aloud.
208	-	O6	Liberalism	Believe laws should be strictly enforced.
209	-	A6	Sympathy	Believe in an eye for an eye.
210	-	C6	Cautiousness	Rush into things.
211	-	N1	Anxiety	Am not easily disturbed by events.
212	-	E1	Friendliness	Avoid contacts with others.
213	-	O1	Imagination	Do not have a good imagination.
214	-	A1	Trust	Suspect hidden motives in others.
215	-	C1	Self-Efficacy	Don't understand things.
216	-	N2	Anger	Am not easily annoyed.
217	-	E2	Gregariousness	Don't like crowded events.
218	-	O2	Artistic Interests	Do not enjoy going to art museums.

219	-	A2	Morality	Pretend to be concerned for others.
220	-	C2	Orderliness	Leave my belongings around.
221	-	N3	Depression	Seldom feel blue.
222	-	E3	Assertiveness	Have little to say.
223	-	O3	Emotionality	Rarely notice my emotional reactions.
224	-	A3	Altruism	Make people feel uncomfortable.
225	-	C3	Dutifulness	Get others to do my duties.
226	-	N4	Self-Consciousness	Am comfortable in unfamiliar situations.
227	-	E4	Activity Level	Like a leisurely lifestyle.
228	-	O4	Adventurousness	Am a creature of habit.
229	-	A4	Cooperation	Insult people.
230	-	C4	Achievement-Striving	Am not highly motivated to succeed.
231	-	N5	Immoderation	Am able to control my cravings.
232	+	E5	Excitement-Seeking	Seek danger.
233	-	O5	Intellect	Have difficulty understanding abstract ideas.
234	-	A5	Modesty	Know the answers to many questions.

235	-	C5	Self-Discipline	Need a push to get started.
236	-	N6	Vulnerability	Know how to cope.
237	+	E6	Cheerfulness	Amuse my friends.
238	-	O6	Liberalism	Believe that we coddle criminals too much.
239	-	A6	Sympathy	Try not to think about the needy.
240	-	C6	Cautiousness	Do crazy things.
241	-	N1	Anxiety	Don't worry about things that have already happened.
242	-	E1	Friendliness	Am not really interested in others.
243	-	O1	Imagination	Seldom get lost in thought.
244	-	A1	Trust	Am wary of others.
245	-	C1	Self-Efficacy	Have little to contribute.
246	-	N2	Anger	Keep my cool.
247	-	E2	Gregariousness	Avoid crowds.
248	-	O2	Artistic Interests	Do not like concerts.
249	-	A2	Morality	Take advantage of others.
250	-	C2	Orderliness	Am not bothered by messy people.
251	-	N3	Depression	Feel comfortable with myself.
252	-	E3	Assertiveness	Don't like to draw attention to myself.

253	-	O3	Emotionality	Experience very few emotional highs and lows.
254	-	A3	Altruism	Turn my back on others.
255	-	C3	Dutifulness	Do the opposite of what is asked.
256	-	N4	Self-Consciousness	Am not bothered by difficult social situations.
257	-	E4	Activity Level	Let things proceed at their own pace.
258	-	O4	Adventurousness	Dislike new foods.
259	-	A4	Cooperation	Get back at others.
260	-	C4	Achievement-Striving	Do just enough work to get by.
261	-	N5	Immoderation	Never spend more than I can afford.
262	-	E5	Excitement-Seeking	Would never go hang gliding or bungee jumping.
263	-	O5	Intellect	Am not interested in theoretical discussions.
264	-	A5	Modesty	Boast about my virtues.
265	-	C5	Self-Discipline	Have difficulty starting tasks.
266	-	N6	Vulnerability	Readily overcome setbacks.
267	-	E6	Cheerfulness	Am not easily amused.
268	-	O6	Liberalism	Believe that we should be tough on crime.

269	-	A6	Sympathy	Believe people should fend for themselves.
270	-	C6	Cautiousness	Act without thinking.
271	-	N1	Anxiety	Adapt easily to new situations.
272	-	E1	Friendliness	Keep others at a distance.
273	-	O1	Imagination	Have difficulty imagining things.
274	-	A1	Trust	Believe that people are essentially evil.
275	-	C1	Self-Efficacy	Don't see the consequences of things.
276	-	N2	Anger	Rarely complain.
277	-	E2	Gregariousness	Seek quiet.
278	-	O2	Artistic Interests	Do not enjoy watching dance performances.
279	-	A2	Morality	Obstruct others' plans.
280	-	C2	Orderliness	Am not bothered by disorder.
281	-	N3	Depression	Am very pleased with myself.
282	-	E3	Assertiveness	Hold back my opinions.
283	-	O3	Emotionality	Don't understand people who get emotional.
284	-	A3	Altruism	Take no time for others.
285	-	C3	Dutifulness	Misrepresent the facts.

286	-	N4	Self-Consciousness	Am able to stand up for myself.
287	-	E4	Activity Level	React slowly.
288	-	O4	Adventurousness	Am attached to conventional ways.
289	-	A4	Cooperation	Hold a grudge.
290	-	C4	Achievement-Striving	Put little time and effort into my work.
291	-	N5	Immoderation	Never splurge.
292	-	E5	Excitement-Seeking	Dislike loud music.
293	-	O5	Intellect	Avoid difficult reading material.
294	-	A5	Modesty	Make myself the center of attention.
295	-	C5	Self-Discipline	Postpone decisions.
296	-	N6	Vulnerability	Am calm even in tense situations.
297	-	E6	Cheerfulness	Seldom joke around.
298	-	O6	Liberalism	Like to stand during the national anthem.
299	-	A6	Sympathy	Can't stand weak people.
300	-	C6	Cautiousness	Often make last-minute plans.