Are smarter people more procedure invariant? An online experiment investigating the effects of cognitive ability on preference reversal

Mark Wainwright[1]

June 2011
Dissertation, University of Nottingham

Abstract

In the preference reversal literature, models of stochastic choice and error are increasingly endorsed as explanations of the phenomenon. In this paper, I document an online experiment examining whether there exists the negative link between cognitive ability and observation of preference reversal that an error-led approach may imply. I find that cognitive ability has no effect on whether subjects exhibit PR or not, but that when subjects do so, the extent of reversal is strongly and negatively linked to cognitive ability. I find that part of this effect is caused by increased simple error making amongst lower cognitive ability subjects, but argue that further research is needed to establish the full nature of the association, incorporating more detailed notions of error and cognitive ability.

JEL classification: C90, D01, D81

Keywords: preference reversal, cognitive ability, errors, decision making, risky choice

1. Introduction

Preference reversal (henceforth PR), the tendency for revealed preferences in decision-making tasks to depend systematically on the processes used to elicit them, has become an accepted empirical regularity since observed by Lichtenstein & Slovic in 1971. Extensive testing has established that the effect is robust and systematic—in experiments, subjects often choose 'P-bets', with high probabilities and low prizes, over '$-bets', with low probabilities and high prizes, but nonetheless give the latter a higher valuation. The opposite of this reversal is observed far less frequently. Taking these results at face value, they present a convincing case for the violation of expected utility theory, and as Grether & Plott argue, "that no optimization principles of any sort lie behind even the simplest of human choice behavior" (1979, p. 623).

Focus has since shifted from robustness tests to investigating explanations of PR, which tend to fall into three categories: 1) violations of the preference axioms of transitivity, independence or reduction; 2) failures of procedure invariance; and 3) error-making, stochastic choice and imprecision .

Within the latter, a key result is due to Schmidt & Hey (2004), who find that many PRs can be explained by errors made in valuation tasks, and argue that simple error has been both overlooked and under-examined as an explanation for PR, in ways that it has not for other violations of rational theory (p. 208). If error is truly a significant cause of PR then it is worth investigating the criteria for and characteristics of error, and what effect these might have on observation of PR.

In this paper, I address the question of whether PR is negatively linked to the cognitive ability of the individual, and if so, whether this link is caused by greater errors being made by those of lower cognitive ability. On a broader level, this seeks to answer the question of whether PR is a consequence of bounded rationality, with cognitive ability being a factor that limits the ability of individuals to express their preferences consistently and accurately across different elicitation procedures that standard economic theory would consider equivalent.

The results from my experiment ultimately show that cognitive ability has no significant effect on the probability that subjects will make a PR, but that when subjects do so, subjects of a lower cognitive ability give valuations that are significantly further away from consistency with their choices. Using a notion of error as demonstrated by choosing a dominated option, I claim that part of this association can be explained by a higher degree of error amongst low ability subjects.

1.1. Literature review

The effect of cognitive ability on preferences like risk aversion and time discounting is beginning to receive some

[1] *Email address*: markwainwright1@gmail.com

attention in the literature, with systematic links often being found. Benjamin *et al.* (2006) study the effects of cognitive ability on "short term discounting" and "small stakes risk aversion", which they assert are "preference anomalies" (p. 2), and found that subjects with higher cognitive ability are significantly less likely to exhibit such anomalies. Frederick (2005) found a similar result—that individuals with higher scores in a "cognitive reflection test" demonstrate reduced future discounting and risk aversion.

Frederick notes that cognitive ability has rarely been applied to decision making, and justifies further research through the fact that "a neglected aspect does not cease to operate because it is neglected" (p. 25). Whilst the effect of cognitive ability on preferences like risk aversion and time discounting is receiving increased attention in the literature, to my knowledge its effect on PR has never been examined. Since significant links between error-making and PR, and cognitive ability and error-making have been shown, there is every reason to suggest that studying its effects on PR is worthwhile.

Even in the early origins of PR, bounded rationality considerations were present. To broaden their set of explanations for PR, Slovic & Lichtenstein (1983) reference work by March (1978), which presents several weaknesses of rational theory. March argues that, whilst rational theory assumes current and future preferences are "exogenous, stable and known with adequate precision to make decisions unambiguous" (p. 589), the extent to which this is true is limited by cognitive ability—"man [is] not smart enough to be rational" (p. 588). He advocates bounded rationality models as a better approach to decision making, due to his assertion that "limitations of [...] information capacity affect information processing about preferences just as they affect information processing about consequences" (p. 598). March also touches on another sense of bounded rationality—that cognitive ability limits the ability to express preferences, which suggests the relevance of procedure invariance as an explanation of PR. "Decisionmakers [sic] are driven to techniques of limited rationality by the exigencies of the situation in which they find themselves" (p. 598).

Several models of how this might be the case have been proposed. Error models, for example, imply that the ability to make decisions consistent with rational preferences is limited by error making in responses. For this theory to be a candidate explanation of PR, errors would need to be more prevalent on either the choice or the valuation task; symmetrical errors on both would be unlikely to produce the robust one-way reversal.

A consideration of the cognitive loads associated with each task suggests that this might be the case. Schkade & Johnson (1989) use the time taken on each task as a proxy for the effort expended (p. 210), and find that valuation tasks require significantly more effort from the subject. Considering that a procedure-invariant view would suggest the choice task requires more time because it involves the consideration of two alternatives rather that one, this is a significant result.

The results of Shiv & Fedorikhin (1999) lend themselves to a psychological explanation of why greater cognitive load may lead to more errors. They find that people are more likely to make "impulsive" decisions when faced with the increased cognitive load, which could lead to greater chance of error and a less precise expression of true preferences. Schmidt & Hey (2004) present a result that is compatible with this interpretation—that many PRs can be shown to be caused by errors in the valuation task and very few by choice errors. They reach this conclusion using an "extended error model" that identifies decisions as stochastic or deterministic; when responses varied over repetitions of the same task, they consider at least one response an error (p. 209). Their conclusion, combined with previously discussed results on the relative cognitive load of the valuation task, implies that cognitive ability may limit subjects' ability to make valuations without error. Whether this could contribute to PRs is indeterminate from existing research, however.

2. Experimental design

To test my hypothesis, I ran an online experiment with tasks designed to test PR and cognitive ability. The experiment made use of the open source survey package LimeSurvey, although its format and style had more in common with that of a laboratory experiment. The experiment consisted of four sections: demographic questions (see Section 2.1), choice and valuation tasks (see Section 2.2) and a cognitive ability test (see Section 2.3).

2.1. Demographic questions

The experiment was designed to be entirely and transparently anonymous, so no personal details were collected beyond basic demographic data, which were used as control variables in my statistical analysis. The questions were limited to sex, age (in 10-year ranges) and the subject's field of study if they are a university student. I chose not to ask for more detailed information because of the importance of subjects being satisfied of their anonymity, especially due to the potential sensitivity of their cognitive ability score.

2.2. Choice and valuation tasks

My methods of investigating PR are very similar to the canonical method introduced by Lichtenstein & Slovic (1971), with some additions. I used six sets of two gambles, replicating the payoff and probability parameters employed by Cubitt *et al.* (2004), which they derived from scaling up payoffs of Tversky *et al.* (1990). Each set contains a 'P-bet', with a high probability of winning a small prize, and a '\$-bet', with a lower probability of a higher prize, although the two gambles' expected values are similar ($P_p\pi_p \approx P_\$\pi_\$$ where P_x is the probability of winning in gamble x, and π_x is the prize if gamble x is won). The sixth set featured a "dominated option", where P is strictly preferable to $, which I used to measure error, in the manner of Cubitt *et al.* (2004, p. 715).

Because Grether & Plott (1979) found that task ordering does not significantly affect outcomes, all choice tasks were carried out together, followed by the valuation tasks, in order to maximise simplicity, and so that each type of task could be explained separately, to reduce misunderstanding.

In the choice section, subjects were presented with six tasks in which they were shown both gambles and were asked to pick which they prefer, or to indicate that they were indifferent. I allowed indifference because of the argument of Grether & Plott (1979)—that not doing so may create reversals only because of "a systematic resolution of indifference on the part of subjects forced to record a preference" (p. 626). I discuss the implications that this might have for my analysis in Section 2.5.1.

In the valuation section, subjects faced 12 screens showing the same gambles individually, and were asked for their reservation price for each gamble. Responses were given through a slider, which enforced of a minimum value of zero and a maximum value of the prize of the gamble, to demonstrate clearly to subjects that valuations above the gamble's prize cannot be a minimum reservation price, regardless of the subject's level of risk aversion. For example, valuation of a £32 gamble allowed responses the range 0–32, inclusive.

A particular emphasis was placed throughout on the fact that subjects were being asked for the minimum price they would accept (*"Before you submit your answer, ask yourself 'is that the smallest amount of money I would accept to not play that gamble?'"*). This was primarily to avoid the problem of strategic responses, since a natural reaction to being asked for a selling price may be to give a reservation price above one's true valuation of the gamble (Grether & Plott, 1979, p. 626). With this emphasis, and the fact that a lack of monetary payoffs meant that there was no incentive to overstate valuations, I do not expect this to be a problem.

In all choice and valuation tasks, probabilities were expressed in the manner of Schmidt & Hey (2004)—through pie charts representing the probabilities of winning. This has the benefit of making probabilities easier to visualise than if they were solely represented by a number, without excluding subjects who may prefer taking cues from the numbers.

Within each section, tasks were presented to each subject in a random order. This was with the intention of controlling for learning effects and the likelihood of subjects making fewer errors as they become more familiar with the tasks and response modes. Because each task was presented to roughly the same proportion of subjects first (and so on), the impact of dynamic effects such as learning is precluded. It also meant that subjects could not easily associate identical gambles with each other between tasks, reducing the likelihood of a 'experimenter's bias' arising, if subjects 'saw through' the experiment and made decisions solely because they believed they were in line with what I expected.

2.3. Cognitive ability test

The cognitive test closely resembled the one used by Sousa (2010). It comprised 12 questions, in four categories designed to assess different reasoning skills: quantitative reasoning, sequential reasoning, verbal reasoning and Frederick's "cognitive reflection" (2005).

I used a broad test because my research question involves general cognitive ability, meaning it is important to gain a measure of a wide range of skills. The cognitive reflection test is also particularly relevant to my experiment, both because it was specifically designed for comparison with decision-making outcomes, and because it discriminates between the ability of subjects to apply the detailed, extended and "reflective" thinking that the valuation task may require.

As with previous tasks, the order of questions was randomised. Also, in line with Sousa (2010), subjects were given a maximum of one minute to answer each question. Although Sousa does not justify his use of this limit, I implemented it with the intention of ensuring that scores reflected fundamental differences in cognitive ability, rather than the willingness (or ability) of subjects to spend extended amounts of time considering the answer. Although Frederick (2005) does not discuss this issue either, I believe gaining control over time constraints is especially important with the cognitive reflection test, in which the propensity for the correct answer to be given is strongly linked to the time taken considering it.

2.4. Subject pool

My subject pool mainly consisted of friends and family whom I asked to participate. In turn, they often recommended the experiment to others. As I illustrate in Section 3, this generated a subject pool with a more heterogeneous range of characteristics than is often found in laboratory experiments, which is usually limited to students of a young age. A particular advantage of this is that, if students are not fully representative of the wider population in terms of decision making, my results may have more external validity than those generated from narrower samples. It is worth noting, however, that general economic theories of decision making and preference are not conditional on demographic factors, so my experiment is no better a test of existing theories than a lab experiment in this sense.

2.5. Data analysis

My main data analysis strategy was to run regressions of measures of PR on cognitive ability and other control variables, to establish whether the null hypothesis that cognitive ability has no significant effect on PR can be

Table 1: Categorisations of behaviour

Valuation task		$V(\$) > V(P)$	$V(\$) = V(P)$	$V(\$) < V(P)$
	P	"Standard reversal"	"Indifference"	"Consistent P"
Choice task	$	"Consistent $"	"Indifference"	"Counter reversal"
	Indifferent	"Indifference"	"Indifference"	"Indifference"

rejected. I used data from set 6, the dominated option, to investigate whether error making is linked to cognitive ability in a similar manner, and whether it can explain PR in the other tasks. These processes involve several issues, which are examined in the next three sub-sections.

2.5.1. Categorisation of behaviour

To draw conclusions about PR, I categorised behaviour into one of five categories, and will later report the frequency of each, for each set and in aggregate.

Consistent P. P is chosen, and valued higher than $.

Consistent $. $ is chosen, and valued higher than P.

Standard reversal. P is chosen, and $ is valued more highly.

Counter reversal. $ is chosen, and P is valued more highly.

Indifference. Indifference is chosen and/or P and $ are valued equally.

Much existing research does not allow indifference in the choice task, and thus concludes that choosing P and valuing P and $ equally is permissible as a consistent preference. For example, Cubitt *et al.* explicitly state that they take this approach, because they do not allow indifference (2004, p. 718). However, since I allowed indifference, such a set of decisions would be clearly inconsistent. Considering this, one approach could be to treat valuation indifference following a non-indifferent choice as a PR.[2] However, this may be too stringent a requirement, and inflate the PR rate (indeed, only two observations represented indifference consistent to both tasks). Because of this, I include observations with indifference in either task in a fifth category, indifference. This effectively separates out observations that contain equal valuations and means they are considered neither consistent nor a reversal, and will therefore not factor into my analysis.[3]

For additional clarity, Table 1 presents a summary of the criteria for each category, for each combination of decision in the two tasks.

[2]For example, if a subject chose P but gave equal valuations to P and $, this could be judged a standard reversal, because $ is valued higher than the choice indicates it should be. The only consistent response to choice indifference in this interpretation is to give P and $ exactly equal values.

[3]Tversky *et al.* also do not allow indifference but take the same step as me in "exclud[ing] tied prices from the analysis" to avoid inflating the PR rate (1990, p. 209). Papers that do allow indifference are less clear.

2.5.2. Measures of preference reversal

In order to compare cognitive performance and PR, I require a formalised method of quantifying it. I propose three distinct 'measures', which were calculated for each gamble set. Each subject's overall measure was then given by a simple average of their measures in each of the five gamble sets.

I. Frequency of reversal. The amount of sets in which the subject exhibited a standard or counter reversal. Can be considered a measure of the probability that the subject will make a PR in any given set.

II. Extent of reversal. If a PR has occurred, the amount by which the subject's valuation of the unchosen gamble exceeds that of the chosen gamble. If decisions are consistent, this measure is declared as zero. For each gamble set, defined as (where $V(.)$ represents the valuation given to a gamble):

$$|V(\$) - V(P)|$$

III. Relative extent of reversal If a PR has occurred, the extent of reversal expressed as a percentage of the valuation of the chosen gamble (which is therefore the lower of the two), and therefore defined as follows, conditional on PR:

$$\frac{|V(\$) - V(P)|}{\min\{V(\$), V(P)\}}$$

I used a concept of the extent of reversal with the intention of being able to observe more subtle differences in PR between subjects rather than a binary observation of whether a reversal has occurred. In addition, it allowed me to differentiate between minor divergences from consistency, which are more likely to be consequences in a margin of error in expression than large valuation differentials are, rendering them a more significant violation of expected utility theory. Measure II seems to be equivalent to Grether & Plott's 'magnitude of reversal' (1979, pp. 632-633). Measure III was used to acknowledge that the extent of reversal should not just be examined in absolute terms; for example, a small reversal would be more significant against lower valuations than high.

Since measuring PR through a uniform weighting of gamble sets makes the implicit assumption that an equivalent reversal in each set is an equally significant violation, a concern could be that differences between expected values and probabilities could render reversals in one set more significant in some way than in another. However, by using gambles that are all of the same format, and that are relatively similar in terms of probabilities (ranging from 19-50% for $-bets and 81-97% for P-bets), I do not expect this point to be to the detriment of my results.

2.5.3. Control variables

To increase confidence that any observed link between cognitive ability and PR is solely caused by cognitive ability and not interaction with other variables, I used several control variables: 1) sex; 2) age; and 3) whether the subject studies economics.

Age was elicited in the form of 10-year ranges, to increase anonymity. To use these data in my regressions, I took the midpoint of the subject's age range as a proxy for their actual age.

I also included an 'economics student' dummy in my regressions to control for a specific form of experimenter's bias that I propose may otherwise have an effect. Economics students are more likely to have studied experimental or behavioural economics and thus be aware of the PR phenomenon, and may recognise the methods that I use to investigate it. If this is the case then they may be more likely to pointedly avoid making inconsistent decisions than the average individual may. This may also occur more generally when economics students are more aware of theories of decision making, and base their responses more closely on expected value calculations than is typical. To maximise external validity and robustness, I control for any effects along these lines.

3. Results

In the 23 days that my experiment was open for participation, it received 103 full responses. Subjects took an average of 15 minutes and 40 seconds to complete the experiment. The median age of subjects was 27, 66% of subjects were male, 49% were students and 13% studied economics.

3.1. Preference reversal

Table 2 presents a summary of the results obtained from the choice and valuation tasks, in each set and in aggregate. The last two columns show the results aggregated across sets 1-5; set 6 is the dominated choice with the specific purpose of measuring error, and is analysed in Section 3.4.

Having placed each set for each subject into one of the five categories defined in Section 2.5.1, the data show a strong preponderance of PR. Whilst 45.4% of observations were consistent decisions, 45.7% were PR of either type. The key proof that PR is systematic lies in the asymmetry of standard and counter reversals, which is strongly illustrated in my results: 80.1% of all P choices are contradicted by valuation, whilst just 4.1% of $ choices are contradicted; of all PRs, 96.2% were in the 'standard' direction. These results are broadly in line with existing research.

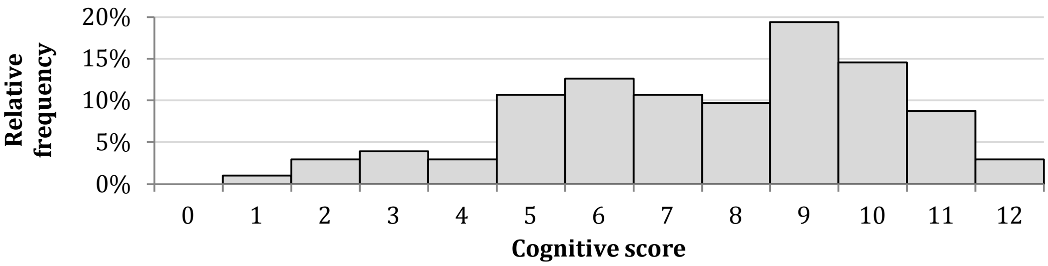
The main divergence from existing results lies in the choice task; in my experiment, 54.8% of subjects chose P and 42.3% chose $, but in most other research, the choice of $ is more frequent (Grether & Plott, 1979; Cubitt *et al.*, 2004, *etc.*). This seems to suggest that, without direct monetary incentives, subjects overestimate their risk aversion. However, since the more important comparison of choice and valuation closely replicates existing research, this is unlikely to be a cause for concern.

Relatively few subjects expressed indifference in either task; of 515 observations, indifference was chosen in 2.9%, and valuations were equal in 6.4%. The infrequency of indifferent choices perhaps justifies why many studies choose not to allow indifference, even though a higher percentage of subjects indicated they were indifferent through valuation. Moreover, the low incidence of valuation indifference

Table 2: Summary of choice and valuation results

Gamble sets	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Agg. (1-5)	
Choice								
P	61	59	64	59	39	98	282	54.8%
$	40	35	37	44	62	0	218	42.3%
Indifferent	2	9	2	0	2	5	15	2.9%
Categories of behaviour								
Consistent P	12	8	8	5	2	55	35	6.8%
Consistent $	37	27	34	42	59	0	199	38.64%
Standard reversal	48	44	52	48	34	14	226	43.88%
Counter reversal	3	3	2	0	1	0	9	1.75%
Indifference	3	21	7	8	7	34	46	8.93%
Observations	103	103	103	103	103	103	515	

Figure 1: Distribution of cognitive test scores



means that my potentially contentious approach—that indifference cannot contribute to either a consistent decision or a reversal (see Section 2.5.1)—is unlikely to have a significant impact on my results.

3.2. Cognitive ability

In the cognitive test, scores spanned the entire range from 1 to 12 out of a maximum 12 points, with a median of 8 and mean of 7.63. The standard deviation of scores was 2.55. Although scores were skewed towards the higher end, the frequency of scores declines almost to zero at both ends of the distribution, so there is no reason to suggest that the test was too easy nor too difficult. The full distribution is shown in Figure 1.

3.3. Cognitive ability and preference reversal

3.3.1. Non-parametric tests

For a basic, summative analysis of the data, subjects were divided into two cognitive groups: 'high', which includes subjects who scored above the median score and 'low', which includes those who scored below or equal to the median score. Descriptive statistics were then calculated for the measures of reversal within each group.

From the results in Table 3, it is clear that by all measures, the low cognitive ability group exhibits a greater degree of PR than the high. Mann-Whitney U tests were used to test the null hypotheses that the two groups are drawn from the same distribution with the same median, in order to rule out chance as an explanation for this difference. This null hypothesis can be rejected at the 1% level for the second and third measures, but only at the 10% level for the first measure. Therefore, the key results here are that low cognitive ability subjects' extents of reversal are higher than that of high cognitive ability subjects by £0.78, or 26.8% of the value given to their chosen gamble.

3.3.2. Regression analysis

A more precise approach is necessary to investigate the link between PR and cognitive ability further. In this section, I propose a model that defines the measure of reversal as a function of the subject's cognitive ability score and the control variables of age, gender, and whether the subject is an economics student. The first three columns of Table 4 report the ordinary least squares estimates obtained from regressions that have each of the three measures of reversal as the dependent variable.

The effect of cognitive ability on the first measure is not significantly different from zero, which is unsurprising given the weak significance in the aggregated non-parametric tests. However, it does have a statistically significant effect on the second and third measures at the 5%

Table 3: Measures of reversal by cognitive group

Cognitive groups	Low	High	Mann-Whitney test statistic
I. Frequency of reversal	0.507	0.396	$z = -1.688^{*}$
	(0.283)	(0.337)	($p = 0.091$)
II. Extent of reversal	1.860	1.073	$z = -2.993^{***}$
	(1.932)	(1.400)	($p = 0.003$)
III. Relative extent of reversal	0.550	0.282	$z = -3.006^{***}$
	(0.563)	(0.353)	($p = 0.003$)
Observations	56	47	

Notes: Cells report group mean and standard deviation (in parentheses). Mann-Whitney test p-values are two-tailed.

Table 4: Regressions of reversal measures and error-making on cognitive score

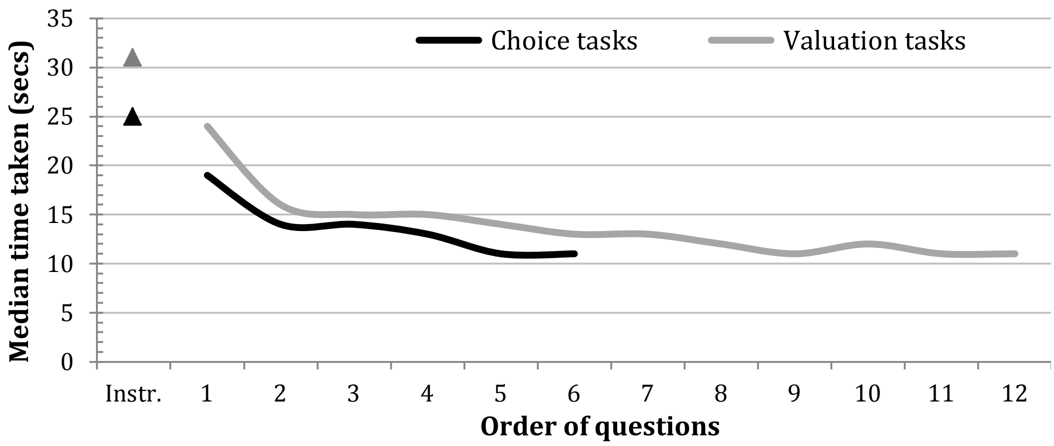
Dependent variable	Measures of preference reversal			Dominated
	I. Frequency	II. Extent	III. Relative extent	valuation dummy
Constant	0.655***	3.544***	1.099***	0.553***
	(0.112)	(0.614)	(0.171)	(0.127)
Cognitive score	-0.006	-0.134**	-0.048***	-0.038***
	(0.012)	(0.066)	(0.018)	(0.014)
Male dummy	-0.236***	-1.201***	-0.314***	-0.034
	(0.063)	(0.348)	(0.097)	(0.072)
Age	0.000	-0.006	-0.002	-0.003
	(0.002)	(0.012)	(0.003)	(0.002)
Economics student dummy	-0.066	-0.226	-0.131	-0.034
	(0.088)	(0.482)	(0.134)	(0.010)
Observations	103	103	103	103
R-squared	0.157	0.194	0.225	0.113

Notes: Cells report coefficient and standard error (in parentheses). $^{*}: p < 0.1$;$^{**}: p < 0.05$;$^{***}: p < 0.01$.

Table 5: Regressions of reversal measures on dominated valuations

Dependent variable	Measures of preference reversal		
	I. Frequency	II. Extent	III. Relative extent
Constant	0.613***	2.540***	0.716***
	(0.092)	(0.506)	(0.140)
Dominated valuation dummy	0.018	0.0652*	0.301**
	(0.086)	(0.475)	(0.132)
Male dummy	-0.242***	-1.1314***	-0.349***
	(0.062)	(0.344)	(0.095)
Age	0.000	-0.006	-0.002
	(0.002)	(0.121)	(0.003)
Economics student dummy	-0.073	-0.344	-0.168
	(0.087)	(0.481)	(0.133)
Observations	103	103	103
R-squared	0.155	0.176	0.213

Figure 2: Time spent on choice and valuation tasks



Notes: The question order on the horizontal axis represents the order in which each subject faced each question, e.g. 1 represents the first question answered, etc.

and 1% levels respectively. The coefficients indicate that a unit increase in cognitive ability score decreases the extent of reversal by £0.13 and the relative extent by 4.8%.

3.4. Cognitive ability and error

The next question to ask is whether the propensity to make this error in valuation is related to cognitive ability. Here I use a simple notion of error employed by Cubitt *et al.*—that subjects are deemed to have made an error where they choose or value the dominated option in set 6 over the dominating option (2004, p. 715). Because the P-bet in this case offers a higher probability than the $-bet of an equal prize, any preference for the latter can only represent a mistake or a misunderstanding, assuming only that subjects prefer greater expected wealth to less. In my experiment, no subjects chose the dominated option (although 4.9% expressed indifference between the two), but 13.6% valued it more highly. This is very similar to Cubitt *et al.*'s results of 4.4% and 13% respectively, when monetary values were solicited (2004, pp. 722-723).

The rightmost column of Table 4 shows the ordinary least squares estimates of a model regressing the dummy variable of whether the subject violated dominance in valuation on cognitive ability and the control variables. These results show that cognitive ability has a significant (but minor) negative effect on violations of dominance: a unit increase in cognitive test score makes the subject 3.8% less likely to make this error ($p < 0.01$). However, only 14 subjects made this error in valuation, providing a limited sample from which to draw conclusions.

A further regression investigating whether errors in set 6 have an association with PR in the other sets produced the coefficient estimates shown in Table 5. Subjects who violated dominance in valuation had a 30.1 percentage point higher relative extent of reversal than those who did not, across the other sets ($p < 0.05$) and a higher absolute extent of £0.65 ($p < 0.1$). This suggests that PR can be at least partly attributed to errors, as observed through the dominated set where they are most transparently observable. This corroborates the finding of Schmidt & Hey (2004)—that some PRs can be attributed to errors in the valuation task, but seldom in the choice task.

Insofar as violations of dominance represent error, a combination of this and the earlier result linking cognitive ability to error leads us to the conclusion that higher error making amongst lower cognitive ability subjects can partially explain the negative association between cognitive ability and PR.

3.5. Time taken

My data on the time subjects took to answer each question are in line with the results of Schkade & Johnson (1989). As Figure 2 shows, subjects generally spent longer valuing a single gamble than choosing between two. Although the difference is not large, the null hypothesis that the median difference between the time spent on choice and valuation tasks is zero can be rejected at the 10% level using a Wilcoxon signed-rank test[4], which may provide an indication that the valuation tasks require more effort. Moreover, this result differs from a potential conclusion of standard economic theory and procedure-invariance—that subjects should spend longer on choice tasks, because they are appraising two gambles rather than one (Schkade & Johnson, 1989, p. 210).

Using Schkade & Johnson's claim that time represents "an approximate measure of total effort" (1989, p. 210), this implies that valuation tasks involve a higher cognitive load and require more effort from the subject. Whilst

[4]The signed-rank test was used because it does not assume that the samples are independent—the time that each subject spends on the choice and valuation tasks is likely to be strongly interrelated.

the fundamental unobservability of effort makes this measure little more than an approximation, the fact that subjects took an average 6 seconds longer to read the instructions for the valuation task does add strength to the claim (significant at the 5% level under a Wilcoxon signed-rank test).

Despite the noted differences in cognitive load, there is little support for a claim that between-subject variation in effort has any connection to PR *per se*. Multiple regressions of the measures of reversal on the time spent on each task and reading the instructions yielded only small coefficients that were never significantly different from zero at the 10% level. Nor is there any evidence that any measure of time taken has a significant effect on the propensity to make errors, or that time taken is linked to cognitive ability in any way. Therefore, although valuation tasks seem to involve a greater cognitive load, as expected, variation between subjects in the time spent answering questions cannot explain PR directly.

4. Discussion

I previously argued that my online experiment may have an advantage over more traditional economic experimentation in that it provided a wider subject pool than is considered standard (see Sections 2.4 and 3). In addition, my PR results replicated existing findings fairly closely, even though my experiment lacked many procedures that are common to most existing research, including monetary incentives and the all-round control of the laboratory. However, since my methodology varied in significant ways to the 'tried and tested', it is important to consider whether the results I have obtained should be taken seriously.

First and foremost, my experiment did not offer any direct incentives to the subjects, either in the form of a outcome-contingent payoff or of a 'turn up' fee. The power of monetary payoffs in experimental economics is arguably to replicate the incentives and constraints that participants would be subject to in more naturally occurring settings. In PR experiments, incentives are typically designed to be "truth-revealing" in that they are incentive-compatible, meaning it is in subjects' interest to reveal their preferences accurately (Berg *et al.*, 2010). It could also be argued that incentives encourage subjects to put the same amount of thought and care into their decisions as they would in a more naturally occurring setting. However, the second wave of PR research has shown that incentives do not have a singificant effect on observation of PR (Berg *et al.*, 2010, p. 443), and where 'imaginary' payoffs have been used, as in my experiment, results do not differ fundamentally from treatments in which actions are incentivised.

Moreover, a possible benefit of my methodology is that it sidesteps issues of incentive compatibility and, in particular, the independence axiom. Since the commonly-used Becker-DeGroot-Marschak method assumes independence, some have raised questions about whether PR results produced using this mechanism could be inaccurate, if independence was found not to hold. Indeed, Tversky *et al.*. note that "there may be no incentive compatible scheme for the elicitation of selling prices that does not rely on the independence axiom" (1990, p. 207). Since my experiment did not involve incentives in any sense, it precludes the possibility of confounding effects caused by failures of independence and other intricacies of whether procedures can be truly incentive compatible.

In my experiment, it is clear that subjects had no incentive not to reveal their true valuations, even though they were not being paid to do so. Because of this, the focus must be on whether a lack of incentives had an effect on effort that could have spuriously caused the link between cognitive ability and preference reversal that I observed. This could have been the case if low ability subjects responded to a lack of incentives with a greater reduction than higher ability subjects did, and thus made more errors and PRs. However, because I found no significant link between cognitive ability and a proxy for effort (the time spent on tasks), and no link between variations in effort and preference reversal or error, there is no evidence to suggest that my result could be an artefact of unincentivised subjects.

Perhaps the most significant feature that is truly specific to the laboratory is heightened control. However, in the case of my experiment, it is unclear what the merits of tight control would have been. In experimental games, where subjects interact with each other and influence each other's payoffs, the power of the laboratory is that it can enforce complete anonymity and a lack of communication, meaning the only incentives that the subjects face are those that are set and controlled by the experimenter. Whilst my experiment did offer full anonymity, it is less clear whether subjects will have worked in isolation, and what effects this could have had on my results. However, since my experiment concerned individual decision-making, it is less integral that subjects do not communicate than it would be if they were interacting together. Regardless, the preliminary instructions did request that subjects not discuss the experiment until afterwards and that this would ensure the integrity of my results. Given this, and the assumption of a degree of goodwill, I do not expect this issue to be a significant issue.

These arguments notwithstanding, it is impossible to be sure that my experimental procedures did not adversely affect the validity of my result in some way. Ideal follow-up research should test the robustness of my results by replicating existing PR experiments more closely—in a laboratory setting, with the isolation of subjects and making use of an incentive-compatible payoff scheme such as the 'ordinal payoff scheme', which is less dependent on independence (Tversky *et al.*, 1990, p. 207). As Berg *et al.* certify, "replication is a cornerstone of the scientific method" (2010, p. 461).

In addition, the validity of my claims is limited by the extent to which the 12-question cognitive test is a good proxy for general cognitive ability. Since the test included questions covering four distinct types of reasoning skill, I believe that it represents an adequate measure of cognitive ability, within the context and constraints of my experiment. However, future research into this question could add additional certainty by using a more comprehensive and peer-reviewed method of measuring cognitive ability, perhaps using a standardised intelligence quotient (IQ) test.

5. Conclusions

To address the question of whether cognitive ability is linked to preference reversal, I ran an experiment that combined the canonical procedures testing for PR, and a simple test of cognitive ability. Perhaps unsurprisingly, there exists no link between cognitive ability and the binary measure of whether subjects exhibit PR or not. However, because it is clearly evident that cognitive ability is not the sole determinant of PR, I used measures that allowed me to examine PR more finely. Under these measures, I found a large and significant negative relationship between cognitive ability and the absolute and relative extents of reversal, meaning subjects of lower cognitive ability gave valuations significantly further from consistency with choice than those of higher ability.

My data also bear insight into the mechanisms behind the link between PR and cognitive ability. The effect is partially explained by the higher propensity of low cognitive ability subjects to make basic errors, as revealed through violations of dominance in the valuation task. This is likely to be because valuation tasks involve a higher cognitive load than choice tasks, as evidenced by subjects spending significantly more time on them, and by *ex-post* discussions with subjects. This assertion is backed up by evidence that PRs and errors are not caused by between-subject variation in effort put into tasks, implying issues of cognitive ability and load are more relevant as a cause of errors and PRs.

My experiment was not designed to discriminate between other possible explanations for PR, and thus cannot make any grand statements about which direction the literature should take in future. However, since my data show that a proportion of subjects make even simple errors, it is clear that error models and stochastic choice must play a part in explaining PR. The insight that my research brings to this branch of the literature is that such errors are linked to cognitive ability, to some extent. This suggests an approach to error that is centred on bounded rationality, with cognitive ability being considered a limitation to subjects' ability to express preferences without a stochastic component, particularly on the valuation tasks. This connection is clearly illustrated by the fact that, of the subjects in the higher quartile of cognitive scores, just one violated dominance, whilst in the lower quartile, nine made this error.

I must reiterate that the notion of error I employ is narrow and simplistic, and only captures one basic way in which subjects could be judged to have made an error. Further research should extend knowledge of how error and stochastic choice are affected by cognitive ability, through experiments designed to detect error in a more sophisticated manner, such as in Schmidt & Hey (2004). This approach should also investigate which specific skills, or facets of cognitive ability, are linked to error making and PR, and together with a more detailed analysis of errors, would pave the way towards an empirical theory or understanding that explains my findings, insofar as they are accurate.

The intention of the paper was to answer a simple empirical question of whether preference reversals and cognitive ability are negatively linked, and I have shown that this is the case. Herein lies the main contribution of my research; whilst cognitive ability is clearly not the sole determinant of PR, it undoubtedly plays more of an explanatory role than previously thought. My results are not heavily prescriptive of a future approach to PR research, but they do prompt further research in pursuit of a more complete understanding of the effect that I have observed. I believe that I have also shown that online experiments can be a viable experimentation medium in certain cases, and even that they can provide distinct benefits over the university laboratory, primarily in terms of allowing people with a wide range of characteristics and backgrounds to participate.

Acknowledgements and remarks

My thanks go primarily to my supervisor Professor Robin Cubitt at the University of Nottingham for his valuable and insightful guidance and advice on my research. I also express my gratitude to Sergio Sousa for his cooperation in sharing unpublished details of his paper with me, including the answer key.

In addition, I must also thank Tim Harford for graciously promoting my experiment on Twitter, and all of the 103 individuals who generously donated 20 minutes of their time to it and, I believe, proved that people do not always need to be incentivised to give well thought-out answers that can be of use to economists.

I strongly welcome comments and feedback on my research.

Appendices

Full datasets and copies of the instructions given to subjects can be downloaded from

http://markwainwright.com/dissertation.

References

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2006). Who is 'behavioral'? Cognitive ability and anomalous preferences.

Berg, J. E., Dickhaut, J. W., & Rietz, T. A. (2010). Preference reversals: The impact of truth-revealing monetary incentives. *Games and Economic Behavior*, 68(2), 443–468.

Cubitt, R. P., Munro, A., & Starmer, C. (2004). Testing explanations of preference reversal. *The Economic Journal*, 114(497), 709–726.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.

Grether, D. M. & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, 69(4), 623–638.

Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), 46–55.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics*, 9(2), 587–608.

Schkade, D. A. & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, 44(2), 203–231.

Schmidt, U. & Hey, J. D. (2004). Are preference reversals errors? An experimental investigation. *Journal of Risk and Uncertainty*, 29(3), 207–218.

Shiv, B. & Fedorikhin, A. (1999). Heart and mind in conflict: the interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, 26(3), 278–292.

Slovic, P. & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *The American Economic Review*, 73(4), 596–605.

Sousa, S. (2010). Are smarter people really less risk averse?

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 80(1), 204–217.

November 2010 - June 2011
This version published 6th June, 2011.