Preference Reversals: The Impact of Truth-Revealing Monetary Incentives. A Methodological Note.

Ferdinand M. Vieider*

Ludwig-Maximilians-University Munich, Germany

Climate Policy Initiative, DIW Berlin, Germany

05—June—2010

Abstract:

In a recent effort aimed at aggregating the existing evidence on the effect of financial incentives on preference reversals, Berg, Dickhaut and Rietz [Berg, Joyce E., John W. Dickhaut, & Thomas A. Rietz (2010). Preference Reversals: The Impact of Truth-Revealing Monetary Incentives. *Games and Economic Behavior*, 68(2), 443–486], claim that incentives designed to elicit true preferences under expected utility theory do reduce preference reversals. I reanalyze their data and show that their claim is not tenable when proper meta-analytic techniques are employed. Indeed, their results are found to derive from a number of statistical flaws, including a double count of some extreme effects deriving from a lack of statistical independence of supposedly independent observations, the co-variation and hence confound of important decision parameters with incentives, and several more. While their results thus need to be reconsidered, the method they propose holds great promise if executed properly. This note tries to raise the awareness of the technical issues involved, in order to encourage the use of state of the art meta-analytic techniques.

Keywords: preference reversal; uncertainty; risky choice; meta-analysis; methodology

JEL classifications: C91, D81, C19

* fvieider@gmail.com. The author is indebted to Marie Claire Villeval and Peter Wakker for helpful comments.

1. Introduction

In 1971, Sarah Lichtenstein and Paul Slovic published a seminal paper describing preference reversals between pairs of bets. They presented subjects with two bets, one providing a high probability of winning a relatively small prize (the *P-bet*), and one providing a low probability of winning a relatively large prize (the *$-bet*). When asked to choose between these two bets, a majority of subjects preferred the P-bet. When the same subjects where however subsequently asked to state their minimum selling price, or willingness to accept (*WTA*), for each of the two bets, most subjects expressed a preference for the $-bet by attributing it a higher price.

Preference reversals have since proven to be extremely resilient to attempts at debiasing by economists who tried to depict the finding as an artifact of psychological research. Grether & Plott (1979) famously changed several experimental aspects that they deemed suspect with the declared goal of uncovering the artificiality of these findings. The most important one of these changes consisted in the provision of high financial incentives. However, all their efforts proved fruitless—preference reversals persisted and their frequency was even slightly increased under real incentives. Even though ever-larger amounts of money were subsequently spent on disproving the phenomenon through the provision of significant financial incentives, the results were generally disappointing.

Recently, a paper by Berg, Dickhaut, & Rietz (2010; henceforth *BDR*) put forth an important new attempt at summarizing the power of monetary incentives to reduce the incidence of preference reversals, as well as several patterns of choice and pricing connected to preference reversals. The latter study brings new evidence to the table, inasmuch as it attempts to quantitatively aggregate and compare the effects of incentives *between studies* in an observational way. They find a marginally significant reduction of preference reversals induced by what they call "truth-revealing" monetary incentives, in addition to several other secondary findings on choice patterns and conditional preferences. They then proceed to examining different models, reaching the conclusion that under truth-revealing incentives the best fit is achieved by their noisy maximization model.

The present note concerns itself exclusively with the empirical findings of BDR. Aggregating findings from existing studies is a very promising technique, which in economics has not yet received the attention it deserves. While BDR's attempt at shedding light on the preference reversal phenomenon by aggregating existing evidence is laudable, I will argue that their statistical methods are flawed in several ways. Once these statistical issues are accounted for, no significant effects of incentives on preference reversals can be found. Furthermore, all but one of the other effects BDR find also lose their significance when proper meta-analytic techniques are applied. Finally, the one remaining significant effect can be shown to hinge on questionable choices in the construction of the index, so that it is not clear what that index is supposed to represent. These results indeed agree with the evidence from existing *within study* variations of incentives, which strongly suggest that monetary incentives—"truth-revealing" though they may be—are generally ineffective when it comes to reducing preference reversals.

The present note is structured as follows. Section 2 presents a brief introduction to the meta-analytic method. It then proceeds to discussing several flaws in BDR's methodology, each one of which taken by itself is sufficient to invalidate their main hypothesis—that "truth-revealing" incentives reduce the incidence of preference reversals. Section 3 employs the techniques discussed in section 2, and re-analyzes all the effects found by BDR. Section 5 contains a discussion of the issues that have ben raised. Section 6 concludes this short note.

2. Methodological Issues: Applying proper Meta-Analytic Techniques.

While this is neither the place nor the time for a complete exposition of the meta-analytic method, it seems imperative to quickly refer to the main statistical tools it applies and thus to describe the power it confers in aggregating findings from existing studies. Meta-analysis is widely used in medicine to aggregate effects from several randomized trials (Glass, 1976; Rosenthal & di Matteo, 2001). Although it is by now widely used in psychology, economists have only recently started to apply it (e.g. Engel, 2010; Murphy *et al.*, 2005; Oosterbeek, Sloof, & van de Kuilen, 2004). Given the extent of the literature existing on many issues in experimental economics, meta-analysis is an extremely powerful instrument that can be used to analyze and aggregate existing findings in order to resolve potential controversies. In this sense, meta-analysis may well be the single most promising innovation for the development of our understanding of human behavior as applied to economic issues.

The analysis of BDR is unfortunately biased by the fact that a number of statistical errors are committed in their data analysis. We will list these errors below, together with an illustration of the problems deriving from them and a quick discussion of possible solutions. Only after having discussed all the potential issues arising in their analysis will we present a re-analysis of their data using state of the art meta-analytic techniques. In doing this, we will concentrate on the comparison of hypothetical studies to the ones providing "truth-revealing" incentives (i.e., incentives designed to elicit true preferences under expected utility theory in BDR's definition), since this is the main comparison used by the authors. The purpose of this exercise is not so much to criticize BDR, who have undertaken the commendable task of aggregating findings from the literature in a quantitative way; the goal is much rather to make experimental economists aware of the potential pitfalls when aggregating studies, and even more to make them aware of the immense promise held by properly implemented meta-analytic techniques to shed light on obscure or controversial issues in our field.

2.1 Biased Sampling of Studies

The sample of studies BDR use for their investigation is biased. While I do not necessarily agree with their restrictive inclusion criteria (almost exact replications of the Lichtenstein & Slovic,1971, experiments), to a certain extent those are a matter of scope of the research question and hence a matter of preference. The reasons for their restrictive choice can however be questioned, all the more so since the study uses between study observations and would thus hugely benefit from a larger sample—indeed we will see that the small sample of BDR is one of their greatest problems. For instance, the exclusion of studies that used different elicitation methods seems odd, since such differences could be controlled for in a regression, and would thus bring important additional information to the table.

What is worse, the authors explicitly exclude studies that have not been published, thus introducing a potential publication bias (Rosenthal, 1991). This bias derives from the fact that studies with statistically significant results or confirming the authors' hypotheses are more likely to be published than insignificant results or results contradicting pre-existing hypotheses (see Dickersin, 2005, for a review). Excluding unpublished results may thus lead to a bias in overall findings (Sterne *et al.*, 2001). Running the Egger *et al.* (1997) test for publication bias results in a highly significant indication of publication bias ($p=0.006$). This test has however limited validity for between study analyses as the one by BDR, and is difficult to interpret. The shortcomings from the selection bias are thus best analyzed in the context of other statistical issues that we will now proceed to examine.

2.2 Violation of statistical independence.

One important statistical issue in BDR is that some of the effects they report are not statistically independent from each-other. This violation of statistical independence takes place on several different levels. The most consequential of these violations is that BDR include several effect sizes obtained from the same studies using the same subjects in their analysis. This effectively weights those studies double, and has the same effect as treating each decision obtained from a subject as independent in a statistical test, thus artificially increasing the degrees of freedom.

Figure 1 displays the proportion of preference reversals found by the different studies included in BDR, with the dashed line indicating the weighted (see below) mean preference reversal rate. As can be seen clearly in the graph, the supposed lowering of preference-reversal rates under truth-revealing incentives fund by BDR is entirely due to two papers—Berg *et al.* (1995) and Selten *et al.* (1999). The second sessions in the two Berg *et al.* (1985) experiments however use the same subjects as the first sessions. From a statistical point of view, those two effects are thus not independent from the session 1 results. Given that their session 2 studies are effectively reproducing the session 1 results obtained with the same subjects, these findings are counted double and can thus account in part for the incentive effect BDR claim to find. Indeed, if one introduces weighting to account for this lack in independence (Borenstein *et al.*, 2009), the supposed effect of incentives on preference reversals is no longer significant.

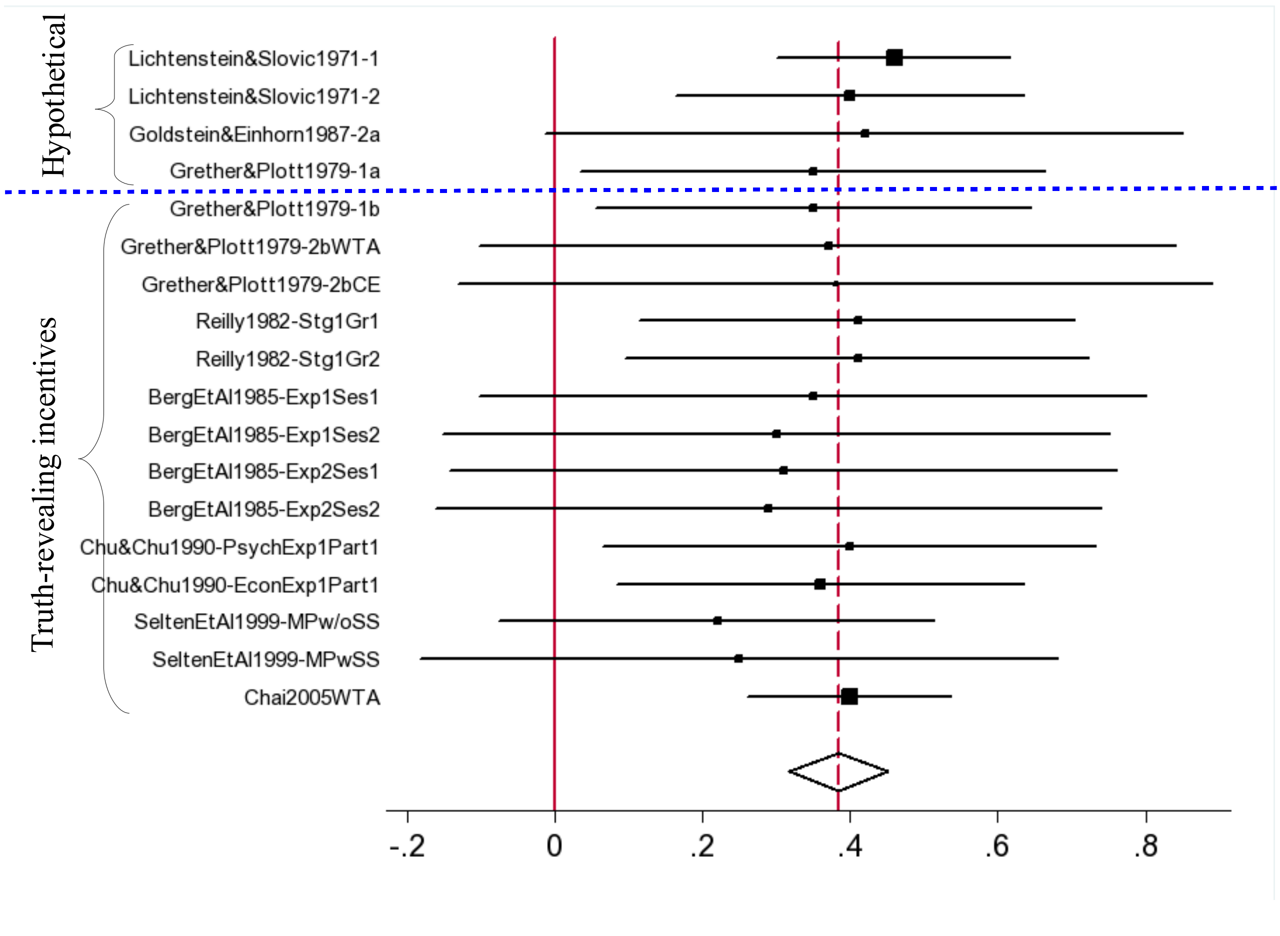

Fig. 1: Forest plot of the BDR studies: hypothetical and "truth-revealing"
Squares indicate the effect found by the given study, and the horizontal lines indicate the 95% confidence intervals around those effects. The rhombus and corresponding vertical dashed line indicate the aggregate effect size (preference reversal rate). The area of the square indicates the standard error of the study, and is used in meta-analysis to calculate weights (see below for a discussion).

2.3 Unobserved between-study variations

Within-study, experimental, tests of the effect of an independent variable are generally much more powerful than results from a between-study comparison such as the one used by BDR (Egger, Schneider, & Smith, 1998; Higgins & Thompson, 2004). One reason for this is that unobserved elements of procedure can generally be assumed to be constant within the same study. Between studies however such elements may easily change. While some such elements can and should be obtained from the methodological description of the study and can serve as moderator variables (see below), it is generally accepted that many changes between studies cannot be captured in this way. Egger *et al.* (1998) thus observe how between-study meta-analyses "yield estimates of association which may deviate from true underlying relationships beyond the play of chance. This may be due to the effects of confounding factors, the influence of biases, or both" (p. 141). It is thus generally seen as desirable to have a large number of studies when between-study effects are considered in order to filter out such chance deviations as much as possible. Unfortunately, BDR use but 18 effects for their between study comparison. Only four of these concern hypothetical choices. And not all of them are statistically independent.

The problems arising from this can again be illustrated by looking at figure 1. One can indeed see that most studies hover about the aggregate preference reversal rate. Notable exceptions to this rule are the first study by Lichtenstein & Slovic (1971), which displays a somewhat larger reversal rate; the four studies by Berg *et al.* (1985), and the two studies by Selten *et al.* (1999), which all display smaller reversal rates. This however shows us immediately how the reduction of preference reversals found for truth-revealing incentives is driven entirely by the results obtained by only two groups of researchers. And again, effect sizes obtained from one same research group cannot be seen as strictly independent from each-other (Borenstein *et al.*, 2009). This goes to show that inserting a simple control for research group in the regression—as a rough approximation for a better indicator of potential differences in experimental design or procedure—is sufficient to eliminate the effect of truth-revealing incentives claimed by BDR.

2.4 Adding moderator variables

We have seen above how different studies are likely to vary some elements of procedure or even some parameters of the decision problem. In order to control for this in a meta-analysis, an effort needs to be made to encode such variations through the inclusion of as many moderator variables as possible in the regression. Not including such moderators risks introducing confounds that could have been avoided, with the consequent impossibility of attributing any effects found clearly to one and only one factor—in this specific case, truth-revealing incentives. In the case of preference reversals, potential moderators include whether WTA was used as a pricing mechanism, or rather some other valuation mechanism like willingness to pay (highest buying prices) or certainty equivalents; whether the task was carried out on computer or using paper and pencil; whether the subject pool consisted of undergraduate students or some other specific population subgroup; what specific kind of mechanism was used to incentivize valuations; and any other elements of the experimental design or task that may influence behavior. Unfortunately, given the small sample of studies included by BDR, such moderator variables are unlikely to have enough variation to have much explanatory power. Nevertheless, an effort was made to encode at least some of these variables to be added to the BDR data (see appendix for the data table with added moderators).

Potentially extremely important moderator variables in the case of preference reversals are constituted by the parameters of the prospects amongst which subjects are called to choose. Indeed, the different probabilities of winning in the P-bets and $-bets, as well as the different prizes, and whether the lotteries used include losses or not, appear all to be very relevant in determining choice patterns. So are the expected values of the prospects, and the differences in terms of expected values of the two prospects in a choice-pair. Unfortunately, detailed data about these issues could not be gathered due to the fact that choice and pricing patterns for single bet pairs are not reported in the papers, but only average data. Average values over many different prospects are too rough an approximation to yield additional insights. Nevertheless, one may speculate that e.g. the particularly low average probability of winning offered by the P-bet in Berg *et al.* (1985) experiment 2 may influence choice patterns ($p=0.68$ on average against $p\approx0.90$ in all other studies). And similar speculations may apply to the studies by Selten *et al.* (1999), whose average probability of winning in the $-bet of $p=0.20$ is significantly lower than the average probability in all but one other study (see table A2 in the appendix). More about this will be said in the discussion.

2.5 Weighing studies with their inverse standard error

BDR do not weight the studies included in their regressions in any way. To see why this may be a problem, take again a look at figure 1. The area of the squares indicating effect sizes represents the weight attributed to a study, which derives from the number of subjects the result is based on. Given the absence of weighting in the BDR regression, big studies like the Lichtenstein and Slovic (1971) study 1 or the study by Chai (2005) are given the same weight as much smaller studies.

To see why this is problematic, we need to be aware of what we try to do when analyzing a phenomenon statistically. By looking at a sample of observations, we try to estimate the true population mean (if there is no *one* true mean but rather a distribution of true means, this does not affect the general validity of the argument). How precise our estimate will be depends on the sample size, with the sample mean converging to the population mean as N goes to infinity. However, the smaller our sample size, the larger the potential bias in the estimation of the population mean. Now imagine that we have two groups of studies we want to compare. If these groups contain only a few effect sizes which are each based on a small sample, then any biases deriving from the small samples may carry over to the general effects if this is not accounted for. For this reason, effect sizes are normally weighted by the inverse of the standard error when they are aggregated (Hedges & Olkin, 1985).

Such weighting can be used in the context of a fixed-effects model or a random effects model. In a fixed-effects model, the simple inverse of the standard error is used as a weight for the study. This method is used whenever one assumes that there is one and only one underlying true effect size that needs to be estimated. If however effect sizes may follow a certain distribution and only fall into an interval, or if our best estimates of the true effect size would fall into such an interval because there are some factors we cannot control for that are likely to influence our estimates, then it is best to use a random effects model. The latter relatively overweights effects from small samples and relatively underweights effects from large sample (Borenstein *et al.*, 2009).

Even though BDR claim a certain homogeneity of their sample, we have seen that heterogeneity is in fact introduced by the fact that some studies vary the method by which data are collected (paper versus computer), the elicitation method of selling prices and whether selling or buying prices are elicited, and the probabilities and prizes of the lotteries, in addition to many potential variations that are not apparent from the writing of the papers. In order to account for this, it is generally preferable to use a random effects model. Notice that for the data at hand this model is in fact closer to BDR's method than a fixed effects model would be, and that it will thus be more favorable to their results. We encoded the number of subjects who participated in each study and added it to the BDR data-set in order to allow for such weighting.

3. Reanalyzing the data applying state of the art meta-analysis

I will now re-analyze the data from BDR using the improvements discussed above, and hence applying state-of-the-art meta-analytic techniques. Notice however how all studies used by BDM are kept in the data set, so that the discussed issues of statistical independence are not yet accounted for in the following analysis. We will look at the effect of truth-revealing incentives on preference reversals, and on all the other dependent variables analyzed by BDR, to see if any effects may be significant when meta-analytic techniques are applied to them. Following BDR, no corrections for multiple testing are used in the results reported here. For matters of convenience, we will here follow their order of analysis in section 4 (4.1 to 4.3). Their original statics are reported in square brackets for easy comparability.

Effect on preference over bets

Meta-regressing[1] the percentage of P-bet choices on truth-revealing incentives, we find a non-significant effect with p=0.221 [BDR: p=0.0041]. Adding progressively other moderator variables such as *practice* (whether subjects had made some practice decisions before the ones reported), *WTA* (a dummy indicating whether willingness to accept was used for valuations, or something else), *computer* (a dummy indicating whether the experiment was computerized or executied using paper and pencil), or *info* (a dummy indicating whether additional information about expected values was provided) does not improve the results (p=0.549), and none of the moderators shows a significant effect. As to the average absolute difference between P-bet preference measures, a

1 The *metareg* command in Stata was used to conduct the analysis—see Harbord & Higgins (2004) for technical details.

regression analysis shows a non-significant result with p=0.161 [BDR: p=0.0257]. Adding moderator variables further increases this p-value to p=0.301.

Effects on reversal rates

Carrying out a meta-regression of preference reversal rates on truth-revealing incentives, we do not find a significant result at p=0.383 [BDR: p=0.0893]. Including moderator variables in the analysis results in p=0.668. This is indeed consistent with the conclusion of Camerer & Hogarth (1999), and with the (statistically much stronger) evidence from direct studies of incentive effects (in the truth-revealing category, Grether & Plott, 1979)—financial incentives do not affect the incidence of preference reversals.

Effects on conditional reversal rates

Meta-regressing the difference in reversal rates conditional on pricing on truth-revealing incentives, we find a significant effect of p=0.005 [BDR: 0.0058]. The effect remains highly significant if moderator variables are included. This effect is thus indeed strong enough so that it should be taken into serious consideration. To do this, let us first take a closer look at the dependent variable that has been found significant.

The dependent variable found to be significant above is the rate of preference reversals conditional on the two bets being priced differently. This is a composed index, so that we need to take a closer look at what this difference in conditional reversal rates actually is. Table 1 shows the typical way in which preference reversals are analyzed. The two elements the composed index above consists of are (1) c/(a+c), which indicates the proportion of preference reversals conditional on the P-bet being priced higher; and (2) b/(b+d), the rate of preference reversals conditional on the $-bet being priced higher. BDR then proceed to subtracting (2) from (1), thus coming up with the index found to be significant above. However, what is the meaning of this index?

	P-bet priced higher	$-bet priced higher
P-bet chosen	Cell a consistent	Cell b typical reversal
$-bet chosen	Cell c inverse reversal	Cell d consistent

Table 1: Possible combinations of choices and pricing: overview

Table 2 shows the means of the two conditional reversal rates that are used to construct the index as well as the index itself for hypothetical studies and studies using truth-revealing incentives (the complete data table with cell frequencies by study can be found in the appendix). While the reversal rate conditional on the $-bet being priced higher [b/(b+d)] does indeed decrease with truth-revealing incentives, the proportion of reversals conditional on the P-bet being priced higher [c/(a+c)] *increases* with the provision of truth-revealing incentives. Regressing these two conditional rates on incentives, however, none of these two effects is significant. Nevertheless, this begs for the question why the composition of these two opposing effects all of a sudden produces a significant effect of incentives.

	c/(a+c)	b/(b+d)	[c/(a+c)] – [b/(b+d)]
hypothetical	0.29	0.44	–0.15
trust-revealing incentives	0.40	0.32	0.09

Table 2: Changes of conditional reversal rates induced by incentives.

The answer lies in the particular aggregation technique used by BDR. The index of conditional reversals found significant above is constructed through the subtraction of b/(b+d) from c/(a+c). This results in a difference of the index of roughly 24 percentage points when comparing hypothetical to incentivized studies—an effect that is indeed highly significant. Notice however how this subtraction effectively ends up counting an opposing effect as an agreeing one. Indeed, if one looks at the two composing elements of the index, one can see how the 12 percentage points decrease in the reversal rate conditional on the $-bet being priced higher [b/(b+d)] when incentives are provided is almost exactly compensated by the 11 percentage points *increase* in reversal rates conditional on the P-bet being priced higher. The effect found is thus a mere artifact of the subtraction, which ends up counting an opposing effect as an agreeing one! Considering that BDR summed unconditional reversal rates, a sum would have been the correct procedure also for conditional reversal rates. The significant effect found is thus due purely to this mistake in the construction of the index.

4. Discussion

While the aggregation of existing studies is indeed a very promising avenue in economics, the data provided by BDR fail to support their claim that preference reversals are significantly reduced by the provision of trust-revealing incentives. This is due to several statistical errors they commit. The most consequential errors at the base of their mistaken conclusions can be found in two somewhat related issues—1) the excessive restrictions of the sample that resulted in the small sample size; and 2) the failure to weight direct experimental tests of the effect of incentives more heavily than more spurious between-study evidence.

While between-study aggregations of evidence are certainly worthwhile and can add significant new evidence to the existing literature, direct tests of a hypothesis in a perfectly controlled environment such as a laboratory experiment do always constitute a more reliable test. It is thus good practice to compare any evidence obtained from between-study aggregations of data to such experimental evidence. When the two bodies of evidence are not found to coincide, one needs to carefully consider the reasons why this may be so. If it is difficult to find such reasons, then the between study effect that has been found is likely to be caused by confounds that could not or have not been observed or controlled for in the meta-analysis.

The sample size issue is closely related to the latter point. Indeed, the larger the number of studies included in the meta-analysis, and the larger the number of research groups that have worked on a topic, the less likely it is that potential confounds due to the between-study nature of the analysis introduce systematic bias into the analysis. In addition to that, it is good practice to control for as many potential differences in studies as possible. For example, Engel (2010) conducted a large between-study meta-analysis on the dictator game, including data from 616 treatments stemming from over 100 different experiments. He also encoded a large number of potential moderator variables, which allowed him to both add evidence to the existing experimental literature on the moderator variables concerned, and to explain a large proportion of the variance in effect sizes. This goes to show that between-study aggregations of evidence can actually constitute an asset, since they will uncover evidence about the effect of moderator variables that cannot be seen from the experiments it aggregates in separation.

Overall, the fact that incentives—truth-revealing though they may be—do not show an effect on preference reversals is not all too surprising. This finding indeed agrees not only with the evidence from experiments testing this directly for the type of preference reversals studied by BDR, but for a larger class or preference reversals, including for instance framing effects (Kühberger *et al.*, 2002). Also, it may be explained by the fact that incentives tend to focus the decision maker's attention on the single decision problem at hand, while coherence between different tasks is not directly incentivized. In this sense, other cognitive motivators such as requiring subjects to justify their decisions may prove more fruitful in this context. Indeed, they have been found to work in context such as preference reversals between gain and loss frames in which incentives have proved ineffective (Takemura, 1994; Vieider, 2010). Testing such a mechanism may thus be a fruitful path for future research, since to my best knowledge no study of the effect of justification requirements on the occurrence of the classical type of preference reversals between P-bets and \$-bets exists to date.

While the presumed effect of incentives on preference reversals conditional on pricing has been shown to be the product of a mistake in the aggregation of the two conditional indices, the fact that the latter vary substantially in opposite directions under real incentives seems worth some additional considerations. Unfortunately, in the absence of additional data one cannot conclude with certainty what lies behind these changes. I would however speculate that it could be one of two issues, or a combination of the two. The first one is that the decision parameters in at least some of the studies providing incentives are different from the ones in the hypothetical studies—and particularly, that they provide different probabilities of winning in the P-bet and the $-bet. In the presence of probability weighting, this may easily lead to changes in overall choice and pricing patterns, and thence to difference in conditional reversal rates. Indeed, we have seen above that some *average* probabilities are quite different from the ones typically used in the hypothetical studies. The exact effect of this can however not be determined due to the absence of detailed data.

The second potential explanation lies in changes in risk attitudes provoked by the provision of real incentives. Indeed, monetary incentives and stakes have been found repeatedly to influence risk attitudes, generally by increasing risk aversion/decreasing risk seeking (Etchard-Vincent & L'Haridon, 2008; Holt & Laury, 2002). This may thus lead to differences in pricing and choice patterns, and thence to a shift in the patterns of preference reversals. Indeed, a reduction in risk seeking for small probability prospects caused by incentives (Kachelmeir & Shehata, 1992; Lefebvre *et al.*, 2009) may well cause the $-bet to be less over-priced, thus reducing the base for preference reversals conditional on the $-bet being priced higher. It is however important to underline once again how this does not produce a reduction in the proportion of preference reversals, but rather a shift from preference reversals conditional on pricing the $-bet higher to preference reversals conditional on pricing the P-bet higher. In the absence of more detailed data however, a precise answer whether this is indeed the correct explanation for the patterns found remains unfortunately elusive.

6. Conclusion

BDR have made a laudable attempt at summarizing the existing evidence on the effect of incentives on preference reversals. Unfortunately, the reliance on a flawed statistical methodology has led them to come up with conclusions that cannot be sustained based on the data they report. The problem in a sense is that they did not go far enough, by including as many studies as possible in their analysis and controlling for potential moderating factors. The hope is that by pointing out the potential pitfalls as well as the huge potential of aggregating the evidence from existing studies, the awareness and popularity of a methodology which may well be the single most promising tool to further our knowledge in the burgeoning field of behavioral analysis may be increased.

Appendix: BDR data

Reference	Inc	N	SE	learn	WTA	Info	comp	P-prob	$-prob	Choices:	Prices	AbsDiffP	Reversal	CCR-Pbet	CCR-$bet	CCR-diff	PCR-Pbet	PCR-$bet	PCR-diff
Lichtenstein&Slovic1971-1	0	173	0.08	0	1	0	0	0.895	0.335	0.51	0.12	0.39	0.46	0.83	0.06	0.77	0.27	0.48	-0.21
Lichtenstein&Slovic1971-2	0	74	0.12	1	0	0	0	0.895	0.335	0.53	0.39	0.14	0.4	0.51	0.27	0.24	0.33	0.44	-0.11
Goldstein&Einhorn1987-2a	0	23	0.22	0	1	0	0	0.910	0.330	0.52	0.31	0.21	0.42	0.61	0.21	0.4	0.33	0.45	-0.12
Grether&Plott1979-1a	0	44	0.16	0	1	0	0	0.910	0.330	0.49	0.26	0.23	0.35	0.59	0.11	0.48	0.22	0.39	-0.17
Grether&Plott1979-1b	1	46	0.15	0	1	0	0	0.910	0.330	0.36	0.18	0.18	0.35	0.73	0.13	0.6	0.46	0.32	0.14
Grether&Plott1979-2bWTA	1	20	0.24	0	1	0	0	0.910	0.330	0.38	0.25	0.13	0.37	0.66	0.19	0.47	0.48	0.33	0.15
Grether&Plott1979-2bCE	1	18	0.26	0	0	0	0	0.910	0.330	0.38	0.19	0.19	0.38	0.75	0.15	0.6	0.5	0.35	0.15
Reilly1982-Stg1Gr1	1	45	0.15	0	1	0	0	0.895	0.335	0.39	0.31	0.08	0.41	0.62	0.27	0.35	0.53	0.36	0.17
Reilly1982-Stg1Gr2	1	41	0.16	0	1	1	0	0.895	0.335	0.45	0.22	0.22	0.41	0.3	0.16	0.14	0.41	0.4	0.01
BergEtAl1985-Exp1Ses1	1	22	0.23	1	1	0	0	0.910	0.330	0.39	0.33	0.06	0.35	0.48	0.24	0.24	0.44	0.3	0.14
BergEtAl1985-Exp1Ses2	1	22	0.23	1	1	0	0	0.910	0.330	0.32	0.23	0.09	0.3	0.39	0.15	0.24	0.44	0.25	0.19
BergEtAl1985-Exp2Ses1	1	22	0.23	1	0	0	0	0.680	0.270	0.42	0.5	0.08	0.31	0.73	0.33	0.4	0.39	0.23	0.16
BergEtAl1985-Exp2Ses2	1	22	0.23	1	0	0	0	0.680	0.270	0.48	0.55	0.07	0.29	0.77	0.35	0.42	0.33	0.24	0.09
Chu&Chu1990-PsychExp1Part1	1	39	0.17	0	1	0	0	0.890	0.450	0.45	0.4	0.06	0.4	0.49	0.32	0.17	0.44	0.38	0.06
Chu&Chu1990-EconExp1Part1	1	55	0.14	0	1	0	0	0.890	0.450	0.4	0.34	0.06	0.36	0.48	0.25	0.23	0.43	0.32	0.11
SeltenEtAl1999-MPw/oSS	1	48	0.15	0	1	0	1	0.900	0.200	0.4	0.31	0.1	0.22	0.61	0.1	0.51	0.19	0.22	-0.03
SeltenEtAl1999-MPwSS	1	24	0.22	0	1	1	1	0.900	0.200	0.5	0.38	0.13	0.25	0.63	0.13	0.5	0.17	0.3	-0.13
Chai2005WTA	1	186	0.07	0	1	0	1	0.950	0.190	0.43	0.2	0.23	0.4	0.27	0.15	0.12	0.42	0.4	0.02

Table A1: BDR data with added moderators. The table reproduces the data from BDR's table 3, with the columns from "choices" to "PCR-diff" indicating their corresponding dependent variables. *Inc* is a dummy variable indicating the presence of truth-revealing incentives; *N* indicates the number of subjects on which the result is based, and *SE* the standard error; *learn* indicates whether learning rounds have taken place prior to the experiment the data are based on; *WTA* indicates whether selling prices have been elicited or some other mechanism has been used; *Info* indicates whether subjects were given additional information about expected values; *comp* indicates whether a computer was used rather than paper and pencil; *P-prob* and *$-prob* indicate the average probability of winning afforded by the bets included in the study for the P-bet and the $-bet respectively.

Reference	incentive	a	b	c	d	sum
Lichtenstein&Slovic1971-1	0	0.08	0.42	0.03	0.46	1
Lichtenstein&Slovic1971-2	0	0.26	0.27	0.13	0.34	1
Goldstein&Einhorn1987-2a	0	0.2	0.31	0.1	0.38	1
Grether&Plott1979-1a	0	0.2	0.29	0.06	0.45	1
Grether&Plott1979-1b	1	0.1	0.26	0.08	0.55	1
Grether&Plott1979-2bWTA	1	0.13	0.25	0.12	0.5	1
Grether&Plott1979-2bCE	1	0.09	0.29	0.09	0.53	1
Reilly1982-Stg1Gr1	1	0.15	0.24	0.17	0.44	1
Reilly1982-Stg1Gr2	1	0.13	0.31	0.09	0.46	1
BergEtAl1985-Exp1Ses1	1	0.19	0.2	0.15	0.47	1
BergEtAl1985-Exp1Ses2	1	0.13	0.19	0.1	0.58	1
BergEtAl1985-Exp2Ses1	1	0.31	0.11	0.19	0.39	1
BergEtAl1985-Exp2Ses2	1	0.37	0.11	0.18	0.34	1
Chu&Chu1990-PsychExp1Part1	1	0.22	0.23	0.17	0.37	1
Chu&Chu1990-EconExp1Part1	1	0.19	0.21	0.15	0.45	1
SeltenEtAl1999-MPw/oSS	1	0.25	0.16	0.06	0.54	1
SeltenEtAl1999-MPwSS	1	0.31	0.19	0.06	0.44	1
Chai2005WTA	1	0.11	0.32	0.08	0.48	1

Table A2: cell proportions. The table shows the proportions falling into the various cells described in table 1, for hypothetical studies and studies with trust-revealing incentives only. The data are taken from BDR's table 2, with the difference that I report proportions instead of number of choices, i.e. their numbers divided by N, the total number of choices.

References

Berg, Joyce E., John W. Dickhaut, & John R. O'Brian (1985). Preference Reversal and Arbitrage. In V. Smith (ed.), "Research in Experimental Economics", vol. 3.

Berg, Joyce E., John W. Dickhaut, & Thomas A. Rietz (2010). Preference Reversals: The Impact of Truth-Revealing Monetary Incentives. *Games and Economic Behavior*, 68(2), 443–486.

Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins, & Hannah R. Rothenstein (2009). Introduction to Meta-Analysis. Wiley & Sons, London.

Camerer, Colin F., & Robin M. Hogarth (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty, 19(1)*, 7-42.

Chai, Xiaoyong (2005). Cognitive Preference Reversal or Market Price Reversal? *Kyklos 58(2)*, 177–194.

Chu, Yun-Peng, & Ruey-Ling Chu (1990). The Substance of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note. *American Economic Review 80(4)*, 902–911.

Dickersin K. (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication Bias in Meta-Analysis—Prevention, Assessment and Adjustments. West Sussex (UK): John Wiley & Sons. 356 p.

Egger, Matthias, George Davey Smith, Martin Schneider, & Christoph Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal 315*, 629-634.

Egger, Matthias, Martin Schneider, & George Davey Smith (1998). Meta-analysis Spurious precision? Meta-analysis of observational studies. *British Medical Journal 316*, 140-144.

Engel, Christoph (2010). Dictator Games: A Meta Study. Preprints of the Max Planck Institute for Research on Collective Goods, Bonn 2010/7.

Etchart-Vincent, Nathalie, & Olivier l'Haridon (2008). Monetary Incentives in the Loss Domain and Behaviour toward Risk: An Experimental Comparison of Three Rewarding Schemes Including Real Losses. Working Paper

Fehr-Duda, Helga, Adrian Bruhin, Thomas F. Epper, & Renate Schubert (2010). Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size. Journal of Risk and Uncertainty 40(2), 147–180.

Glass, Gene V. (1976). Primary, Secondary, and Meta-Analysis of Research. Educational Researcher 5(10), 3-8.

Goldstein, William M., & Hillel J. Einhorn (1987). Expression Theory and the Preference Reversal Phenomena. *Psychological Review 94(2)*, 236-254.

Grether, David M., & Charles R. Plott (1979). Economic Theory of Choice and the Preference

Reversal Phenomenon. *The American Economic Review 69(4)*, 623-638.
Harbord, Roger, & Julian Higgins (2004). METAREG: Stata module to perform meta-analysis regression. Boston College Department of Economics. Available at: http://ideas.repec.org/s/boc/bocode.html
Hedges, Larry V., & Ingram Olkin (1985). Statistical Methods for Meta-Analysis. Academic Press Inc., London.
Higgins, J., & S.G. Thompson (2004). Controlling the Risk of Spurious Findings from Meta-regression. Statistics in Medicine 21, 1539–1558.
Kachelmeier, Steven J., & Mohamed Shehata (1992). Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China. *American Economic Review 82(5)*, 1120-1141.
Kühberger, Anton, Michael Schulte-Mecklenberg, & Josef Perner (2002). Framing Decisions: Hypothetical and Real. *Organizational Behavior and Human Decision Processes 89(2)*, 1162-1175.
Lefebvre, Mathieu, Ferdinand M. Vieider, & Marie Claire Villeval (2009). Incentive Effects on Risk Attitude in Small Probability Prospects. Working Paper, Laboratoire GATE, University of Lyon.
Lichtenstein, Sarah, & Paul Slovic (1971). Reversals of Preference Between Bids and Choices in Gambling Decisions. Journal of Experimental Psychology 89(1), 46-55.
Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, & Darryl Weatherhead (2005). A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics 30(3)*, 313-325.
Oosterbeek, Hessel, Randolph Sloof, & Gijs van de Kuilen (2004). Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics 7(2)*, 171-188.
Reilly, Robert J. (1982). Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design. *American Economic Review 72(3)*, 576-584.
Rosenthal, Robert (1991). Meta-Analytic Procedures for Social Research. Sage Publication, Newberry Park, CA.
Rosenthal, Robert, & M. R. DiMatteo (2001). Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology 52*, 59–82.
Selten, Rainer, A. Sadrieh, & Klaus Abbink (1999). Money does not induce risk neutral Behavior, but binary lotteries do even worse. *Theory and Decision 46(3)*, 211–249.
Sterne, Jonathan A. C., Matthias Egger, George Davey Smith (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British*

Medical Journal 323, 101-105.

Takemura, Kazuhisa (1994). Influence of Elaboration on the Framing of Decision. *Journal of Psychology 128(1)*, 33.

Vieider, Ferdinand M. (2010). Separating Real Incentives and Accountability. Working Paper, Laboratoir Gate, Université de Lyon.