EXPLAINING THE BDM—OR ANY RANDOM BINARY CHOICE ELICITATION MECHANISM—TO SUBJECTS[†]

PAUL J. HEALY[*]

January 8, 2020

CONTENTS

I. Introduction 1
II. Explaining the BDM Mechanism for Eliciting Values 2
III. Eliciting Cardinal Utilities 3
IV. Eliciting Beliefs About an Event E 4
V. Eliciting Risk Aversion via Holt-Laury 5
VI. Eliciting an Entire Preference Relation 6
VII. Understanding Monotonicity and Incentive Compatibility 7
VIII. Notes & Some Related Literature 8
References 9

I. INTRODUCTION

Experimenters often find the BDM mechanism confusing to explain to subjects, and consequently don't trust its use. It is often phrased as some sort of auction, or in terms of contingent payments based on some random numbers. But there's a simpler way of explaining it as a list of binary choices, which I believe is very intuitive. In fact, there is an entire class of elicitation mechanisms—which I call *random binary choice (RBC)* mechanisms—that are procedurally identical to the BDM mechanism and can be explained in the same way. As we show in Azrieli et al. (2018), they are all incentive compatible under the assumption of monotonicity, which simply says that subjects never choose dominated gambles. RBC mechanisms *do not* require expected utility preferences *unless* the subject also obeys the reduction of compound lotteries (or some similar axiom).[1] For evidence that this list presentation is effective, see Holt and Smith (2016).

[†]This comes from my work on epistemic game theory, entitled "Epistemic Game Theory Experiments: Utility Elicitation and Irrational Play". If you want to cite something, cite that. That paper previously circulated as two separate projects, "Epistemic Conditions for the Failure of Nash Equilibrium" and "Preferences, Rationality, and Belief Updating in Extensive-Form Games: Experimental Evidence".

[*]Dept. of Economics, The Ohio State University, Columbus, OH; healy.52@osu.edu.

[1]See Section VII and Azrieli et al. (2018) for details.

II. Explaining the BDM Mechanism for Eliciting Values

Suppose I want to run the BDM mechanism to find the subjects' values for a cheeseburger. The following (between the lines) is exactly what I would show to subjects:

I am going to ask you the following list of questions:

Q#		**Option A**		**Option B**
1	Would you rather have:	Cheeseburger	or	$0.01
2	Would you rather have:	Cheeseburger	or	$0.02
3	Would you rather have:	Cheeseburger	or	$0.03
⋮	⋮	⋮	⋮	⋮
1,999	Would you rather have:	Cheeseburger	or	$19.99
2,000	Would you rather have:	Cheeseburger	or	$20.00

In each question you pick either Option A (the cheeseburger) or Option B (the money). After you answer all 2,000 questions, I will randomly pick *one* question and pay you the option you chose on that one question. Each question is equally likely to be chosen for payment. Obviously you have no incentive to lie on any question, because if that question gets chosen for payment then you'd end up with the option you like less.

I assume you're going to choose the Option A in at least the first few questions, but at some point switch to choosing Option B. So, to save time, just tell me at which dollar value you'd switch. I can then 'fill out' your answers to all 2,000 questions based on your switch point (choosing Option A for all questions before your switch point, and Option B for all questions at or after your switch point). I'll still draw one question randomly for payment. Again, if you lie about your true switch point you might end up getting paid an option that you like less.

At which dollar value would you switch? ________________

The switch point they announce is their dollar value for the cheeseburger. That's all there is to it. I show why this is incentive compatible under fairly weak assumptions in Section VII.

The choice of endpoints ($0.01 and $20) is of course arbitrary; we just need the range to contain all possible valuations.

Past work on value elicitation has shown evidence of a WTP-WTA gap due to an endowment effect. Notice that the procedure given above does not endow the participant with anything; they simply choose between two options. Presumably this avoids any endowment effects.

III. Eliciting Cardinal Utilities

Suppose I assume the subject satisfies expected utility and I want to know their cardinal utility for a cheeseburger. If I set $u(\$0.00) = 0$ and $u(\$20.00) = 1$ (which is without loss under expected utility), then there is some p^* for which

$$u(\text{Cheeseburger}) = (1 - p^*)\underbrace{u(\$0.00)}_{=0} + p^*\underbrace{u(\$20.00)}_{=1} = p^*.$$

Thus, I can know their cardinal utility if I can find this indifference probability (often called the probability equivalent, as opposed to the certainty equivalent). And I can elicit that using the BDM procedure above, but with probabilities instead of certain dollar payments:

Q#		**Option A**		**Option B**
1	Would you rather have:	Cheeseburger	or	1% chance of \$20
2	Would you rather have:	Cheeseburger	or	2% chance of \$20
3	Would you rather have:	Cheeseburger	or	3% chance of \$20
⋮	⋮	⋮	⋮	⋮
99	Would you rather have:	Cheeseburger	or	99% chance of \$20
100	Would you rather have:	Cheeseburger	or	100% chance of \$20

The instructions are almost identical to the BDM mechanism. Again, they announce a switch point, I'll draw one question randomly, and they'll get paid based on their answer to the chosen question. If they satisfy expected utility, their switch point probability will be their $u(\text{Cheeseburger})$.

The choice of the \$20 prize is arbitrary. All that matters is that every subject prefer Option A in the first row and Option B in the last row.

Interpreting the switch point as a cardinal utility value obviously requires expected utility. But even if the subject doesn't satisfy expected utility, the elicited switch point will still be a meaningful measure of their 'value' for the cheeseburger. And it will be elicited truthfully as long as they obey monotonicity (see Section VII) and probabilistic sophistication (Machina and Schmeidler, 1992).

For the lotteries in Option B it would be ideal to use a physical randomizing device.

In practice cardinal utilities would probably be elicited for dollar amounts rather than cheeseburgers. That's fine; to elicit $u(\$10)$ simply replace every instance of "Cheeseburger" with "\$10.00". In a game, if I want the cardinal utility of outcome $(\$5, \$10)$ (meaning, \$5 for you, \$10 for your opponent), I would have Option A be $(\$5, \$10)$ and Option B be a $p\%$ chance of $(\$20, \$20)$. One player's elicitation is chosen for payment and the resulting payment is made to both players.

IV. ELICITING BELIEFS ABOUT AN EVENT E

Suppose you want to know the subject's belief that some event E will occur. You could use a scoring rule, but that requires risk neutral expected utility. Instead, use an RBC mechanism. This was first proposed by Grether (1981) (though it really comes straight out of Savage, 1954) and studied formally by Karni (2009). Again, it's just an RBC mechanism so we can run it like the BDM described previously. The questions you'd ask are as follows:

Q#		**Option A**		**Option B**
1	Would you rather have:	\$20 if E occurs	or	1% chance of \$20
2	Would you rather have:	\$20 if E occurs	or	2% chance of \$20
3	Would you rather have:	\$20 if E occurs	or	3% chance of \$20
⋮	⋮	⋮	⋮	⋮
99	Would you rather have:	\$20 if E occurs	or	99% chance of \$20
100	Would you rather have:	\$20 if E occurs	or	100% chance of \$20

Under probabilistic sophistication (Machina and Schmeidler, 1992), we can interpret the switch point p^* as as the subject's probability belief of E. Even if they don't satisfy probabilistic sophistication, the switch point will serve as a measure of the subjects 'likelihood' of event E. Under monotonicity (see Azrieli et al., 2018) it will be elicited truthfully.

An important caveat is that you need to be able to observe (and verify to the subjects) whether or not E occurs before the end of the experiment. I am not aware of a strictly incentive compatible way to elicit beliefs for events that cannot be observed and verified; I imagine it's impossible.

If you have multiple events $(E_1, E_2, \ldots, E_n)$, you can elicit the subject's probability for each, using the above procedure for each E_i. Then you randomly pick one $i \in \{1, \ldots, n\}$ to be paid, and pay for that elicitation as described above. This will still be incentive compatible under monotonicity. You may or may not restrict the subject's probabilities to sum to one; that's up to you.

You can even elicit second-order beliefs in a game using this mechanism. Or, at least, a coarsening of those beliefs. Let D_1 be the event that Player 1 plays 'Defect', and D_1^c its complement. Ask Player 2 their beliefs about D_1. Now, define B_2^{25} as the event that Player 2's elicited belief was between 0% and 25%, B_2^{50} as the event that Player 2's elicited belief was between 25% and 50%, and so on. Then elicit Player 1's belief over B_2^{25}, B_2^{50}, B_2^{75}, and B_2^{100}. The result is (a coarsening of) Player 1's second-order belief.[2]

[2]Really you'd want to elicit beliefs over each $B_2^i \times D_2$ and $B_2^i \times D_2^c$ to capture correlation in the hierarchy.

V. ELICITING RISK AVERSION VIA HOLT-LAURY

The well-known Holt-Laury method of eliciting risk preferences is just another example of an RBC mechanism:

Q#		Option A		Option B
1	Would you rather have:	(10%, \$2; 90%, \$1.60)	or	(10%, \$3.85; 90%, \$0.10)
2	Would you rather have:	(20%, \$2; 80%, \$1.60)	or	(20%, \$3.85; 80%, \$0.10)
3	Would you rather have:	(30%, \$2; 70%, \$1.60)	or	(30%, \$3.85; 70%, \$0.10)
⋮	⋮	⋮	⋮	⋮
9	Would you rather have:	(90%, \$2; 10%, \$1.60)	or	(90%, \$3.85; 10%, \$0.10)
10	Would you rather have:	(100%, \$2; 0%, \$1.60)	or	(100%, \$3.85; 0%, \$0.10)

Here, (10%, \$2; 90%, \$1.60) means a 10% chance of \$2.00 and a 90% chance of \$1.60, for example.

One row is chosen randomly for payment. You can have them answer every question or simply announce a switch point. The subject's risk attitude is then inferred from their switch point. Often a functional form is assumed for the subject's cardinal utility (such as CARA) and a parameter estimated from the switch point. But even if that functional form is incorrect the switch point will still be a meaningful measure of the subject's tolerance for risk. And it will be elicited truthfully as long as the subject obeys monotonicity (see Azrieli et al., 2018).

Obviously you can change the exact lotteries given, the number of rows, whether the probabilities vary or the dollar amounts vary, *et cetera*. Again, the Option B lotteries should be played out using a physical, objective randomizing device such as a Bingo cage.

VI. ELICITING AN ENTIRE PREFERENCE RELATION

Suppose there is a finite set of objects in $X = \{a, b, c, \ldots, m, n\}$ and you want the subject's preference relation over X.[3] A simple way to elicit that is to ask them to rank all objects in X, and then you pick two objects randomly and pay them the one they rank higher. This was done in Bateman et al. (2007) and Crockett and Oprea (2012), for example.

But that method is theoretically equivalent to the following RBC mechanism:

Q#		**Option A**		**Option B**
1	Would you rather have:	a	or	b
2	Would you rather have:	a	or	c
3	Would you rather have:	a	or	d
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-1$	Would you rather have:	a	or	n
n	Would you rather have:	b	or	c
$n+1$	Would you rather have:	b	or	d
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n-3$	Would you rather have:	b	or	n
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\frac{(n-1)n}{2}$	Would you rather have:	m	or	n

In other words, you ask them every possible pairwise comparison and pick one randomly for payment. Normally we'd ask for a switch point to reduce the number of decisions down to one, but this list has no natural ordering. Instead, having the subject provide one unified ranking does shrink their decision from $(n-1)n/2$ questions down to one. In practice that's how you'd use the mechanism. The point of this exercise is just to show that it is equivalent to an RBC, and therefore is incentive compatible under the exact same assumption (monotonicity) as all the other RBC mechanisms.

[3] If X is infinite you can use this procedure to elicit the preference relation over some finite grid on X.

VII. Understanding Monotonicity and Incentive Compatibility

All RBC mechanisms are incentive compatible under monotonicity. Here I provide some intuition for this claim; see Azrieli et al. (2018) for full details.

Suppose there are six binary questions of the form $\{a_i, b_i\}$ for $i \in \{1, \ldots, 6\}$. The subject announces their choice in each and the experimenter rolls a die to determine which is paid. Assume the subject's true preference is $a_i \succ b_i$ in each i. Now consider the (random) payment he gets if he announces truthfully, versus what he gets if he lies on the third question:

Question Chosen for Payment:	1	2	3	4	5	6
Payment if Truthful:	a_1	a_2	a_3	a_4	a_5	a_6
	$=$	$=$	$\curlyvee$	$=$	$=$	$=$
Payment if Lying on 3rd Question:	a_1	a_2	b_3	a_4	a_5	a_6

Here's the important point:

The random payment he gets by lying is <u>dominated</u> (state-by-state) by the random payment he gets when he tells the truth.

This is obviously true for any lie, not just ones with a single deviation. So as long as the subject doesn't prefer dominated random payments, he will never lie. This leads to the following definition and result from Azrieli et al. (2018). In this definition, a random payment is an 'act' f that maps the chosen question (denoted by ω) into a choice object.

Definition 1 (Dominance & Monotonicity)**.** Say that act f *dominates* act g (according to $\succeq$) if $f(\omega) \succeq g(\omega)$ at every ω. A subject satisfies *monotonicity* if 'f dominates g' implies that 'f is preferred to g'. (Note: Nothing is assumed if f does not dominate g.)

Theorem 1. If a subject satisfies monotonicity then any RBC mechanism will be incentive compatible (the subject will report truthfully).

Notice: *Incentive compatibility does not require expected utility.* But there is a caveat: If the choice objects are lotteries then monotonicity combined with the reduction of compound lotteries implies expected utility preferences over the space of choice objects. Thus, if we do not trust expected utility *but do trust reduction*, then we should not trust the RBC mechanisms. Similarly, if choice objects are (possibly ambiguous) acts then monotonicity combined with the 'order reversal' axiom of Anscombe and Aumann (1963) implies ambiguity neutrality. Thus, if we believe subjects are ambiguity averse but we believe they satisfy order reversal, then we should not trust the RBC mechanisms. See Azrieli et al. (2018) for a more complete discussion of this issue.

VIII. Notes & Some Related Literature

In theory it doesn't matter what distribution you use to draw the question for payment, as long as each question has positive probability of being drawn (or, more generally, is non-null in the sense of Savage, 1954). In practice you want the subject to believe each question might matter, so the uniform distribution seems like the best (and simplest) choice. Alternatively, you may choose a distribution that puts more probability on the questions that are more likely to be at the subjects' switch points, since the subject will be nearly indifferent on those questions. The increased probability will offset that indifference. But such a non-uniform distribution will probably be more complicated to explain.

In a belief elicitation setting, Holt and Smith (2016) compare performance of the RBC mechanism presented as a list versus the same mechanism presented as depending on the realization of a randomly-drawn probability. Subjects report posterior beliefs upon observing a signal, and those reported posteriors are compared to the Bayesian posterior. Deviations from the Bayesian posterior are smallest with the list presentation, though both presentations outperform the (binarized) quadratic scoring rule.

In Brown and Healy (2018) we show that subjects do not report truthfully when we present the RBC as one list on a single screen, but do report truthfully when each row is given on a separate screen and in a random order. We argue that the list causes subjects to view all the choices as one large choice rather than separate choices, and this causes them to report a different answer. In most applications of the RBC mechanisms, however, asking each row separately is simply not practical. Given the results of Holt and Smith (2016), I would just stick with the list presentation. Even in Brown and Healy (2018) the chosen row ended up being an unbiased estimate of the subject's true preferences, so the failure of incentive compatibility caused no real problem in that particular experiment. It's just that Brown and Healy (2018) show it *might* be a problem in other experiments.

Some experimenters will randomly select one subject for payment. This is also incentive compatible under monotonicity, so using an RBC mechanism and paying one randomly-selected subject is incentive compatible if and only if the RBC mechanism is incentive compatible when all subjects are paid.[4] In other words, there is no loss (in theory) to using this procedure. In practice, however, it does 'dilute' the probability with which each decision is chosen, which may alter behavior.[5]

[4] There is one subtle caveat: Paying one randomly-selected subject alters the state space of the randomization device because it now must select a question number and a subject number. I assume here that if monotonicity holds for one state space, it holds for the other.

[5] If behavior is altered then either monotonicity is violated or actual behavior does not conform to a fixed preference relation. The latter occurs if choice is stochastic, for example.

The BDM mechanism comes from Becker et al. (1964). Cardinal utility elicitation was first done by Mosteller and Nogee (1951), though they paid for every decision. I'm not sure who was the first to elicit cardinal utilities using an RBC mechanism. To my knowledge, belief elicitation using an RBC mechanism was first done by Grether (1981), and independently discovered by several subsequent authors, including Holt (1986), Mobius et al. (2013), and Karni (2009). But it really just comes straight out of Savage (1954), who also mentions the idea of paying for one randomly-selected question and attributes it to Wallis and Allais. The risk aversion elicitation comes from Holt and Laury (2002), though there are many other methods suggested, some of which are also equivalent to (variations of) an RBC mechanism.

There are several papers offering tests and critiques of mechanisms that pay for one randomly-chosen decision. These include Starmer and Sugden (1991), Cubitt et al. (1998), Cox et al. (2014b,a), Harrison and Swarthout (2014), Baillon et al. (2014), and Brown and Healy (2018). A full review of this literature is far beyond the scope of this note.

References

Anscombe, F. J., Aumann, R. J., 1963. A definition of subjective probability. Annals of mathematical statistics, 199–205.

Azrieli, Y., Chambers, C. P., Healy, P. J., 2018. Incentives in experiments: A theoretical analysis. Journal of Political Economy 126 (4), 1472–1503.

Baillon, A., Halevy, Y., Li, C., 2014. Experimental elictation of ambiguity attitude using the random incentive system, University of British Columbia working paper.

Bateman, I., Day, B., Loomes, G., Sugden, R., 2007. Can ranking techniques elicit robust values? Journal of Risk & Uncertainty 34, 49–66.

Becker, G. M., DeGroot, M. H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behavioral Science 9, 226–232.

Brown, A., Healy, P. J., 2018. Separated decisions. European Economic Review 101, 20–34.

Cox, J., Sadiraj, V., Schmidt, U., 2014a. Alternative payoff mechanisms for choice under risk. International Advances in Economics Research 20 (3), 239–240.

Cox, J., Sadiraj, V., Schmidt, U., 2014b. Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms. PLOSone 9 (3), e90742.

Crockett, S., Oprea, R., 2012. In the long run we all trade: Reference dependence in dynamic economies, University of California Santa Barbara Working Paper.

Cubitt, R. P., Starmer, C., Sugden, R., 1998. On the validity of the random lottery incentive system. Experimental Economics 1, 115–131.

Grether, D. M., 1981. Financial incentive effects and individual decisionmaking, California Institute of Technology Social Science working paper 401.

Harrison, G. W., Swarthout, J. T., 2014. Experimental payment protocols and the bipolar behaviorist. Theory and Decision 77, 423–438.

Holt, C. A., 1986. Preference reversals and the independence axiom. American Economic Review 76, 508–515.

Holt, C. A., Laury, S. K., December 2002. Risk aversion and incentive effects. American Economic Review 92 (5), 1644–1655.

Holt, C. A., Smith, A. M., 2016. Belief elicitation with a synchronized lottery choice menu that is invariant to risk attitudes. American Economic Journal: Microeconomics 8 (1), 110–39.

Karni, E., 2009. A mechanism for eliciting probabilities. Econometrica 77 (2), 603–606.

Machina, M. J., Schmeidler, D., 1992. A more robust definition of subjective probability. Econometrica, 745–780.

Mobius, M. M., Niederle, M., Niehaus, P., Rosenblat, T., 2013. Managing self-confidence: Theory and experimental evidence, Stanford Univesrity working paper.

Mosteller, F., Nogee, P., 1951. An experimental measurement of utility. Journal of Political Economy 59, 371–401.

Savage, L. J., 1954. The Foundations of Statistics. John Wiley & Sons, New York, NY.

Starmer, C., Sugden, R., 1991. Does the random-lottery incentive system elicit true preferences? an experimental investigation. American Economic Review 81, 971–978.