cbess

Centre for Behavioural and
Experimental Social Science

UEA

University of East Anglia

CORRELATION NEGLECT AND CASE-BASED DECISIONS

Benjamin Radoc

Robert Sugden

Theodore L. Turocy

Correlation neglect and case-based decisions[*]

Benjamin Radoc

Economics Department

Ateneo de Manila University

Quezon City 1108, Philippines

Robert Sugden

School of Economics and CBESS

University of East Anglia

Norwich, United Kingdom NR4 7TJ

Theodore L. Turocy[†]

School of Economics and CBESS

University of East Anglia

Norwich, United Kingdom NR4 7TJ

November 27, 2017

Abstract

We test the conjecture that case-based reasoning may be used, not only when state spaces are undefined, but also when relevant probabilities can be derived by Bayesian inference from observations of random processes. Our experiment elicits participants' valuations of a lottery after observing realisations of this and another lottery. Depending on the treatment, participants know that the two lotteries are independent, positively correlated, or negatively correlated. We find no evidence of correlation neglect indicative of case-based reasoning. However, in the negative correlation treatment, valuations cannot be explained by Bayesian reasoning, while stated qualitative judgements about chances of winning can.

Keywords: case-based decision, correlation, experiment

JEL Classifications: C91, D03, D81

1 Introduction

Most theories of choice under uncertainty that have been proposed by economists or decision theorists are closely related to expected utility theory, and often are generalisations of that theory. (For one survey, see Sugden, 2004.) In these theories, uncertainty is represented by a set of states of the world, any one of which might obtain. Alternative acts available to an agent are represented as different assignments of consequences to states. Decision-making is conceptualised as a process of evaluating acts in terms of the subjective values that the agent attributes to their consequences and the probabilities or subjective weights that he or she assigns to the events in which those consequences occur.

However, case-based decision theory (CBDT), proposed by Gilboa and Schmeidler (1995), is based on a fundamentally different representation of decision problems. In CBDT, there is no state space and no concept of probability. The agent is not assumed to know anything about the outside world except what he has actually experienced as the results of previous decision making. The agent uses neither forward-looking hypothetical reasoning ("What will happen if I choose X?") nor backward-looking counterfactual reasoning ("What would have happened if I had chosen X?"). He simply evaluates each currently available act in terms of the consequences he has in fact experienced as a result of choosing that act (or in some variants of the theory, choosing similar acts) in previous decision problems that he perceives to be similar to the problem at hand.

Gilboa and Schmeidler (1995, pp. 606, 622) present CBDT and expected utility theory (EUT) as "complementary theories". They argue that CBDT is normatively most defensible and descriptively most plausible when "states of the world are neither naturally given, nor can they be simply formulated". Decision making in such circumstances is decision under *ignorance*, as contrasted with decision under *risk* (i.e. with known probabilities) and decision under *uncertainty* (i.e. with known states of the world but unknown probabilities). For decision under ignorance, Gilboa and Schmeidler argue, "the very language of expected utility models is inappropriate". If this complementarity claim were taken at face value, any idea of testing CBDT and EUT *against one another* would be out of place.

Up to now, most experimental tests of CBDT have been designed on the premise that CBDT and EUT are complementary theories. For example, Ossadnik et al. (2013) set up an experimental environment of "structural ignorance" (p. 212), and compare the explanatory power of CBDT with that of three alternative criteria for decision under ignorance – maximin, maximax, and the pessimism-optimism criterion of Hurwicz (1951). Similarly, Grosskopf et al. (2015) start from the explicit premise that CBDT "is not proposed as an alternative to or a generalization of [EUT]" (p. 640), and use an experimental environment in which "EUT is not a reasonable alternative decision making procedure" (p. 652). They test CBDT against the null hypothesis of random

choice and against a very simple heuristic, related to the "Take the Best" algorithm of Gigerenzer and Goldstein (1996). Unlike Ossadnik et al. and Grosskopf et al., who test specific parameterised forms of CBDT, Bleichrodt et al. (2015) report experimental tests of predictions derived from a very general, non-parameterised version of CBDT. But they too presuppose that CBDT is intended to be applied to situations in which states cannot be specified (p. 2), and design their experiment accordingly.[1]

In contrast, the starting point for our experimental research was the conjecture that CBDT might have predictive power in situations in which states are well-defined and objective prior probabilities are known, but expected-utility decision making requires the construction of posterior probabilities by Bayesian inference from observations of random processes. In such situations, the "correctness" of Bayesian reasoning about probabilities is uncontroversial. Nevertheless, such reasoning can be cognitively demanding and its implications can be counter-intuitive. It is well-known that human judgements about probability often contravene Bayesian principles in predictable ways, for example because of the use of availability and representativeness heuristics. (Tversky and Kahneman, 1973, 1983) Charness and Levin (2005) report evidence of deviations from Bayesian reasoning that are consistent with one of the simplest reinforcement learning rules, the win-stay-lose-shift heuristic. Given that case-based reasoning requires much less cognitive sophistication and is well-adapted to naturally-occurring problems of decision under ignorance, the hypothesis that human beings are predisposed to use it is psychologically plausible and worthy of investigation.

Since our conjecture has not been endorsed by the proposers of CBDT, we cannot structure our enquiry as a test of that theory. Our methodological strategy is to test predictions *of EUT* in situations in which there are intuitive reasons, derived from the underlying principles of CBDT, for expecting those predictions to fail in specific ways. This general strategy, used in combination with disparate intuitions, has led to many important developments in decision theory. The Allais paradox, common ratio effect, Ellsberg paradox, and preference reversal phenomenon were all first discovered by researchers who recognised potential limitations of EUT but who, at the time of discovery, were not in a position to propose a comprehensive alternative theory. These robust violations of EUT achieved the status of "exhibits" which informed the subsequent development of alternative decision theories.[2] These exhibits show non-random patterns in deviations of actual behaviour from the predictions of EUT. (For example, the Allais paradox involves comparisons between responses to two binary choice problems that EUT treats as equivalent. The lotteries in one of these problems stochastically dominate the lotteries in the other. Individuals' choices are

---

[1]Guilfoos and Pape (2016) report an experimental test of CBDT which uses a very different methodology. They compare actual behaviour in a repeated Prisoner's Dilemma experiment with simulated data created by a "software agent" programmed to use case-based reasoning.

[2]This methodological strategy, and its role in the development of experimental economics, is discussed by Bardsley et al. (2010, Chapter 4).

systematically more risk-averse in the first problem than in the second.) Such an exhibit provides evidence that some non-random mechanism, not encompassed by EUT, is at work, but is not to be interpreted as confirming any fully-specified alternative theory. Our research was designed to have the potential to create exhibits of this kind which might inform the development and application of CBDT.

Our experiment tests two related intuitions about how case-based reasoning might lead to systematic deviations from the behaviour predicted by EUT. The first of these intuitions derives from a fundamental property of CBDT – *act separability*. In CBDT, experiences are encoded in memory as *cases*; each case consists of a problem, the act that was chosen in that problem, and the result of that choice (measured in utility units). Given a new problem, a decision-maker assesses each available act by recalling the previous cases in which that act was chosen, and weighting the result of each of those choices by a measure of the similarity between the problem in which that choice was made and the new problem. This algorithm requires that the agent maintains a separate memory for each act. Each of these memories needs to store only results that have actually been experienced as a result of the choice of the relevant act. If memory is organised and used in this way, there is no way to retrieve information that could show positive or negative correlation between the results of different acts. Thus, one might expect case-based reasoners to differ from Bayesian reasoners by neglecting information about correlation.

The second intuition derives from the fact that probability judgements have no role in CBDT. Case-based reasoning does not lead to the formation of probability judgements that can be classified as "correct" or "incorrect" according to Bayesian principles. Instead, by moving directly from memory (encoded without reference to states or probabilities) to decisions, it circumvents the whole process of forming probability judgements. When an agent's case-based reasoning leads to a violation of EUT, an outside observer may be able to conclude that the agent has behaved *as if* she were trying to maximise expected utility but had made erroneous probability judgements, but the agent herself may have no perception of making or endorsing the judgments that the observer attributes to her. This raises the possibility that agents might in fact endorse probability judgements that are systematically different from those that are revealed in their decisions when those decisions are analysed in the theoretical framework used by EUT. It is of course debatable whether such a difference can properly be called a violation of EUT. (Opinions differ about whether "probability" in EUT refers to an agent's actual beliefs, or is merely part of a formal representation of her decision-making behaviour.) But a pattern of predictable differences between stated and revealed probabilities would be a surprising phenomenon calling for explanation.

These two lines of investigation have the potential to be mutually corroborating. Suppose that, in some experiment, participants' decisions are found to be insensitive to variations in relevant information about correlation. CBDT would offer a possible explanation for that observation. How-

ever, another possibility might be that the participants were Bayesian reasoners who had misunderstood the information given to them, perhaps because of weaknesses in the experimental design. But suppose it were also found that participants' stated probabilities showed Bayesian sensitivity to information about correlation. That would suggest that participants understood that information, used it when forming probability judgements, but failed to use in when making decisions. That would give additional credence to CBDT's explanation of correlation neglect in decisions.

The remainder of the paper is organised as follows. Section 2 describes act separability under CBDT, and Section 3 discusses stated and revealed probabilities under EUT. Section 4 presents the experimental design; Section 5 discusses the hypotheses, and Section 6 presents the results. Section 7 concludes with a further discussion.

2   Act separability and correlation neglect

In the core version of CBDT presented by Gilboa and Schmeidler (1995), the primitives are *problems*, *acts* and *results*. A problem is interpreted as a description of a unique decision situation in which the agent faces a non-empty set of acts, one of which must be chosen. The choice of a particular act in a particular problem leads to a result. A *case* is a triple (q, a, r) where q is a problem faced, a is the act chosen in that problem, and r is the result of that choice. The agent's memory M is a set of cases, interpreted as those cases that he has in fact experienced. There is a function $u(\cdot)$ which assigns a real-valued *utility* measure $u(r)$ to every possible result r. A utility value of zero is interpreted as the agent's aspiration level. There is a function $s(\cdot, \cdot)$ which assigns a real-valued non-negative *similarity* coefficient $s(p, q)$ to every ordered pair of problems (p, q). This coefficient is interpreted as a measure of the similarity of q to p, viewed from p. The theory is concerned with the behaviour of an agent who faces a new problem p, given a memory M. The agent chooses the available act a that maximises

$$U(a) = \sum_{(q,a,r) \in M} s(p, q)u(r),$$

where the summation over the empty set is treated as having a value of zero.

Now consider how this model deals with a simple case involving assets with returns that are potentially correlated. Suppose there are two lotteries L_1 and L_2. The agent faces a sequence of decision problems; in each problem, he faces one of the two lotteries and chooses whether or not to play it. Each time he plays either of these lotteries, the result is either *hit*, with constant utility $u_H > 0$ or *miss*, with constant utility $u_M < 0$. Not betting leads to a utility of zero. If he bets, he immediately learns whether the outcome was hit or miss. Consider an agent who has faced at least

5

$2n$ such prior problems (where $n \geq 1$), and has chosen to play each lottery exactly n times.[3] He has experienced h_1 hits from L_1 and h_2 hits from L_2. As a final problem, he faces some lottery L_i, $i \in \{1, 2\}$, and has to choose whether or not to play it.

A natural CBDT representation of this situation (*Model 1*) would treat "play L_1" and "play L_2" as distinct acts a_1 and a_2. (Since not playing either lottery always gives zero utility, it makes no difference whether that option is represented as one act or two.) It would also be natural to assume that there is some parameter $\sigma > 0$ such that, for $i \in \{1, 2\}$ and for all problems p_i, q_i in which the choice is between playing or not playing L_i, $s(p_i, q_i) = \sigma$. Then whether the agent chooses to bet in the final problem is determined by the sign of

$$U(a_i) = \sigma[h_i u_H + (n - h_i u_M)]$$

Notice that $U(a_i)$ is independent of h_j for $j \neq i$, and that $\partial U(a_i)/\partial h_i = \sigma(u_H - u_M) > 0$. Thus, the final choice is determined solely by the number of hits that have been experienced on whichever lottery is faced in the final problem.

An alternative representation (*Model 2*) would treat "play some lottery" as a single act and model the difference between L_1 and L_2 in terms of a difference in similarity.[4] Assume there are parameters σ, σ', with $\sigma > 0$ and $\sigma \geq \sigma' \geq 0$, such that for all problems p_i, q_i, q_j, where $i \neq j$, $s(p_i, q_i) = \sigma$ and $s(p_i, q_j) = \sigma'$. Then whether the agent chooses to bet in the final problem is determined by the sign of

$$U(a_i) = \sigma[h_i u_H + (n - h_i u_M)] + \sigma'[h_j u_H + (n - h_j u_M)]$$

For any distinct i and j, $\partial U(a_i)/\partial h_i = \sigma[u_H - u_M] > 0$ and $\sigma(u_H - u_M) \geq \partial U(a_i)/\partial h_j = \sigma'(u_H - u_M) \geq 0$. Thus, the final choice is potentially determined by the numbers of hits that have been experienced on both lotteries. Hits on the lottery that is faced in the final problem have a strictly positive weight that is strictly greater than the weight for hits on the other lottery; the latter weight may be zero but cannot be negative.

Notice that, in deriving the conditions under which the agent chooses to bet in the final problem, we have made no assumptions about correlation between the two lotteries. In Model 1, the case-based reasoner's final choice depends only on his experience of the lottery he is actually facing. This is compatible with Bayesian reasoning only in the case of a Bayesian agent with a prior belief

---

[3] By assuming that the agent's memory contains the same number of cases of choosing each lottery, we avoid the need to distinguish between the core version of CBDT and the "average performance" variant in which $U(a)$ is a similarity-weighted *average* (rather than *sum*) of experienced utilities (Gilboa and Schmeidler, 1995, pp. 619–620).

[4] Model 2 represents this difference in terms of similarity between *problems involving L_1* and *problems involving L_2*. Because these sets of problems are disjoint, this is equivalent to treating "bet on L_1" and "bet on L_2" as distinct acts and then, as in the Gilboa and Schmeidler (1995, pp. 635–638) "alternative model", defining similarity as a relationship between (problem, chosen act) pairs.

that the two lotteries are independent. In Model 2, the case-based reasoner's final choice may also be influenced by his experience of the other lottery, but only in the direction that corresponds with a Bayesian belief in *positive* correlation, and only to the extent that the two lotteries are perceived as subjectively similar. To assume that similarity judgements systematically incorporate prior beliefs about correlation would be inconsistent with one of the fundamental principles of CBDT – that decision-makers use only knowledge that is derived from direct experience. This thought points to a class of situations in which the underlying intuitions of CBDT suggest that there might be systematic violations of EUT. These are situations in which there is a mismatch between salient similarity cues and information about correlation that is not embedded in experiences of decision making. Our experimental design attempts to create such situations.

It is already known that people have a tendency to neglect asset correlation, but the possible connection between this tendency and CBDT has not been explored. For example, in an asset allocation experiment, Kallir and Sonsino (2009) found that participants focused their attention on individual asset returns and that the resulting portfolio decisions did not take into account return correlations. Eyster and Weizsäcker (2011) found that even when equipped with correlation information, participants regarded assets independently and resorted to the *1/n heuristic* (or naïve diversification) when allocating investment funds to individual securities. Similarly, in a hypothetical investment choice experiment, Hedesström et al. (2006) observed that participants focused on individual asset volatility rather than on portfolio volatility. Resulting portfolios were inappropriately diversified and had higher volatility. Correlation neglect has been found to be sensitive to the magnitude of the stakes involved in the decisions. In portfolio experiments with low stakes, Kroll et al. (1988) found that while participants were aware of the correlation in stock returns, correlation information was not reflected in their portfolio choices. However, when the stakes were significantly increased, participants managed to effectively diversify their asset holdings and the resulting portfolio choices were closer to the predictions of mean-variance optimisation. In contrast to the experiments we have just described, our experiment uses a design that allows us to investigate attitudes to similarity while controlling and manipulating similarity cues.

3   Stated probabilities and revealed probabilities

EUT legitimates a simple *certainty equivalence procedure* for eliciting an agent's subjective ranking of the probabilities of two events. Consider any two (non-null) events E_1 and E_2. Consider consequences that are measured in money units, and assume that larger consequences are always preferred to smaller ones. Choose any two consequences x and y such that $x > y$. Let xE_iy ($i \in \{1, 2\}$) denote the act that gives x if E_i obtains and y otherwise. Given a suitable continuity assumption, EUT implies that there exist $z_1, z_2 \in (x, y)$ such that the agent has the preferences

$z_1 \sim xE_1y$ and $z_2 \sim xE_2y$. (We use $\sim$ to denote indifference.) Then the subjective probability of E_1 is greater than (equal to, less than) that of E_2 if and only if z_1 is greater than (equal to, less than) z_2. Since this derivation of probability rankings does not depend on either the independence or transitivity axioms of EUT, it remains valid for almost all recognised forms of non-expected utility theory. Variants of this procedure are widely used in experimental economics to elicit probability judgements.

However, CBDT does *not* legitimate the certainty equivalence procedure. Since CBDT explains individuals' decisions without making any reference to events or probabilities, there is no way of using that theory to read off the probabilities of events from observations of decisions. Indeed, the logic of CBDT suggests that the idea of trying to infer attitudes to events from decisions is misguided. Notice that, if EUT holds, we can use *any* pair of non-indifferent consequences to elicit an agent's probability ranking of E_1 and E_2, and the resulting ranking will be the same. In this sense, the certainty equivalence procedure elicits attitudes to events that are independent of the decision problems in which those attitudes are elicited. But this need not be true for a CBDT agent. For such an agent, indifference between two acts in a particular decision problem is determined by the agent's memory of cases that were similar to that problem; if the problem is changed, the relevantly similar cases can change too.

Once one recognises that there are theories of choice under uncertainty that make no use of probabilities, it becomes a significant research question to ask whether individuals' stated probability judgements are the same as the *revealed* probabilities that are elicited by the certainty equivalence procedure. Viewed within the conceptual framework of EUT, systematic inconsistencies between stated and revealed probabilities would be surprising, and would raise doubts about the construct validity of the concept of probability used in EUT. There would be particular reason for this kind of doubt if, in a setting in which correct probabilities could be formed by Bayesian reasoning, stated probabilities had Bayesian properties but revealed probabilities did not.

Our experiment is designed to detect instances of correlation neglect that might be induced by case-based reasoning. Such behaviour, were it to occur, would be picked up as non-Bayesian properties of revealed probabilities. As an additional diagnostic tool, we investigate how far revealed probabilities are consistent with stated probabilities.

In experimental economics, it is standard practice to incentivise survey questions by reformulating them as decision problems with material (usually monetary) consequences. In the present case, however, the whole point of the enquiry is to discover whether stated probabilities differ from those that are revealed in decision problems. We believe that this is one of the class of "significant problems in economics that appear to be capable of experimental investigation only in nonincentivised designs" identified by Bardsley et al. (2010, pp. 336–337). The logic of our investigation requires that the elicitation of stated probabilities is *not* incentivised.

4 Experimental design

Our experiment used a set-up similar to that analysed in Section 2. In designing the experimental interface,[5] we tried to set a level playing field for investigating the prevalence of Bayesian and case-based reasoning. We avoided all explicit references to probability. Information about probability was always conveyed by describing physical randomising devices, but these descriptions were designed to make the translation between physical properties and objective probabilities as simple as possible. Judgements about the "chances" of a winning outcome were elicited on a qualitative scale. Thus, participants were not primed to think about probability, but no obstacles were placed in the way of participants who were predisposed to think in this way. The two lotteries seen by any participant were displayed and described in exactly the same way, except for two differentiating features – colour (blue or yellow) and position (left or right on the participant's computer screen). Our background assumption was that, for participants who used Bayesian reasoning, it would be obvious that colour and position provided no information about probabilities or payoffs. However, by making these irrelevant features visually salient, we made it more likely that participants who used case-based reasoning would maintain separate memories of their experiences of the two lotteries.

Each participant was informed about two lotteries, described as the *blue game board* and the *yellow game board*. Within an experimental session, the boards were the same for all participants. At appropriate times, these boards were displayed on participants' screens, vibrantly coloured in blue or yellow. The blue board always appeared on the left side of the participant's screen and the yellow board always appeared on the right. Each game board had 100 numbered *boxes*, corresponding with different numbered balls that might be drawn from a bingo cage (the same cage for both boards). Each box on each board had a predetermined value of either GBP 20 (a *winning box*) or zero (a *losing box*), but this value was not visible to participants until the box was "opened." At the start of the session, all boxes were closed.

The experiment had two parts. In Part 1, each participant played ten *sample rounds*, five using the blue board and five using the yellow board, in random order. These were described as "samples that will give you the opportunity to learn as much as you can about the game boards." In each sample round, the relevant game board was displayed on participants' screens. One ball was drawn from the bingo cage, without replacement. The corresponding box on the board was opened to show its value, with a green background if it was a winning box and a red background if it was a losing box. It remained open only until the end of the round. In this way, each participant learned the values of five of the 100 boxes on each board, selected at random subject to the constraint that all ten opened boxes would have different numbers. Because no more than one box was

---

[5]The experiment was programmed in z-Tree. (Fischbacher, 2007)

open at any time, participants could access the information revealed in the sample rounds only by attending to each round as it occurred and by memorising its outcome. This design feature ensured that participants accumulated memory through experience over time, as is usually assumed in interpretations of CBDT. Although the sample rounds were not decision problems in the strict sense of CBDT, the framing was designed to encourage participants to think of each round as a demonstration of what they might in fact experience, were they to choose to play the relevant game board. At the end of Part 1, all balls were returned to the bingo cage.

In Part 2, each participant faced a *valuation task* relating to one of the two game boards, selected at random, independently for each participant. She was told that she had the opportunity to play this board – that is, to receive the value of one box on that board, determined by one draw from the bingo cage. The mechanism of Becker et al. (1964) was used to elicit the minimum amount of money that each participant was willing to accept in return for giving up this opportunity. Each participant considered thirty-five possible *offer prices*, ranging from GBP 0.20 to GBP 20, and reported whether she was willing to accept each price. In effect, each participant faced thirty-five binary choice problems, each involving a choice between playing the game board and receiving some amount of money with certainty. No feedback on the outcome of any of these choices was provided until the end of the experiment, when one of the offer prices (selected at random) was revealed as the actual offer price, and participants' decisions conditional on that price were implemented. Irrespective of whether she had chosen to keep or sell the opportunity to play, each participant saw an independent draw (with replacement) from the bingo cage. This determined the number of one box on her board, which was then opened. If she had chosen to keep the opportunity, she was paid the value of this box; if not, she was paid the offer price. All participants received an additional participation fee of GBP 2.

Notice that the valuation task is an instance of the certainty equivalence procedure, as described in Section 3. Thus, if participants behaved according to EUT (or, indeed, according to any of a wide class of non-expected utility theories), the reported valuation of any given participant would be an increasing function of the subjective probability she assigned to winning the final lottery.

At the start of each sample round in Part 1, and also immediately before the elicitation task in Part 2, each participant reported her judgement about "the chance that this game board [i.e. the board relevant for that round or task] will reveal a winning box in this round" on a ten-point Likert scale with end-points labelled "very low" and "very high". These *judgement tasks* were not incentivised.

We used a between-subjects design with three treatments, implementing different properties of correlation between the lotteries. Each session was pre-assigned to one of the three treatments. The differences between treatments can be described in terms of the proportions π_B and π_Y of winning boxes on the final game board on the blue and yellow boards respectively. In each treatment, the

values of π_B and π_Y were determined by a random draw from a joint distribution of (π_B, π_Y). Participants were fully informed about the prior distribution, but were not informed about the actual draw. Thus, they were given sufficient information to construct objective prior probabilities of winning on each board, which could then be updated by Bayesian inference in light of the outcomes of the sample rounds. The method by which this information was communicated to participants is explained in Appendix A.

In all three treatments, and in all realisations of the random mechanism, each board was assigned either ten or thirty winning boxes. We refer to a game board with thirty winning boxes as being a *high type* board (type H), and a game board with ten winning boxes as a *low type* board (type L). Ex ante, a given board was equally likely to be of the high or low type, and therefore the ex ante chance of a given box on the board being a winning box was 0.2. Thus, colours and box numbers had no information content in themselves.

The treatments differed in the joint probability distribution used to assign the types to the two boards.

- In the *independent* treatment, the types of the boards were drawn independently.

- In the *positive correlation* treatment, either both boards were type H, or both boards were type L. Each possibility was equally likely.

- In the *negative correlation* treatment, one board was always of type H and the other of type L. It was equally likely that the blue board was type H and the yellow type L, or vice versa.

For a fully rational Bayesian reasoner, all relevant information about correlation is contained in the initial joint distribution of board types. However, the assignment of winning boxes to boards had certain additional features, designed to help participants to understand the correlation properties of each treatment.

In the positive correlation treatment, the winning numbers were the same for both boards. For example, consider a participant who observes a win on box 36 of the blue board in a sample round of the positive treatment. By abstract Bayesian reasoning from the knowledge that there is perfect positive correlation between π_B and π_Y, she can deduce that the observed win is just as informative about the probability of winning on the yellow board as it is about the probability of winning on the blue board. But our design allows her to make a more direct and more concrete inference, from the knowledge that *box 36* on the blue board is a winning box to the conclusion that that *box 36* on the yellow board is a winning box. In the negative correlation treatment, the winning numbers were different for the two boards. Thus, from the knowledge that box 36 on the blue board was a winning box, a participant could infer that box 36 on the yellow board was a losing box. In the independent treatment, the assignments of winning numbers to the two boards were independent of one another.

As a further aid to understanding, each sample round ended with a screen, described as "sum-maris[ing] what you learned about the game boards in that round." On this screen, participants saw the game board they had just played, with the box just opened coloured green or red and showing "GBP 20" or "GBP 0." At the bottom of the board there was a message reinforcing this information. On the other side of the screen there was a message about the corresponding box on the other game board. For example, consider a round involving the blue game board. Suppose the announced box number was 4 and its value was GBP 20. Then (irrespective of the treatment) the message on the blue game board would be "4 is a winning box on the blue game board." In the positive correlation treatment, the message on the other side of the screen would be "4 is a winning box on the yellow game board." In the negative correlation treatment, it would be "4 is a losing box on the yellow game board." In the independent treatment, it would be "4 may be a winning box or may be a losing box on the yellow game board." Throughout Part 1, the computer screen also displayed a header that constantly reminded participants of the correlation between the two game boards.

Full instructions for the experiment, including examples of screenshots, are given in an online appendix.

5   Hypotheses

Let i index the game board offered to the participant in Part 2, and j denote the other game board. Therefore, $i, j \in \{\text{blue}, \text{yellow}\}$, with $i \neq j$. Let h_i denote the number of winning boxes observed among draws from board i in Part 1, and h_j the number of winning boxes observed among draws from board j in Part 1. The values of h_i and h_j are sufficient to compute the Bayesian posterior probability that board i is type H, which we denote $\rho_i(h_i, h_j)$. Table 1 presents these posteriors for each treatment, for the memories realised in our data.[6] We refer to the pair (h_i, h_j) observed by a participant as their *memory*.

For a given participant, EUT predicts that the stated valuation of a game board will be increasing in the posterior. However, because participants may vary in terms of risk attitudes, this relationship need not be linear, nor the same for all participants. Furthermore, it is unreasonable to expect participants to assess posteriors in a numerically precise way. Therefore, for each treatment I (independent), P (positive correlation), and N (negative correlation), we define corresponding reflexive, complete, and transitive orderings $\succeq_I^E$, $\succeq_P^E$, and $\succeq_N^E$ over memories which are based on the qualitative properties of the Bayesian posterior. For any two memories (h_i, h_j) and (h_i', h_j') we define these orderings as follows:

---

[6]See Appendix B for notes on calculating these.

$\rho_i(h_i, h_j)$	$h_j = 0$	$h_j = 1$	$h_j = 2$
$h_i = 0$	.216	.216	.216
$h_i = 1$	.518	.518	.518
$h_i = 2$	.818	.818	.818

(a) Independent

$\rho_i(h_i, h_j)$	$h_j = 0$	$h_j = 1$	$h_j = 2$
$h_i = 0$	.065	.216	.541
$h_i = 1$	.216	.541	.844
$h_i = 2$	.541	.844	.965

(b) Positive correlation

$\rho_i(h_i, h_j)$	$h_j = 0$	$h_j = 1$	$h_j = 2$	$h_j = 3$
$h_i = 0$	.500	.201	.056	.013
$h_i = 1$	.799	.500	.190	.050
$h_i = 2$	.944	.810	.500	.178
$h_i = 3$	.987	.952	.822	.500

(c) Negative correlation

Table 1: Bayesian posterior probabilities $\rho_i(h_i, h_j)$ as a function of the memory (h_i, h_j), for each treatment.

- In the independent treatment, the posterior is determined entirely by the number of winning boxes h_i observed on board i. We therefore define $\succeq_I^E$ by

$$(h_i, h_j) \succeq_I^E (h_i', h_j') \iff h_i \geq h_i'.$$

- In the positive correlation treatment, the posterior is determined entirely by the total number of winning boxes $h_i + h_j$ observed, irrespective of board. We therefore define $\succeq_P^E$ by

$$(h_i, h_j) \succeq_P^E (h_i', h_j') \iff h_i + h_j \geq h_i' + h_j'.$$

- In the negative correlation treatment, the posterior is increasing in the number of winning boxes observed on board i, and decreasing in the number of winning boxes observed on board j. The difference of the number of winning boxes between the boards, $h_i - h_j$, is not alone sufficient to determine the posterior. However, among memories for which $h_i - h_j = m$ for some m, the posterior varies by only a small amount, compared to the variability of the posterior between two memories for which $h_i - h_j \neq h_i' - h_j'$. We therefore define the ordering $\succeq_N^E$ based on the difference $h_j - h_j$,

$$(h_i, h_j) \succeq_N^E (h_i', h_j') \iff h_i - h_j \geq h_i' - h_j'.$$

We derive relations of strict ordering $\succ^E$ and equivalence $\sim^E$ from $\succeq^E$ in the usual way.

The ranking of a memory is therefore determined by a treatment-specific summary statistic, which we write $M^I(h_i, h_j) = h_i$ for the independent treatment; $M^P(h_i, h_j) = h_i + h_j$ for the positive correlation treatment; and $M^N(h_i, h_j) = h_i - h_j$ for the negative correlation treatment. The summary statistic M^t (for $t \in \{I, P, N\}$) therefore represents the ordering $\succeq_t^E$.

The experimental design assigns each participant at random to a memory (h_i, h_j) determined by the randomly-selected sample box numbers and the game board offered. Each participant k reports a valuation, which we denote $v_k(h_i, h_j)$. Under EUT, and indeed any decision theory in which valuations are monotonic in the Bayesian posterior, valuations will increase in the summary statistic. The following hypothesis is therefore an implication of EUT.

Hypothesis 1. *For each treatment t, the reported valuations are increasing in M^t, the summary statistic of the memory.*

Our interest is in whether actual behaviour deviates from EUT in directions that would be indicative of case-based reasoning. Our design is based on the working assumption that, in all three treatments, participants perceive plays on boards of the same colour as more similar to one another, than to plays on boards whose colours are different. Given this assumption, the underlying

principles of CBDT suggest the following *case-based weighting conjecture*: when participants report valuations of the offered board i, they give greater weight to winning boxes observed on that board than to winning boxes observed on board j. (Wins on board j might be given zero weight, as in Model 1 of Section 2, or positive but lower weight than wins on board j, as in Model 2.) Following the methodological strategy of looking for exhibits (see Section 1), we focus on cases in which this conjecture implies unambiguous biases in behaviour relative to EUT predictions. Such biases are implied only in the positive and negative correlation treatments.

We formalise the case-based weighting conjecture by defining an alternative ranking of memories in the positive correlation and negative correlation treatments.

- In the positive correlation treatment, fix some $m \in \mathbb{Z}_+$ and consider the set of memories $\{(h_i, h_j) : h_i + h_j = m\}$. The case-based weighting conjecture implies a strict ranking $\succ_P^C$ of the members of this set,

$$(h_i, h_j) \succ_P^C (h_i', h_j') \iff h_i + h_j = h_i' + h_j' = m \text{ and } h_i > h_i'. \tag{1}$$

- In the negative correlation treatment, fix some $m \in \mathbb{Z}$ and consider the set of memories $\{(h_i, h_j) : h_i - h_j = m\}$. The case-based weighting conjecture implies a strict ranking $\succ_N^C$ of the members of this set,

$$(h_i, h_j) \succ_N^C (h_i', h_j') \iff h_i - h_j = h_i' - h_j' = m \text{ and } h_i > h_i'. \tag{2}$$

The following hypothesis is an implication of the case-based weighting conjecture:

Hypothesis 2. *For a given treatment* $t \in \{P, N\}$, *consider any set of memories* $\{(h_i, h_j) : M^t(h_i, h_j) = m\}$ *for some* m. *Within this set, reported valuations are increasing in* h_i, *the number of winning boxes observed on the offered board.*

Recall that each participant reported judgements about the "chance" of seeing a winning number prior to each box being revealed, both in Part 1 and Part 2. To allow meaningful comparisons across participants, we normalise each participant's use of the Likert scale relative to that participant's judgement of the chance of winning on the first Part 1 game board. At the stage of the experiment at which this first judgement was reported, all participants had the same information, and that information implied that the objective probability of seeing a winning box was 0.2. For each participant, we define the variable *expectation difference*, which takes on the value $+1$ when the participant reports a higher chance in Part 2 than at the start of Part 1; -1 when the participant reports a lower chance in Part 2 than at the start of Part 1; and 0 when the chances reported in Part 2 and at the start of Part 1 are the same. Expectation difference can be interpreted as a self-reported

15

judgement about whether the probability of seeing a winning box on the relevant board i, given the participant's memory (h_i, h_j), is greater than, less than, or equal to 0.2. Thus, if stated probabilities are consistent with Bayesian reasoning, the following hypothesis will hold:

Hypothesis 3. *For each treatment t, the reported expectation differences are increasing in M^t, the summary statistic of the memory.*

Because CBDT makes no reference to probability, it does not support any particular conjecture about stated probabilities.

6 Results

We conducted a total of thirty sessions in March 2014; there were ten sessions for each treatment. Each session had six to eight participants and lasted 45 minutes. Average earnings were GBP 8.40, and ranged from GBP 2.00 to GBP 22.00. All 226 participants (119 male, 107 female) were recruited from the standing participant pool recruited by the Centre for Behavioural and Experimental Social Science at University of East Anglia, managed via ORSEE (Greiner, 2015).

The experimental design assigned participants randomly to experimental sessions, and therefore to treatments. Furthermore, because the sample results from the game boards were themselves determined at random, the design randomly assigned participants to realised memories (h_i, h_j). Therefore, the data of any pair of memories $(h_i, h_j) \neq (h'_i, h'_j)$ are independent samples.

Our hypotheses make predictions about how valuations and expectation differences change as a function of the observed memory. We base our statistical analysis on nonparametric tests using rank orders. These tests are well-suited to our hypotheses in that our hypotheses only concern trends in valuations as function of h_i and h_j. In addition, because our design uses a discrete set of questions to implement the Becker et al. (1964) procedure, our valuations can be interpreted only as intervals and are therefore ordinal, which is sufficient for the rank order approach.

Our choice architecture did not require participants to give a monotonic response to the questions implementing the Becker et al. procedure. Participants could indicate simultaneously that they accepted some price while rejecting a strictly higher price. Of the 226 participants, 210 provided monotonic responses to the valuation questions. In addition, 4 participants gave responses which were monotonic with the exception of one isolated price, such that, if the response to that price were inverted, the resulting schedule is monotonic. We neglect the isolated non-monotonic response for those 4 participants and include them in our analysis. We drop the remaining 12 participants. For each of the 214 participants in our sample so defined, we define their valuation as the lowest accepted price.[7]

---

[7]More precisely, the revealed valuation falls within the interval between the highest rejected price and lowest

Table (a) Independent

(h_i, h_j) =	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)
Posterior		.216			.518		.818	
RNEUT		286			407		527	
N	8	7	8	8	14	11	8	11
Quartile 1	150	300	270	300	300	260	450	500
Median	260	400	475	425	500	400	675	600
Quartile 3	650	800	750	850	750	500	800	800

(a) Independent

Table (b) Positive treatment

(h_i, h_j) =	(0,0)	(0,1)	(1,0)	(0,2)	(1,1)	(2,0)	(1,2)	(2,1)	(2,2)
Posterior	.065		.216		.541		.844		.965
RNEUT	226		286		416		536		586
N	7	15	13	4	7	4	8	6	8
Quartile 1	200	200	260	475	350	675	525	450	550
Median	220	350	400	500	500	725	650	525	650
Quartile 3	500	700	500	650	750	1375	1000	800	900

(b) Positive treatment

Table (c) Negative treatment

(h_i, h_j) =	(0,3)	(0,2)	(1,2)	(0,1)	(1,1)	(2,2)	(1,0)	(2,1)	(2,0)	(3,0)
Posterior	.013	.056	.190	.201		.500	.799	.810	.944	.987
RNEUT	205	223	276	280		400	520	524	578	595
N	2	9	3	9	7	7	11	4	11	4
Quartile 1	280	400	100	200	500	400	300	550	350	450
Median	490	600	500	450	750	550	500	700	500	500
Quartile 3	700	750	500	600	1200	1100	750	900	800	600

(c) Negative treatment

Table 2: Summary data for valuations by treatment. N is the number of participants to observe each memory. Posterior refers to the Bayesian posterior, given the memory, that the offered board i is of the high type, and RNEUT gives the expected payoff of a single play of board i given the posterior. All values given in pence.

Table 2 summarises the data on stated valuations. For each treatment and memory, we report the median valuation, as well as the first and third quartiles. We group memories for each treatment t according to $\succeq_t^E$, and provide the expected payoff of a single play of board i given the posterior. We note that this expected payoff is lower than the median stated valuation for most memories. This tendency to overvalue a gamble with low probability of winning a large amount is observed, for example, in preference-reversal experiments such as Seidl (2002). In the independent and positive treatments, median valuations generally increase in the posterior probability; the median valuation is roughly 100 pence above the expected payoff. In contrast, in the negative treatment valuations do not appear to vary systematically as a function of the posterior. For example, given the memory $(h_i, h_j) = (2, 0)$, the expected payoff to a play of board i is 578 pence, contrasting with an expected payoff of 223 given the memory $(h_i, h_j) = (0, 2)$. However, the median valuation in $(2, 0)$ is actually lower than that in $(0, 2)$.

To give a more complete picture of how the distribution of valuations changes as a function of the memory, we provide measurements of effect sizes for each pair of memories within a treatment. Consider two memories (h_i, h_j) and (h'_i, h'_j), and let v be a randomly-chosen valuation observed at (h_i, h_j) and v' a randomly-chosen valuation observed at (h'_i, h'_j). The effect size is computed as $Pr(v > v') + \frac{1}{2}Pr(v = v')$. These are reported in Table 3, in which each cell is the probability that a valuation drawn from the row memory is greater than a valuation drawn from the column memory. This effect size is directly related to the test statistic used in the Mann-Whitney-Wilcoxon (MWW) test for equality of distributions. As a convenience, we indicate in Table 3 the pairs of memories for which the effect size is statistically different from one-half using the MWW test.

Because memories are sorted in order of the posterior probability that board i is a high-type board, it would be expected under EUT to see effect sizes less than one-half towards the top-right of the matrix. This pattern is observed broadly in the independent and positive correlation treatments. However, no clear pattern emerges in the negative correlation treatment.

The MWW test is suitable for comparing the distributions of valuations in two groups. In some circumstances, our hypotheses require comparing across three or more groups. For this purpose we adopt the test for trend of Cuzick (1985). This test extends the rank-order calculation of the MWW test to three or more groups; for two groups Cuzick (1985)'s test coincides with MWW exactly. This test requires ordering the groups being compared. The null hypothesis is there is no trend (increasing or decreasing) in the data across the groups, against the alternative hypothesis of a trend.

Result 1. *In the independent and positive correlation treatments, valuations are increasing as a function of the summary statistic $M^t(h_i, h_j)$ of the memory. In the negative correlation treatment, we cannot reject the null hypothesis of no trend.*

accepted price. We identify each interval by its upper endpoint for conciseness.

Posterior			.216			.518			.818	
(h_i, h_j)		(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)
	N	8	7	8	8	14	11	8	11
(0,0)	8	—	.295	.328	.352	.339	.443	.297	.256*
(0,1)	7		—	.563	.473	.454	.623	.375	.325
(0,2)	8			—	.477	.473	.614	.383	.381
(1,0)	8				—	.478	.619	.422	.364
(1,1)	14					—	.675	.411	.367
(1,2)	11						—	.222**	.215**
(2,0)	8							—	.500
(2,1)	11								—

(a) Independent

Posterior		.065	.216		.541			.844		.965
(h_i, h_j)		(0,0)	(0,1)	(1,0)	(0,2)	(1,1)	(2,0)	(1,2)	(2,1)	(2,2)
	N	7	15	13	4	7	4	8	6	8
(0,0)	7	—	.467	.330	.214	.316	.107**	.134**	.298	.134**
(0,1)	15		—	.464	.317	.390	.175**	.237**	.356	.258*
(1,0)	13			—	.308	.379	.058***	.192**	.359	.154***
(0,2)	4				—	.589	.188	.297	.542	.313
(1,1)	7					—	.196	.313	.417	.295
(2,0)	4						—	.641	.750	.734
(1,2)	8							—	.656	.508
(2,1)	6								—	.344
(2,2)	8									—

(b) Positive treatment

Posterior		.013	.056	.190	.201	.500		.799	.810	.944	.987
(h_i, h_j)		(0,3)	(0,2)	(1,2)	(0,1)	(1,1)	(2,2)	(1,0)	(2,1)	(2,0)	(3,0)
	N	2	9	3	9	7	7	11	4	11	4
(0,3)	2	—	.333	.667	.583	.214	.429	.455	.250	.318	.438
(0,2)	9		—	.741	.698	.381	.508	.586	.333	.530	.597
(1,2)	3			—	.463	.190	.238	.333	.083*	.273	.333
(0,1)	9				—	.222*	.325	.384	.181*	.313	.375
(1,1)	7					—	.582	.682	.482	.630	.714
(2,2)	7						—	.571	.375	.526	.589
(1,0)	11							—	.295	.430	.511
(2,1)	4								—	.670	.813
(2,0)	11									—	.568
(3,0)	4										—

(c) Negative treatment

Table 3: Effect sizes for comparisons of valuations, pairwise by outcome of sample draws. Each cell is the probability that a randomly-selected valuation reported given the memory in the row is greater than a randomly-selected valuation reported given the memory in the column. Asterisks indicate significance of the difference using the Mann-Whitney-Wilcoxon test; * at 10%, ** at 5%, *** at 1%.

Comparison	Independent	Positive	Negative
$M = -3$ vs $M = -2$			.381
$M = -2$ vs $M = -1$			.624*
$M = -1$ vs $M = 0$			.421
$M = 0$ vs $M = 1$	.469	.403	.496
$M = 1$ vs $M = 2$	.330**	.295**	.389
$M = 2$ vs $M = 3$		.457	.658
$M = 3$ vs $M = 4$		.438	
Cuzick p	.049	.0004	.577

Table 4: Test for increasing valuations as a function of the summary statistic. Each row reports the effect size of the pairwise comparison of valuations given two summary statistics. Asterisks indicate significance of the pairwise difference using the Mann-Whitney-Wilcoxon test; * at 10%, ** at 5%, *** at 1%.

Support. Table 4 reports an aggregation of the pairwise comparisons as shown in Table 3, where pairwise comparisons are made after grouping valuations by the corresponding summary statistic of the memory. If valuations are increasing in the summary statistic, effect sizes below .5 are expected. This is observed in the pairwise comparisons under the independent and positive correlation treatments. Cuzick's test confirms the trend is significant ($p = .049$ for the independent treatment and $p = .0004$ for the positive correlation treatment). In contrast, the direction of the pairwise comparisons is mixed for the negative correlation treatment, with two of the six pairwise comparisons going in the opposite direction. The null hypothesis of no systematic trend cannot be rejected ($p = .577$). □

Result 2. *There is no significant evidence of the systematic deviations from EUT implied by the case-based weighting conjecture.*

Support. Hypothesis 2 proposes that, in the positive and negative correlation treatments and conditional on a realised value m for the summary statistic, valuations will be increasing in the number h_i of winning boxes observed on the offered board. Conditional on a value of m, we conduct Cuzick's test for trend in h_i, and report the results in Table 5. Each row corresponds to one conditional test. For each row, we repeat the pairwise effect sizes drawn from Table 3. If an increasing trend in h_i were present, we would expect effect sizes in excess of .5. No individual pairwise comparison, nor any conditional test for trend, is significant. Because of the random assignment of participants to memories, each of the conditional tests are independent of each other, and therefore we can combine the results of the conditional tests into a grand test using Fisher's combined probability method. We cannot reject the null hypothesis of no trend overall ($p = .330$ for positive correlation and $p = .410$ for negative correlation). □

| | Pairwise effect size | | |
Conditional on	$h_i = 0$ vs. $h_i = 1$	$h_i = 1$ vs. $h_i = 2$	Cuzick p
$h_i + h_j = 1$	.464		.746
$h_i + h_j = 2$	.589	.196	.203
$h_i + h_j = 3$		.656	.330
		Combined	.330

(a) Positive treatment

| | Pairwise effect size | | |
Conditional on	$h_i = 0$ vs. $h_i = 1$	$h_i = 1$ vs. $h_i = 2$	Cuzick p
$h_i - h_j = -1$	.537		.852
$h_i - h_j = 0$		.418	.607
$h_i - h_j = 1$		.705	.235
		Combined	.410

(b) Negative treatment

Table 5: Test for trends predicted by the similarity conditional on the summary statistic of a memory. The pairwise effect sizes are for comparisons as reported in Table 3. The p-value for each conditional test for trend is reported in each row. The overall p-value reported is calculated using Fisher's combined method.

We now turn to the data on stated judgements. Table 6 shows, for each Bayesian posterior, the breakdown of expectation differences across participants. Overall there is a general trend of decreased optimism, in that around half of the participants report a less optimistic judgement prior to Part 2 than at the start of Part 1. Given that the initial probability of seeing a winning box was only 0.2, the direction of this trend is consistent with the known tendency for rare events to be given greater weight when decisions are made *from description* – that is, on the basis of a priori information about the properties of random mechanisms – than when they are made *from experience* – on the basis of trial-by-trial experience of the realisations of random mechanisms. (Hertwig et al., 2004) In our experiment, judgements about the first Part 1 game board could be made only from description, while judgements about the Part 2 board could take account of both description and experience.

Table 7 presents effect size comparisons on the judgements data. Because we code only the sign of the change in reported judgement from the first period to the last, our measure of judgement is coarse. We therefore aggregate judgements by the level sets $\sim^E$ given by EUT. All treatments show a general pattern of small effect sizes in the upper-right of the table, which indicates a broad consistency of judgements with Bayesian posteriors.

Result 3. *In all treatments, judgements are more optimistic, as measured by expectation differences, when the Bayesian posterior probability is higher.*

$\rho_i(h_i, h_j)$	Decrease	No change	Increase	Total
.216	15	7	2	**24**
.518	23	9	2	**34**
.818	8	4	7	**19**
Total	**46**	**20**	**11**	77

(a) Independent

$\rho_i(h_i, h_j)$	Decrease	No change	Increase	Total
.065	6	1	1	**8**
.216	17	8	4	**29**
.541	9	5	1	**15**
.844	7	4	4	**15**
.965	2	4	2	**8**
Total	**41**	**22**	**12**	75

(b) Positive correlation

$\rho_i(h_i, h_j)$	Decrease	No change	Increase	Total
.013	2	1	0	**3**
.056	7	1	3	**11**
.190	2	0	1	**3**
.201	5	4	2	**11**
.500	8	5	2	**15**
.799	7	1	4	**12**
.810	0	2	2	**4**
.944	3	3	5	**11**
.987	1	1	2	**4**
Total	**35**	**18**	**21**	74

(c) Negative correlation

Table 6: Expectation difference, as a function of posterior.

h_i		0	1	2
	N	20	33	22
0	20	—	.528	.353*
1	33		—	.330**
2	22			—

(a) Independent

$h_i + h_j$		0	1	2	3	4
	N	8	29	15	15	8
0	8	—	.427	.442	.358	.266*
1	29		—	.521	.422	.332
2	15			—	.398	.300*
3	15				—	.425
4	8					—

(b) Positive correlation

$h_i - h_j$		-3	-2	-1	0	1	2	3
	N	3	11	14	15	16	11	4
-3	3	—	.439	.381	.411	.323	.227	.208
-2	11		—	.461	.488	.409	.335	.318
-1	14			—	.533	.435	.351	.330
0	15				—	.402	.312*	.292
1	16					—	.426	.406
2	11						—	.477
3	4							—

(c) Negative correlation

Table 7: Effect sizes for comparisons of judgements, pairwise by outcome of sample draws. Each cell is the probability that a randomly-selected judgement reported after the outcome in the row is greater than a randomly-selected judgement reported after the outcome in the column. Asterisks indicate significance of the difference using the Mann-Whitney-Wilcoxon test; * at 10%, ** at 5%, *** at 1%.

Comparison	Independent	Positive	Negative
$M = -3$ vs $M = -2$			.439
$M = -2$ vs $M = -1$			.461
$M = -1$ vs $M = 0$			.533
$M = 0$ vs $M = 1$	.528	.427	.402
$M = 1$ vs $M = 2$	.330**	.521	.426
$M = 2$ vs $M = 3$		.398	.477
$M = 3$ vs $M = 4$		.425	
Cuzick p	.084	.048	.030

Table 8: Test for increasing judgements as a function of the summary statistic. Each row reports the effect size of the pairwise comparison of judgements given two summary statistics. Asterisks indicate significance of the pairwise difference using the Mann-Whitney-Wilcoxon test; * at 10%, ** at 5%, *** at 1%.

Support. We follow the same approach as in Result 1, using the expectation differences in place of valuations. The details for Cuzick's test for trend are summarised in Table 8. If judgements are more optimistic when the Bayesian posterior is higher, then we expect effect sizes to be less than .5. This is generally the case for all three treatments. Notably, for the negative correlation treatment, this contrasts sharply with the data for the corresponding Result 1 test in Table 4. We reject the null hypothesis of no trend for all treatments, ($p = .084$ for independent, $p = .048$ for positive correlation, and $p = .030$ for negative correlation). □

Our data therefore show a contrast between the patterns in participants' valuations and the patterns in their expectation differences in the negative treatment. Table 2 suggests stated valuations remain high in the negative treatment even when the Bayesian posterior gives a low chance the offered board is of the high type, and Result 1 formalises this statement. However, Result 3 shows expectation differences are significantly more optimistic in situations in which the Bayesian posterior advises they should indeed be more optimistic.

7   Discussion

Our experiment was motivated by the conjecture that human decision makers have a tendency to use case-based reasoning even when events are well-defined and when the objective probabilities of those events can be found by Bayesian reasoning from prior information. More specifically, we tested the "case-based weighting conjecture" that systematic violations of EUT occur when information about correlation is not embedded in experienced decision outcomes and when Bayesian reasoning involves inferences between lotteries that are saliently dissimilar. We designed the experiment in the belief that this conjecture was consistent with the psychological intuitions of CBDT. Had the evidence confirmed that conjecture, we would have interpreted our results as providing support for CBDT. In fact, we found no significant evidence of the effects implied by the conjecture. If we are right to claim that our conjecture reflects the assumed psychology of case-based reasoning, we must conclude that our results count against CBDT.

However, although we find no evidence of the specific form of correlation neglect implied by our conjecture, it is clear our participants found it difficult to perform Bayesian reasoning about decisions that involved negative correlation. More precisely, when reporting *valuations* of an offered lottery i, participants were generally able to recognise the irrelevance of outcomes of lotteries that were independent of i, and to recognise the relevance of outcomes of lotteries that were positively correlated with i; but they failed to recognise the relevance of negative correlation. Surprisingly, however, their *stated probability judgements* about lottery i (expressed in qualitative statements about the "chance" of winning) showed Bayesian responses to independence, positive correlation,

and negative correlation. Our post-hoc conjecture is that the problem was one of cognitive overload. Intuitively, negative correlation is a more difficult concept than independence or positive correlation, and working out one's willingness to exchange a lottery for money is more difficult than merely judging the chance of winning if one plays it. The combination of these two sources of difficulty may have been too challenging for our participants.

This surprising finding raises doubts about the almost universal practice in experimental economics of treating incentivised decision problems as the gold standard for the elicitation of participants' beliefs. If linking the elicitation of beliefs with problems of decision under uncertainty can lead to cognitive overload, experimentalists need to consider the possibility that direct, nonincentivised questions about beliefs might produce more accurate data.

References

Nicholas Bardsley, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer, and Robert Sugden. *Experimental Economics: Rethinking the Rules*. Princeton University Press, 2010.

Gordon Becker, Morris DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232, 1964.

Han Bleichrodt, Martin Filko, Amit Kothiyal, and Peter P. Wakker. Making case-based decision theory directly observable. Working paper, http://people.few.eur.nl/wakker/pdf/cbdt.pdf, 2015.

Gary Charness and Dan Levin. When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review*, 95:1300–1309, 2005.

Jack Cuzick. A Wilcoxon-type test for trend. *Statistics in Medicine*, 4:87–90, 1985.

Erik Eyster and Georg Weizsäcker. Correlation neglect in financial decision-making. DIW Berlin Discussion Paper No. 1104. Available at http://dx.doi.org/10.2139/ssrn.1735339, 2011.

Urs Fischbacher. Z-tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10:171–178, 2007.

Gerd Gigerenzer and Daniel Goldstein. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 193:650–669, 1996.

Itzhak Gilboa and David Schmeidler. Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639, 1995.

Ben Greiner. Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1:114–125, 2015.

Brit Grosskopf, Rajiv Sarin, and Elizabeth Watson. An experiment on case-based decision making. *Theory and Decision*, 79:639–666, 2015.

Todd Guilfoos and Andreas Duus Pape. Predicting human cooperation in the Prisoner's Dilemma using case-based decision theory. *Theory and Decision*, 80:1–32, 2016.

Ted Martin Hedesström, Henrik Svedsäter, and Tommy Gärling. Covariation neglect among novice investors. *Journal of Experimental Psychology: Applied*, 12:155–165, 2006.

Ralph Hertwig, Greg Barron, Elke Weber, and Ido Erev. Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8):534–539, 2004.

Leonid Hurwicz. Optimality criteria for decision making under ignorance. Cowles Commission Discussion Paper, Statistics No. 320, 1951.

Ido Kallir and Doron Sonsino. The neglect of correlation in allocation decisions. *Southern Economic Journal*, 75:1045–1066, 2009.

Yoram Kroll, Haim Levy, and Amnon Rapoport. Experimental tests of the separation theorem and capital asset pricing model. *American Economic Review*, 78:500–519, 1988.

Wolfgang Ossadnik, Dirk Wilmsmann, and Benedikt Niemann. Experimental evidence on case-based decision theory. *Theory and Decision*, 75:211–232, 2013.

Christian Seidl. Preference reversal. *Journal of Economic Surveys*, 16(5):621–655, 2002.

Robert Sugden. *Handbook of Utility Theory Volume 2: Extensions*, chapter Alternatives to expected utility: Foundations and concepts, pages 685–755. Kluwer Academic, 2004.

Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5:207–232, 1973.

Amos Tversky and Daniel Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315, 1983.

A   Determining winning numbers

Before each session of the independent treatment, forty distinct winning numbers were drawn as for the other two treatments; thirty were placed in one envelope and ten in the other. This pair of envelopes was assigned to the blue game board. All balls were returned to the bingo cage and the process was repeated, resulting in a second pair of envelopes, assigned to the yellow game board. A participant drew one of the first pair of envelopes and marked it "blue". The numbers in this envelope became the winning numbers on the blue game board. She then drew one of the second pair of enveloped and marked it "yellow"; the numbers in this envelope became the winning numbers on the yellow game board.

Before each session of the positive correlation treatment, forty distinct "winning" numbers were determined by draws (without replacement) from a bingo cage with 100 balls numbered from 1 to 100. The winning numbers were distributed between two envelopes according to the order in which they had been drawn, thirty numbers being put in one envelope and ten in the other. The envelopes were then sealed. This process was fully explained to participants. At an appropriate stage in the reading aloud of the experimental instructions, the two envelopes were shown to participants and placed in a bag. A participant was asked to draw one envelope from the bag and to mark it "blue and yellow". While the experimenter continued to read the instructions, an assistant entered the numbers in that envelope into the program for the experiment. These became the winning numbers on *both* game boards.

Before each session of the negative correlation treatment, forty distinct winning numbers were drawn in the same way as for the positive treatment; thirty were placed in one envelope and ten in the other. The two envelopes were put in a bag. A participant drew one envelope from the bag and marked it "blue". The numbers in this envelope became the winning numbers on the blue game board. She then drew the second envelope and marked it "yellow"; the numbers in this envelope became the winning numbers on the yellow game board.

In all treatments, after the winning numbers had been entered into the program, all envelopes were posted on a wall. Participants were invited to inspect the contents of the envelopes at the end of the session.

B   Bayesian posteriors

In calculating the posteriors, recall that the samples from the board are drawn without replacement. Therefore, the distribution of hits is given by a suitable parameterisation of the hypergeometric distribution.

In the *independent* treatment, let H denote the state in which board i is of the high type, and

L the state in which it is low. If board i is of the high type, then h_i follows the hypergeometric distribution,

$$P(h_i|H) = \frac{\binom{30}{h_i}\binom{70}{5-h_i}}{\binom{100}{5}}$$

and if board i is of the low type, then h_i follows the hypergeometric distribution,

$$P(h_i|L) = \frac{\binom{10}{h_i}\binom{90}{5-h_i}}{\binom{100}{5}}.$$

In the *positive* treatment, let H denote the state in which both boards i and j are of the high type, and L the state in which both are low. Conditional on state H, the total number of hits $h_i + h_j$ follow the hypergeometric distribution,

$$P(h_i + h_j|H) = \frac{\binom{30}{h_i+h_j}\binom{70}{10-h_i-h_j}}{\binom{100}{10}}$$

and conditional on state L, then $h_i + h_j$ follows the hypergeometric distribution,

$$P(h_i + h_j|L) = \frac{\binom{10}{h_i+h_j}\binom{90}{10-h_i-h_j}}{\binom{100}{10}}.$$

The calculation of probabilities for the *negative* treatment deserves more detailed comment. Let H denote the state in which board i is high and board j is low, and L denote the state in which board i is low and board j is high. Suppose that among the five numbers selected as trials for board i there were h_i winning numbers on board i, and s_j winning numbers on board j. Among the five numbers selected as trials for board j, suppose there were h_j winning numbers on board j, and s_i winning numbers on board i. In our setup, participants do observe h_i and h_j, but do not observe s_i and s_j. Because in our design the winning numbers on the two boards were guaranteed to be distinct, the random variables h_i and h_j are not independent conditional on H or L.

The probability of having h_i winning numbers on board i and s_j winning numbers on board j among the five numbers revealed for board i is given by the multivariate hypergeometric distribution,

$$P(h_i, s_j|H) = \frac{\binom{30}{h_i}\binom{10}{s_j}\binom{60}{5-h_i-s_j}}{\binom{100}{5}}.$$

The probability of having h_j winning numbers on board j and s_i winning numbers on board i among the five numbers revealed for board i, *conditional* on the outcome of the draws on board i,

28

is also a multivariate hypergeometric distribution,

$$P(h_j, s_i|h_i, s_j, H) = \frac{\binom{30-h_i}{s_i}\binom{10-s_j}{h_i}\binom{55+h_i+s_j}{5-s_i-h_j}}{\binom{95}{5}}.$$

Therefore, the total probability in state H of seeing h_i winning numbers among the five draws on board i, and h_j winning numbers among the five draws on board j, is obtained by summing over the unobserved variables s_i and s_j,

$$P(h_i, h_j|H) = \sum_{s_i=0}^{5}\sum_{s_j=0}^{5} \frac{\binom{30}{h_i}\binom{30-h_i}{s_i}\binom{10}{s_j}\binom{10-s_j}{h_j}\binom{60}{5-h_i-s_j}\binom{55+h_i+s_j}{5-s_i-h_j}}{\binom{100}{5}\binom{95}{5}}.$$

By symmetry, $P(h_i, h_j|L) = P(h_j, h_i|H)$. Using these formulas it is straightforward to tabulate the posterior probabilities as given in Table 1.

The calculation above takes care to reflect the fact that the sets of winning numbers on the two boards are disjoint. As an alternative simplifying assumption, one could assume that h_i and h_j are independent conditional on the assignment of the board types.[8] Then, conditional on the board type, h_i and h_j would follow independent hypergeometric distributions, and the posterior probabilities would be

$\rho_i(h_i, h_j)$	$h_j = 0$	$h_j = 1$	$h_j = 2$	$h_j = 3$
$h_i = 0$	.500	.204	.058	.013
$h_i = 1$	.796	.500	.193	.050
$h_i = 2$	.942	.807	.500	.181
$h_i = 3$	.987	.950	.820	.500

These differ only slightly from the results of the precise calculation given in Table 1. The ranking of memories by posterior is not affected.

---

[8]We thank Peter Moffatt for suggesting this.