The cost of asking: how evaluations bias subsequent judgments.

Lee C. White and Emmanuel M. Pothos

City, University of London

Michael Jarrett

INSEAD

Author Note

Lee C. White, Department of Psychology, City, University of London; Emmanuel M. Pothos, Department of Psychology, City, University of London; Michael Jarrett, Organisational Behavior, INSEAD, Europe Campus

Corresponding author. Lee C. White, Department of Psychology, City, University of London, Northampton Square, London, EC1V 0HB, UK. Email: leecwhite@gmail.com Tel: +44 (0)7812 341048

Abstract

A novel decision bias, called the evaluation bias (EB), was reported by White et al. (2014). In a sequence of two stimuli of opposite affective valence, evaluating the first stimulus leads to a more contrasting evaluation for the second one, compared to when the first stimulus is just observed. The EB is consistent with a long tradition of constructive influences or decision biases in questionnaire judgments. The prediction of the EB was based on the application of a quantum probability model, taking advantage of the unique role of evaluations in quantum probability. In the present work, we develop the quantum model so as to examine whether similar predictions are possible in the context of real questionnaires, where precise control over the relative valence of stimulus pairs is impossible. It is shown that an EB prediction can be extracted and we test this prediction in an organizational opinion survey, administered to a range of organizations across four experiments (total N = 868 and 84 organizations) and with two different languages. In all experiments, there was clear evidence for an EB. We examine the result with the quantum model and Hogarth and Einhorn's (1992) belief-adjustment model. Both models can broadly capture the empirical findings and so offer promise for providing a formal understanding of constructive influences.

Keywords: cognitive biases; constructive influences; quantum theories; decision making

The cost of asking: how evaluations bias subsequent judgments

1. Introduction

1.1 General overview

There have been several evocative results that show how a judgment or evaluation can affect subsequent cognitive processing (e.g., later judgments, opinions, or preferences). For example, in the context of a Gallup opinion poll, Moore et al. (2002) famously reported the following result. Consider the two questions ‘Is Clinton honest?’ and ‘Is Gore honest?’. Perhaps unsurprisingly, for the first question, the rate of affirmative responses was 50% and for the second one 68%. When for other participants the same questions were presented in the reverse order, the affirmative responses dropped to 60% for Gore and increased to 57% for Clinton. The size of these differences certainly appears shocking, given the importance opinion polls can have in public life. Order effects can also appear when assessing evidence relating to a hypothesis, including, perhaps worryingly, in diagnosis tasks with participants as medical trainees (Bergus et al., 1998) and in a jury decision making task (McKenzie, Lee, & Chen, 2002; Pennington & Hastie, 1986; Trueblood & Busemeyer, 2011). As another example, making a choice can influence preference for the choice. For example, Sharot, Velasquez, and Dolan (2010; Ariely & Norton, 2008; Brehm, 1956) showed that a blind choice between two options influenced post-choice preference for the options, but not when the choice was dictated by a computer. In other work, there has been evidence that in tasks based on cues or factors, decisions alter the way the cues or factors are perceived, so as to make them more consistent with the decisions. For example, Glöckner et al. (2009) demonstrated what they called coherence shifts. These are changes in subjective cue validities related to a decision, in a direction indicating greater consistency with the decision. Likewise, in a complex hypothetical legal case, Holyoak and Simon (1999) showed that the evaluation of arguments changed to become more consistent with the produced verdict and Simon et al. (2001) generalized this finding under a variety of conditions. But it is not only choices or judgments which can affect subsequent behavior. For example, in affect labelling, expressing an emotion can attenuate this emotion in subsequent statements (Torre & Lieberman, 2018).

Such results can be partly summarized under the labels of question order effects and constructive influences of judgments. They are interesting because they challenge baseline intuitions for objectivity in human judgments. We might think that questionnaire responding should reveal underlying views or attitudes, regardless of preceding questions, and that assessment of evidence should be independent of the order in which that evidence is considered. These are reasonable intuitions in general, but especially so for judgments involving experts, as in the study of Bergus et al. (1998).

The present focus is a finding in this category. Across several experiments, White et al. (2014, 2015) considered pairs of stimuli (e.g. advertisements for smartphones or faces of celebrities) of opposite valence (positive, P, or negative, N; trustworthy or untrustworthy). The second stimulus would always be rated and White et al. (2014, 2015) examined this rating, depending on whether the first stimulus was rated as well (double rating condition) or not (single rating condition). The double rating condition consistently led to an increase in the rating intensity for the second stimulus compared with the single rating condition. That is, consider an affective evaluation task in which stimulus pairs could either be positive then negative (PN) or negative then positive (NP). Then, in the PN condition, the rating of the second stimulus was more negative when the first stimulus was rated as well compared to when it was not rated.

Similarly, in the NP condition an intermediate rating of the first stimulus led to a more positive evaluation for the second stimulus. We can call this finding the Evaluation Bias (EB). We call this effect a 'bias' because evaluating the first stimulus biases the judgment for the second stimulus. The default expectation is that the judgment for the second stimulus should not depend on whether the first stimulus is evaluated or not.

The EB is a surprising finding, since just the difference between evaluation vs. observation for the first stimulus can robustly lead to a more intense impression for the second stimulus for the same participant responding twice to exactly the same stimulus. The apparent importance of the intermediate evaluation contrasts with research showing that just observing a stimulus should be sufficient for the automatic formation of an impression of the stimulus's affective valence, independent of other cognitive processes, without fully processing the features of the stimulus, and regardless of the familiarity of the stimulus (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Damasio, 1994; Duckworth, et al., 2002; Fazio, et al., 1986; Greenwald, et al., 1989; LeDoux, 1996; Zajonc, 1980). Clearly, the intermediate evaluation changes something, even though affective information from the first stimulus would be available just from observing the first stimulus.

The EB has been replicated a number of times (White et al., 2014, 2015). There exists familiar terminology to characterize questionnaire decision biases. Recency vs. primacy refers to whether the first or the last question has more influence. Assimilation vs. contrast describes whether the relative influence of two questions is one of convergence or divergence. In general, we prefer to not employ such terms for the EB, because their use depends on which condition (the single or the double rating one) is considered the baseline. For convenience, below we sometimes employ such terms when the context makes one condition the natural baseline.

Instead, the EB is essentially an effect of evaluation. The theoretical challenge is to explore how existing theory might offer guidance regarding possible explanations for the EB. We consider several potentially relevant ideas.

First, a drive to reduce cognitive dissonance could lead to constructive influences and coherence shifts in judgment (Festinger, 1957; Glöckner et al., 2009; Sharot et al., 2010). For example, when making a choice, cognitive dissonance could arise from tension or regret from having to abandon some of the original available alternatives. The possibility of constructive influences is not limited to a cause from cognitive dissonance, they could arise in alternative ways.

Second, for stimuli evaluated in a fixed order, evaluating an earlier stimulus could impact on the processing of subsequent ones, because of recognition or fluency processes. For example, evaluating a stimulus could lead to increased availability of information about the stimulus (Goldstein & Gigerenzer, 2002; Lewandowsky & Smith, 1983), which could in turn affect the familiarity and fluency of subsequent ones (Allport & Lepkin, 1945; Schwarz et al., 2007, for an overview). In turn, increased fluency can influence evaluation, but note that the direction of this influence has been inconsistent across studies (Sanna et al., 2002).

Third, the inclusion/exclusion model (IEM; Schwarz & Bless, 1992; Bless & Schwarz, 2010) seeks to explain context effects in feature-based evaluative judgements. According to the model, the way previous information is used at the time of a judgement will lead to one of two effects. One effect is assimilation, when the information is used to form a representation of the target. The other effect is contrast, when the information is used to form a representation of a standard against which the target is compared.

Fourth, classical probability theory (CT) has provided an influential framework for decision making (Oaksford & Chater, 2009; Tenenbaum et al., 2011). Regarding question order effects, if we denote as A, B the events corresponding to yes to questions A, B, then the probability of A and then B is given by $Prob(A)Prob(B|A)$. But, $Prob(A)Prob(B|A) = Prob(A\&B) = Prob(B)Prob(A|B)$, because conjunction in CT is commutative. That is, baseline CT cannot inform order effects However, one could extract question order effects from CT through appropriate conditionalization, so that the probability of A and then B is written as $Prob(A\&B|order\ 1)$, which can be different from $Prob(A\&B|order\ 2)$. An analogous approach could be adopted for constructive influences. As an aside, note that many of the so-called probabilistic fallacies can be made consistent with baseline CT, in ways analogous to the above.

Finally, an influential idea concerning question order effects is that earlier judgments reveal thoughts or perspectives which can affect later ones (Asch, 1946; Schwarz, 2007; Wang & Busemeyer, 2013). Partly based on this idea, Hogarth and Einhorn (1992; McKenzie et al., 2002) developed their belief-adjustment model, according to which a belief state is adjusted based on an initial anchor and the subsequent pieces of information which are encountered. The model distinguishes between whether there is an end of sequence (EoS) or step by step (SbS) consideration of the evidence. In an EoS process, a single judgment is made after all evidence has been presented; the belief state changes only once, after this single judgment. In a SbS process, a judgment is made after each piece of evidence; the belief state changes after each judgment. Therefore, Hogarth and Einhorn's (1992) incorporates an assumption of constructive influences, that is, changes to the belief state as a result of judgments.

The diversity of this literature reveals a theoretically rich landscape, but also a challenge in terms of a degree of interchangeability between various ideas. For example, question order effects could reflect contextuality, activated thoughts (cf. Asch, 1946), constructive influences (cf. Sharot et al., 2010) or any combination of these effects. Within existing empirical results discriminatory conclusions can be hard. One advantage of the EB is that the simplicity of the paradigm makes it easier to exclude certain viewpoints. This, in turn, enables a focus on the essential idea that can provide a satisfactory explanation.

It is hard to explain the EB in terms of cognitive dissonance, because there is no overt link between the two stimuli (participants are not told that the stimuli are organized in pairs and each stimulus is presented independently of the others). Likewise, because the stimuli are presented as independent, it is hard to see how the EB could arise from coherence shifts (Glöckner et al., 2009), since these assume a set of judgments contributing together towards a hypothesis. However, the EB could reflect a constructive influence arising from the judgment of the first stimulus. That is, the first stimulus could just drive a corresponding change in the opinion; while this idea appears appealing, it is clearly incomplete without further elaboration.

Availability or fluency accounts are also challenged by the EB. There has been some debate on how fluency impacts on later judgments (Sanna et al., 2002). Nevertheless, in the EB paradigm, a reasonable approach would be as follows. Consider the PN condition. The evaluation of the second stimulus might depend on ease of processing of the features of this stimulus. Which features benefit from ease of processing would be affected by which features were primed when processing the first stimulus, since evaluation of the first stimulus would increase the memory strength and availability of its features. Putting these assumptions together, if the first stimulus is evaluated, there will be more P features available when considering the second stimulus, compared to when the first stimulus is not evaluated. That is, in the PN condition, evaluating the first stimulus will increase fluency for any positive features the second, N stimulus has, thereby making the second stimulus appear more positive. This line of reasoning predicts an effect opposite to the EB. However, it has to be said that without a more formal approach to fluency the above line of reasoning is not watertight.

The IEM can allow a prediction about the difference between the single and double rating condition if the participant's cognitive representation of the first stimulus is different depending on whether or not it was rated. According to the IEM (Schwarz & Bless, 1992; see also the set/reset model; Martin, 1986; Martin & Shirk, 2007), the perceived accessibility, representativeness and relevance of the first stimulus will determine whether it is either included in the cognitive representation of the target leading to assimilation or used to construct a representation of the standard against which the target is compared leading to contrast. Primed concepts are included in a representation of a target, which leads to assimilation effects, when the participant is not aware of their influence. When the participant is aware of the potential influence of a prime, then the primed concepts are excluded from the representation of the target, leading to contrast effects. For example, Mussweiler and Neumann (2000) observed contrast effects for externally provided primes and assimilation effects for internally generated primes, because, they argued, the external primes were more obviously a potential source of contamination, with respect to the subsequent judgement about an ambiguously described person. The key point in these and similar studies, as Clore and Colcombe (2003) noted, is that it is the participant's attributions regarding the source of information that determines whether a contrast vs. assimilation effect is observed.

There are obvious differences between the methodology used in White et al.'s (2014, 2015) experiments and those used in both subliminal and blatant priming studies. Typically, in the priming experiments, prime and target are related, the target is often ambiguous and the priming task is different from the subsequent judgement task for the target stimulus. In White et al.'s (2014) experiments, the first and second stimuli were unrelated, they were unambiguously positive or negative, and the judgement task was identical for all stimuli. With a degree of stretching the original IEM ideas, one could propose that the first stimulus impacts on the second one in a way that makes the participant less aware of this influence in the single rating condition; but in the double rating condition, rating the first stimulus makes the participant more aware of the influence, producing increased contrast. So, the model of White et al. (2014) could be seen as a formalization of IEM ideas, as applied specifically for the EB paradigm.

Fourth, considering a baseline CP account for the EB, even though we can write $Prob(second\ measurement\ positive\ |first\ stimulus\ measured) \neq$ $Prob(second\ measurement\ positive\ |first\ stimulus\ not\ measured)$, this conditionalization does not allow us to distinguish between the EB (the first probability less than the second) and an effect opposite to the EB (the first probability higher than the second). The problem with applying CT to the EB and to constructive judgments generally, is that in baseline CT there is no native mechanism for incorporating the role of judgments. In CT, probabilities reflect epistemic uncertainty, so a judgment or evaluation is assumed to reveal what is already true. To accommodate a constructive influence one would need to postulate some additional mechanism on top of the baseline CT process (Pothos & Busemeyer, 2013). It is important to note that cognitive modelers employing CT have been pursuing elaborations better suited to the study of cognition, e.g., incorporating linguistic and pragmatic influences (Goodman et al., 2015) or bounded rationality considerations (Lieder & Griffiths, 2019). It is possible that these more sophisticated approaches would be able to account for order effects and constructive influences in more natural ways.

Finally, the EB appears consistent with the idea that evaluating the first stimulus creates a context that is different from when just observing it (Asch, 1946; Schwarz, 2007), but any attempt along such lines will need to be reconciled with evidence regarding automatic generation of affective information (e.g., Damasio, 1994; Zajonc, 1980). That is, any explanation based on contextuality will need to consider how the process of making an evaluation differs from observing it. Perhaps more promising is Hogarth and Einhorn's (1992) belief-adjustment model, since this is a formal model which incorporates a constructive influence. A key objective of the model was to provide a systematic attempt to organize question order effects in a single framework. However, the EB cannot obviously be an order effect since it concerns a pair of stimuli presented in the same order. The distinction in Hogarth and Einhorn's (1992) model between SbS and EoS processes fits well with the EB experimental paradigm. The SbS process can be thought of as equivalent to the double rating condition, because in the latter there is a judgment after each stimulus. Analogously, the EoS process is equivalent to the single rating condition, because in the latter two stimuli are presented, but a judgment is made only at the end. A potential complication with these analogies is that Hogarth and Einhorn (1992) considered pieces of evidence all bearing on a single final hypothesis. By contrast, in the EB paradigm stimuli were presented independently of each other – participants were never given any indication that the judgment for one stimulus should impact on that for another. We will ignore this complication and simply assume that Hogarth and Einhorn's (1992) model is applicable to the EB paradigm.

Hogarth and Einhorn’s (1992) review of the relevant literature creates a confusing picture regarding the EB. They noted that EoS can induce primacy, and SbS can induce recency. This statement appears consistent with the EB. The double rating condition is analogous to SbS and the EB is analogous to a recency effect, if we consider the single rating condition the baseline. However, they also noted that recency is associated with more complex tasks and the paradigm employed for the EB is extremely simple by the standards in Hogarth and Einhorn’s (1992) review.

The application of Hogarth and Einhorn’s (1992) model to the EB paradigm can be approached in different ways. White et al. (2014) carried out one analysis and showed that, under fairly benign assumptions, the model cannot predict the difference between the single and double rating conditions, corresponding to the EB. This was because only two questions are involved; Hogarth and Einhorn’s (1992) model was designed to deal with longer sequences of pieces of information. In the present work, we adopted an alternative approach with Hogarth and Einhorn’s (1992) model, which allows a prediction of constructive influences in the EB paradigm (Appendix 1). We believe this is the first formal application of the model to constructive influences. As we shall see, the success of the model provides an encouraging message regarding the enduring relevance of this classic formalism.

In summary, the EB challenges the applicability of several of the predominant approaches. We have seen how the emergence of the EB as a result of a prior evaluation vs. observation perhaps suggests a constructive influence or contextual effect, but clearly there is a need to further formalize such ideas to unambiguously predict the EB in the observed direction. White et al. (2014) approached this challenge in a technical way, by adopting a modeling framework which requires constructive influences as a result of measurements or judgments. Our objective is to develop this framework and so seek deeper understanding of the EB and the corresponding cognitive principles.

1.2 Quantum theory

We call quantum probability theory (QT) the probability rules from quantum mechanics, without the physics. It is a formal framework for probabilistic inference, much like CT, but based on different axioms (Busemeyer & Bruza, 2011). QT is relevant in the study of constructive influences because a fundamental principle of QT is that a judgment has to alter the underlying state in a certain way. More generally, QT cognitive models have been pursued in cases where behavior appears at odds with baseline CT prescription, e.g., when probabilities appear contextual (Aerts, 2009; Busemeyer & Bruza, 2011; Haven & Khrennikov, 2013; Pothos & Busemeyer, 2013), such as in the case of questionnaire order effects (Trueblood & Busemeyer, 2011; Wang et al., 2014). QT decision models have nothing to do with the controversial ideas regarding a quantum brain (e.g., Litt et al., 2006; Khrennikov et al., 2018) – QT simply provides a coherent set of computational-level principles for cognitive modelling, not unlike CT. Finally, a one sentence introduction to QT is that it is just like CT, but instead of having a single, all-inclusive space of events, events are separated into different 'partitions': within each partition, probabilistic inference is fully classical, but across partitions effects arise which appear as classical errors (Lewandowsky & Kirsner, 2000, Lewandowsky et al., 2002).

We consider the main elements of QT: the state vector, Hilbert space, subspaces, the probability rule, projection, rotation, and the collapse postulate.

In QT, the system is represented by a normalized vector in a Hilbert space. In psychological applications, the system is typically the mental state of a participant prior to going through an experimental manipulation. A Hilbert space is a complex vector space, with some additional convergence properties. A subspace in a Hilbert space is a part of the overall space. For example, in Figure 1, the overall space is two-dimensional. In this two-dimensional space, we can have one-dimensional subspaces, called rays. Subspaces represent question outcomes. For example, in Figure 1, we could ask whether a hypothetical person, Jane, likes a particular font for her party invitations. Both the overall Hilbert space and subspaces can have varying dimensionalities, depending on the complexity of the experimental situation and relevant questions. A fundamental aspect of QT is how to associate probabilities to subspaces. This is done via the Born rule, according to which the probability of producing different outcomes when responding is equal to the squared length of the projection of the state vector to the corresponding subspaces. For example, in Figure 1, assume that Jane is considering the font indicated by Stimulus 1, so that the mental state is represented by a vector along the Stimulus 1 ray. Is she likely to like or dislike Stimulus 1? We ‘lay down’ (project) the Stimulus 1 state vector onto the Like subspace. This projection is indicated by the perforated line. The squared length of this line is the probability we seek. That is, the probability of particular outcomes depends on the *overlap* or projection between the mental state and the corresponding subspace, so that greater overlap implies higher probability (cf. Sloman, 1993). A surprising theorem shows that there is only one consistent way to associate probabilities to events (Hughes, 1989). The probability rule and the associated mathematical theorems warrant the label ‘quantum’, rather than a label along the lines of ‘projective geometry’.

We now reach the key consideration for the present work. The mental state vector can change in two ways. First, when the participant is presented with new information, the vector is rotated in a corresponding way. For example, in Figure 1, if Jane encounters a font she does not like, the mental state vector will rotate towards the Dislike subspace. The degree of rotation will depend on the strength of the stimulus. Second, when a decision or evaluation is made, the mental state vector has to identify with (be projected to) the subspace of the chosen outcome. This is the fundamental collapse postulate in QT. In physics, the collapse postulate has been puzzling, since in some cases elementary particles cannot be said to have *any* properties, such as position or momentum, prior to a measurement. In psychology, it is perhaps easier to accept that judgments sometimes alter mental states. Let us consider again Jane in Figure 1. She considers the font represented by the Stimulus 1 vector. Suppose she is asked whether she likes Stimulus 1 and she answers yes. Then, the state vector will now become a normalized vector along the Like subspace. Therefore, whether Jane makes this judgment or not will impact on subsequent decisions, such as whether she likes another font, indicated by Stimulus 2 in Figure 1.

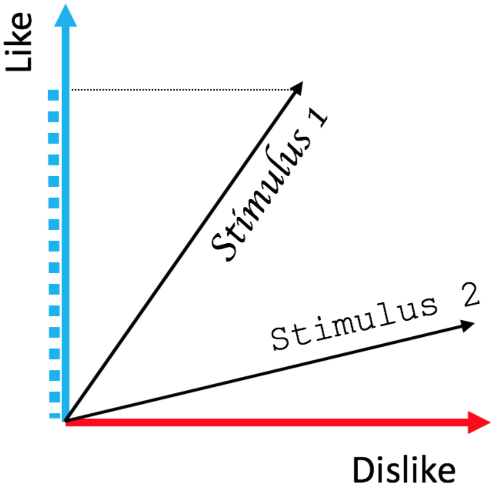

Figure 1. An example of how answering a question alters the mental state. Assume the mental state vector is indicated by Stimulus 1. Then, when answering a question of whether Stimulus 1 is liked or not, the mental state vector is projected along either the Like or Dislike subspace – in the example, we assume the former. The new state vector will be a normalized vector along the Like subspace.

Psychologically, the closest interpretation of these ideas concerns constructive influences in decision making. However, the established motivation for constructive influences typically concerns tension between a particular choice and abandoned alternatives (e.g., Festinger, 1957; Glöckner et al., 2009). Instead, in the case of QT *any* judgment forces a constructive influence on the mental state, as long as the mental state does not already identify with a response outcome (e.g., perfect liking in Figure 1). We also believe the IEM (Schwarz & Bless, 1992) could be formalized in QT terms if it is possible to determine how the difference between making a judgement vs. just perceiving a stimulus maps onto whether the relevant information affecting subsequent judgments is overt vs. covert.

The application of QT to questionnaire responding entails that multiple responses lead to multiple changes in the mental state, with potential for generating systematic response biases (Kvam et al., 2015; Trueblood & Busemeyer, 2011; Yearsley & Pothos, 2016; Wang & Busemeyer, 2013; Wang et al., 2014). Regarding the EB, White et al. (2014, 2015) considered pairs of oppositely-valenced stimuli and the impact of evaluating the first on the evaluation of the second vs. the impact of just viewing the first on the evaluation of the second. They assumed that introducing a stimulus would lead to a change in the mental state towards the corresponding definite response, but evaluating the stimulus would entail an additional change corresponding to the identification of the mental state with the response outcome (Section 1.3) – the first assumption is shared by most dynamical models, but the second one is characteristic of QT.

To summarize, the motivation for considering a QT model for the EB is that it constrains constructive influences from judgments or evaluations to have a specific form. This removes ambiguity regarding whether the EB is predicted to reflect increased or decreased intensity as a result of the intermediate judgment. That is, the EB is a prediction from QT when the stimuli conform to a specific configuration.

1.3 The Evaluation Bias and the quantum model, with general data

White et al. (2014, 2015) demonstrated the EB with stimuli conforming to particular valences, P or N; pilot studies were employed to determine stimulus valence or the stimuli were selected from image libraries with pre-established affective valence. For such stimuli, a schematic application of QT suffices to predict an EB in a straightforward way. In the present case, we wish to apply the same ideas for more general stimuli, as would be likely to be encountered in real contexts.

To fully formalize the QT model for the EB, the first step is to define the state vector, subspaces, and rotations. We do so in terms of a distinction between P and N affect, as in White at. (2014), but clearly the model can apply to any analogous bivalent distinction. As shown in Figure 2, we assume a two-dimensional Hilbert space, so that the rays for P, N affect, the stimuli, and the mental state are all co-planar. The higher the probability of deciding that the stimulus is e.g. N, the more negative the rating for the stimulus.

Consider first the PN condition. Since the first stimulus is P, the initial mental state can be represented with a state vector close to the P subspace, ψ_p. In the single rating condition, introducing the N stimulus leads to a rotation towards the N subspace, so that the mental state becomes $\psi_p{}'$. Then, the rating of the second stimulus depends on the overlap between $\psi_p{}'$ and the N affect subspace. In the double rating condition, the evaluation of the first stimulus makes it likely that the mental state will collapse onto the P ray, so that the new mental state will be a normalized vector along the P ray. Then, when introducing the second stimulus, the mental state is rotated by the same amount as before, to ψ_p'' (cf. Stewart et al., 2005). It can be seen that ψ_p'' is a little closer to the N affect subspace; the judgment for the first stimulus is equivalent to an additional rotation towards the N affect subspace. Therefore, in the double rating condition the second stimulus will be judged as more negative than in the single rating condition. The situation for the NP condition is analogous and shows an EB as a more positive evaluation for the second stimulus, as a result of evaluating the first one. Figure 3 further illustrates the EB, in the PN case.

Why should we place the initial mental state ψ_p in the top left vs. top right quadrant? This is arbitrary and makes no difference to the eventual result. In the PN condition, why should the rotation be clockwise vs. anti-clockwise? We assume that the direction of rotation is the same as the direction of the more likely projection.

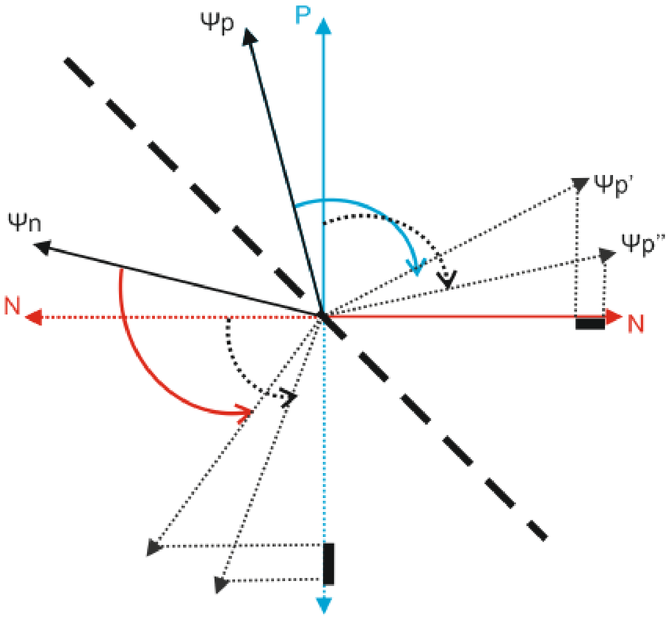

Figure 2. A diagram of the QT EB model. The EB in the PN, NP conditions are indicated by the thick black lines along the N and P rays respectively.

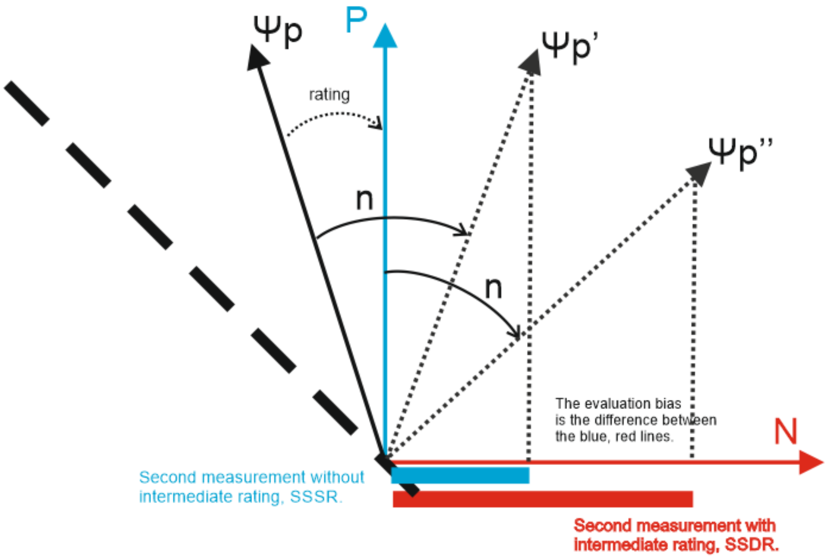
Ψp
P
Ψp'
rating
Ψp''
n
n
The evaluation bias
is the difference between
the blue, red lines.
N
Second measurement without
intermediate rating, SSSR.
Second measurement with
intermediate rating, SSDR.

Figure 3. A more detailed illustration of the EB occurring in the PN direction and the assumed QT model representations.

To formalize the QT model, we require a few technical elements and notation. First, a projection operator is a linear operator which takes a state vector and projects it to some subspace. For the P affect, N affect subspaces, the corresponding projection operators will be denoted as P_P and P_N. Then, the projection of the state vector onto e.g. the N affect subspace is $P_N \psi$. Recall, probability is length of projection squared. Therefore, $Prob(N;\psi) = |P_N \psi|^2$. Second, rotations of the mental state vector are computed using unitary operators. In the present case, it suffices to employ $U(n) = \begin{pmatrix} \cos n & \sin n \\ -\sin n & \cos n \end{pmatrix}$, which implements a clockwise rotation of angle n. Define $Perfect_P$ and $Perfect_N$ to be normalized vectors along the P, N subspaces respectively. The introduction of the second stimulus rotates the state vector clockwise if the first stimulus is more likely to be considered P and anti-clockwise if the first stimulus is more likely to be considered N. Also, $\psi_P = U(-rating) \cdot Perfect_P$, that is, the *rating* angle is used to set the initial mental state vector in the top left quadrant. For a different *rating* angle, we will have

$\psi_N = U(-rating) \cdot Perfect_P$. Finally, we introduce the notation FSDR, SSSR, and SSDR, which respectively stand for the ratings concerning first stimulus double rating condition, second stimulus single rating condition, and second stimulus double rating condition; in all cases, a higher value means a more positive rating.

The basic equations of the QT EB model are then, for the NP condition:

$Prob(FSDR; \psi_P) = |P_P \cdot U(-rating) \cdot Perfect_P|^2$………………..………………..Equation 1a

$Prob(SSSR; \psi_P) = |P_P \cdot U(n) \cdot U(-rating) \cdot Perfect_P|^2$…………………….….Equation 2a

$Prob(SSDR; Perfect_P) = |P_P \cdot U(n) \cdot Perfect_P|^2$………………………………..Equation 3a

The basic equations for the PN condition are:

$Prob(FSDR; \psi_N) = |P_P \cdot U(-rating) \cdot Perfect_P|^2$…………………..…………..Equation 1b

$Prob(SSSR; \psi_N) = |P_P \cdot U(-n) \cdot U(-rating) \cdot Perfect_P|^2$………..……….…..Equation 2b

$Prob(SSDR; Perfect_N) = |P_P \cdot U(-n) \cdot Perfect_N|^2$……….……….…………..Equation 3b

These equations assume that the result of rating the first stimulus in the double rating condition is very likely to lead to state $Perfect_P$ in the PN condition and $Perfect_N$ in the NP condition. Unfortunately, in the present case we cannot make this assumption. In Equations 2a, 2b there may be sizeable probabilities to a projection to $Perfect_P$ and $Perfect_N$ regardless of condition. In Equations 3a, 3b likewise there may be sizeable probabilities for rotations characteristic of a PN vs. NP sequence regardless of condition. The equations of the extended QT EB model are:

$Prob(FSDR; \psi) = |P_P \cdot U(-rating) \cdot \psi_P|^2$…………………..…………………..Equation 1c

$Prob(SSSR; \psi) = |P_P \cdot U(n) \cdot U(-rating) \cdot \psi_P|^2 \cdot Prob(FSDR; \psi) + |P_P \cdot U(-n) \cdot U(-rating) \cdot \psi_P|^2 \cdot \big(1 - Prob(FSDR; \psi)\big)$………….……………………..Equation 2c

$Prob(SSDR;\psi) = |P_P \cdot U(n) \cdot Perfect_P|^2 \cdot Prob(FSDR;\psi) + |P_P \cdot U(-n) \cdot Perfect_N|^2 \cdot (1 - Prob(FSDR;\psi))$……………………………………………………….Equation 3c

For example, Equation 2c is a weighted mean of a process assuming the first projection would have been to P and then clockwise rotation towards the N subspace and a process assuming the first projection would have been to N and then anticlockwise rotation towards the P subspace; the weights are determined by the probability that the first projection is to P or N.

We consider a few final issues regarding the application of the model. First, participant ratings were transformed to probabilities through a linear function, so that lower ratings would correspond to lower probabilities for a P evaluation; that is, all probabilities correspond to probabilities that a stimulus is evaluated positively. An alternative approach would have been to use a softmax function (e.g., as in Rehder, 2014, or Trueblood et al., 2017) or relevant extensions to map ratings to probabilities, but this was not considered necessary in the present work. Second, for each participant in the present experiments, we obtained three datapoints, FSDR, SSSR, and SSDR. The QT EB has two parameters per pair of questions (per participant), *rating* (how positively the first question is perceived) and *n* (the impact of the second stimulus on the mental state vector). With Equation 1c, we extract the *rating* parameter; with Equation 2c and the *rating* parameter, we extract the *n* parameter; then, Equation 3c produces the QT prediction for the EB. That is, it is *not* the case that the two available parameters are adjusted to simultaneously fit each group of three data points. Once *rating* and *n* are determined from Equations 2a, 2b, there is no parameter manipulation to improve fit between SSDR and the corresponding QT prediction. Finally, with Equations 1c, 2c, 3c it is no longer necessary to distinguish between PN and NP processes. However, it is advantageous to continue doing so, partly because we think that practical applications are more likely to be driven by approximate, summary predictions regarding the EB, instead of detailed model fits.

With stimuli which are not pre-controlled, we need to specify conditions on the stimuli which can allow classification into PN vs. NP conditions. Both the quantum model and Hogarth and Einhorn's (1992) model suggest that a triplet of judgments {FSDR, SSDR, SSSR} should be assigned to the PN condition if FSDR>SSDR and to the NP condition otherwise. For the quantum model, this condition will consistently produce the EB. Consider first that $Prob(FSDR) \propto \frac{1}{rating}$ and $Prob(SSDR) \propto \frac{1}{n}$, which means that $FSDR > SSDR \Leftrightarrow n > rating$. As shown in Figure 3, the condition $n > rating$ means that ψ_P and ψ_p' are on either side of the P ray, so that $Prob(SSSR) \propto \frac{1}{|n-rating|} \Leftrightarrow SSSR > SSDR$, the latter inequality being the EB in the PN direction. That is, starting from $FSDR > SSDR$ we were led to the EB, $SSSR > SSDR$. For Hogarth and Einhorn's (1992) model, the weights for the SbS process depend on the relative size between the impression from the current stimulus, $s(x_k)$, and the belief state prior to the current stimulus, $R = S_{k-1}$ (see also below). As the model's form depends on the outcome of successive judgments, it makes more sense to use FSDR, SSDR for the assignment of triplets into the PN vs. NP conditions.

Does the assignment of participants into PN vs. NP conditions confound the study of the EB? Consider PN participants, for whom FSDR>SSDR. The EB prediction is that SSSR>SSDR, that is, that the second stimulus in the double rating condition should be more negative (lower) than the rating for the second stimulus in the single rating condition. Knowledge that FSDR>SSDR indicates that the second stimulus will be broadly more negative than the first. However, this knowledge cannot further inform whether SSSR>SSDR or SSSR<SSDR.

Finally, we opted to study the EB in the context of surveys for gathering employee impressions of their organization regarding culture, leadership and performance. Organizational surveys are a suitable choice for several reasons. First, such surveys are commonly employed and often guide organizational communication and strategy. So, the discovery of systematic biases in organization surveys is valuable. Second, questionnaires are also commonly employed in health and clinical settings, but such applications entail unnecessary complications (e.g., regarding participant recruitment). Finally, we had opportunistic availability of large samples of professional respondents, thus avoiding problems with restricting sampling to college students or online participants. With a degree of optimism, one might expect that individuals (managers and professionals) going through an organizational survey about their own organization would provide more thoughtful responses than participants responding to questions involving fictitious stimuli for a small payment (cf. Camerer & Hogarth, 1999, for failures of incentivizing to restore unbiased decision making in indifferent participants).

2. Experimental work

2. 1 Participants and design for all experiments

We carried out four replications of the same design, across different times, geographical locations, and in one case involving a different language. In all cases participants were managers and other senior professionals employed by organizations, responding to questions about their organization.

We describe the participants and design details of all replications together, then consider the common design of all experiments. We then describe the results of the first experiment in detail and relegate detailed descriptions of the results for the other experiments to the Supplementary Materials section). Finally, we fit the QT EB model to the combined dataset.

Participants in Experiment 1a (N=240 from 25 different organizations) were recruited through managers attending an executive education program in October 2014 at an international business school in Europe. Attendant managers were asked to distribute an online questionnaire to employees in their respective organizations. These employees would become the participants in this study. The questionnaire was in English. Participation was voluntary for the managers on the program and the employees in their organization. The survey was anonymous with no personally identifying information collected. This was done to encourage honest responses from employees in the organizations. Organizations were a mixture of industry types (e.g. financial services, manufacturing, pharmaceutical, public sector) and from various countries in Europe, the Middle East and Asia.

Participants in Experiment 1b were recruited through managers attending an executive education program in April 2015 (N=193, 17 different organizations) at an international business school in Europe.

Participants in Experiment 1c were recruited through managers attending an executive education program in September 2015 (N=140, 15 different organizations) at an international business school in Europe.

Participants (N=295, 27 organizations) in Experiment 1d were recruited in a similar manner through managers attending an executive education program in the same business school in September 2015. Organizations were a similar mix of types, as in Experiment 1a, except that all were based in Brazil. The questionnaire employed in Experiment 1d was the same as in the other experiments, but translated into Portuguese.

In all cases, because of the format of the data collection, sample sizes were opportunistic; essentially we included all participants that were available to us from the education programs. Prior to data collection, we were intending to reject samples smaller than 54 participants, since this was the sample size in the original laboratory EB demonstration (White et al., 2014; Experiment 1). But this turned out to never be the case.

The experimental design was mixed with two main independent variables. A between subjects factor, stimulus valence order, had three levels: positive-negative (PN), negative-positive (NP) or equal (EQ). Rating condition was within subjects and had two levels, whether or not the rating of the second question was preceded by an intermediate rating for the first question (double) or not (single).

2.2 Materials and Procedure for all experiments

The questionnaire consisted of 94 questions and asked respondents for their views on various aspects of their organization's strategy, leadership and culture. Of relevance, participants would be presented with a short statement about their organization. In the double rating condition they were asked to think about the general state of their organization (see below) and then asked to respond to a follow-up question about the organization. In the single rating condition they were again asked to think about the general state of their organization but were not asked the follow-up question. Subsequently, in both conditions, they were asked about the strategy of their organization. The strategy question was always presented last, right after the organization one. By analogy with White et al.'s (2014) notation, P, N refer to high and low ratings to these questions and PN to the situation when the rating for the first question was high and for the second low. The two questions were:

Organization:

“Think about the general state of your organization e.g., its overall performance, any financial pressures the organization is facing, the demands of customers, challenges from competitors, changes in technology, political or regulatory issues, market volatility.”

[Organization question] “How do you feel about the general state of your organization?”

Strategy:

[Strategy question] “How do you feel about your business unit's strategy?”

All participants received both the double and single rating conditions (randomized order, across the organizations participating in the survey). In between the two pairs of questions, there were 88 other questions, irrelevant to the present analysis. Thus, the double vs. single rating condition was a within participants manipulation with each participant answering the question about strategy twice. Note, all questions were answered on a seven point Likert scale (very negative to very positive).

There are several considerations which guided the selection of these questions for testing the EB and the QT model. First, the first judgment must have the potential to alter our perspective for the second judgment. Second, the questions must embody a degree of ambiguity. Without some ambiguity, no constructive or QT effects are expected. For example, if participants see a hammer and are asked if there is a hammer, no changes to the mental state are really expected. Note that there is a converse point here, namely that with too much ambiguity the representational assumptions embodied in the QT model are challenged – we take up this point again in the Discussion. Third, the two questions must be broadly thematically related; this relates to the assumption that all subspaces and rotations are restricted to the same two-dimensional space. Finally, the two questions must themselves be presented in an independent way. With the current design, participants were asked each question without reference to the previous one.

Following the coding approach we motivated in Section 1.3, in the double rating condition, if a participant's response to the first question was greater than their response to the second, in the double rating condition, we assigned him/her to the PN condition and vice versa. Note, this assignment was not based on the degree of positivity or negativity, only on the rating of the second question relative to the first. For example, if someone rated the first question 7 and the second 6, although both responses could be seen as positive (i.e. at the high end of the rating scale), the first was less positive than the second and therefore that participant would be allocated to the PN condition. If a participant's response to the two questions was equal, they were assigned to the equal condition (EQ), and this served as a control condition, for which no EB is expected.

2.3 Results for Experiment 1a

The statistical approach is identical for all experiments[1]. First, we conducted a methods check, since we could not directly manipulate the P, N status of different questions. Second, we examined the evidence for the EB, that is, the hypothesis that for PN participants the intermediate judgment reduced the rating for the subsequent N statement; and that for NP

[1] Data and code for all analysis including QT model fits and experiments 1b, 1c and 1d can be found here: https://osf.io/se84w/?view_only=fa9e916b151241dcb5b9f0576f73f8d6

participants the intermediate judgment increased the rating for the P statement. Finally, some participants encountered the single rating condition at the beginning of the survey and the double rating one at the end; for the rest of the participants this was the other way round. We checked that the order in which participants completed the single and double rating condition did not influence the difference in ratings. For assessing all empirical results, we used JASP (JASP Team, 2016) to conduct both Bayesian and standard ANOVAs.

The methods check was carried out to verify that the stimuli assigned to the positive and negative groups were responded to differently. The question is whether the post hoc participant assignment to the valence order condition generated some contrast between positive and negative affect, as intended for a test of the EB. The methods check involved a mixed measures ANOVA with one between participants factor (valence order: PN, NP, EQ) and one within participants factor, comparing the rating of the first question vs. the rating of the second question. The main effects of valence order and first vs. second rating were significant ($F(2,237)=4.39$, $p=0.013$, $\eta^2_p=0.04$, $BF_{10}=2.94$; $F(1,237)=4.29$, $p=0.039$, $\eta^2_p=0.02$, $BF_{10}>100$). Our main interest is the interaction, which was significant ($F(2,237)=365.1$, $p<.001$ $\eta^2_p=0.76$, $BF_{10}>100$) indicating, as desired, a difference between the first and second rating, depending on the order in valence. That is, the first rating would be different depending on whether valence order was NP or PN. The results also indicate that the interaction model was preferred to the main effects model ($BF_{10}>100$). Paired samples t-tests showed that, in the PN condition, ratings for the organization question ($M=5.73$, $SD=0.95$) were significantly higher (i.e., the ratings were more positive) than ratings for the strategy question, $M=4.46$, $SD=1.12$; $t(40)=-16.20$, $p<.001$, all t-tests are two tailed; d=-2.53, 95%, CI[2] for the mean difference = [-1.43, -1.11], BF_{10}>100. In the NP condition, ratings for the organization question (M=4.38, SD=1.26) were significantly lower than ratings for the strategy question, M=5.91, SD=0.81; t(96)=17.33, p<.001; d=1.76, 95% CI for the mean difference = [1.35, 1.70], BF_{10}>100. Overall, for both the PN and NP conditions, the Bayes factors indicate strong evidence for the alternative hypothesis that, in the PN condition the first rating was higher than the second and vice versa for the NP condition, consistent with the intended design. In the EQ condition, the ratings for the organization (M=5.53, SD=1.11) and strategy (M=5.53, SD=1.11) questions were identical and no t-test was required.

In order to examine evidence for the EB, we next conducted a mixed measures ANOVA with two between factors (valence order: PN, NP, EQ and order of presentation: single rated first, double rated first) and one within factor (rating condition: single, double) on participant ratings of the strategy question, which was the second question (Figure 4). There was a main effect of valence order (F(2,234)=12.2, p<.001, η^2_p=0.08, BF_{10}>100) but not of rating condition (F(1,234)=2.51, n.s., BF_{01}=6.66). The crucial valence order × rating interaction was significant (F(2,234)=28.98, p<.001, η^2_p=0.20, BF_{10}>100) indicating, as predicted, that the difference in the rating of the second question between the single and double rating conditions depended on NP vs. PN order. The valence order × rating interaction model was preferred to the main effects model (BF_{10}>100). Paired samples t-tests showed that, in the PN condition, with an intermediate rating, ratings for the strategy question (M=4.46, SD=1.12) were significantly lower than those without the intermediate question (M=5.15, SD=1.06; t(40)=-4.44, p<.001; d=-0.69, 95% CI for the mean difference = [-0.99, -0.37], BF_{10}>100). For those participants who behaved according

[2] We report frequentist confidence intervals unless stated otherwise.

to the hypothesis, the impact of the intermediate question on their rating of the strategy question was associated with a decrease in their rating by 1.39 units on average. In the NP condition, with an intermediate rating, ratings for the strategy (second) question (*M*=5.91, *SD*=0.81) were significantly higher than those without the intermediate question (*M*=5.35, *SD*=1.28; *t*(96)=5.43, *p*<.001; *d*=0.55, 95%, CI for the mean difference = [0.35, 0.76], BF_{10}>100). For those participants who behaved according to the hypothesis, the impact of the intermediate question on their rating of the strategy question led to an increase in the rating by 1.47 units on average. In the EQ condition, ratings for the strategy question with an intermediate rating (*M*=5.53, *SD*=1.11) were no different from ratings for the strategy question without the intermediate rating (*M*=5.65, *SD*=1.03; *t*(101)=-1.83, n.s., 95%, CI for the mean difference = [-0.25, 0.01], BF_{01}=1.82). Finally, to determine whether the order of presentation influenced the results we looked at the valence order × rating × order of presentation interaction. The model including this three-way interaction was essentially indistinguishable from an identical model, but without the three-way interaction, BF=0.96. We therefore conclude that order of presentation is not an important variable in the analyses and we do not consider further.

Overall, for both the PN and NP conditions, the Bayes factors suggest strong evidence for the role of the intermediate rating in the response to the second question, in the predicted direction. For the EQ condition the Bayes factor supports the null hypothesis. The results of Experiments 1b, 1c, and 1d replicate this pattern (Supplementary Materials).

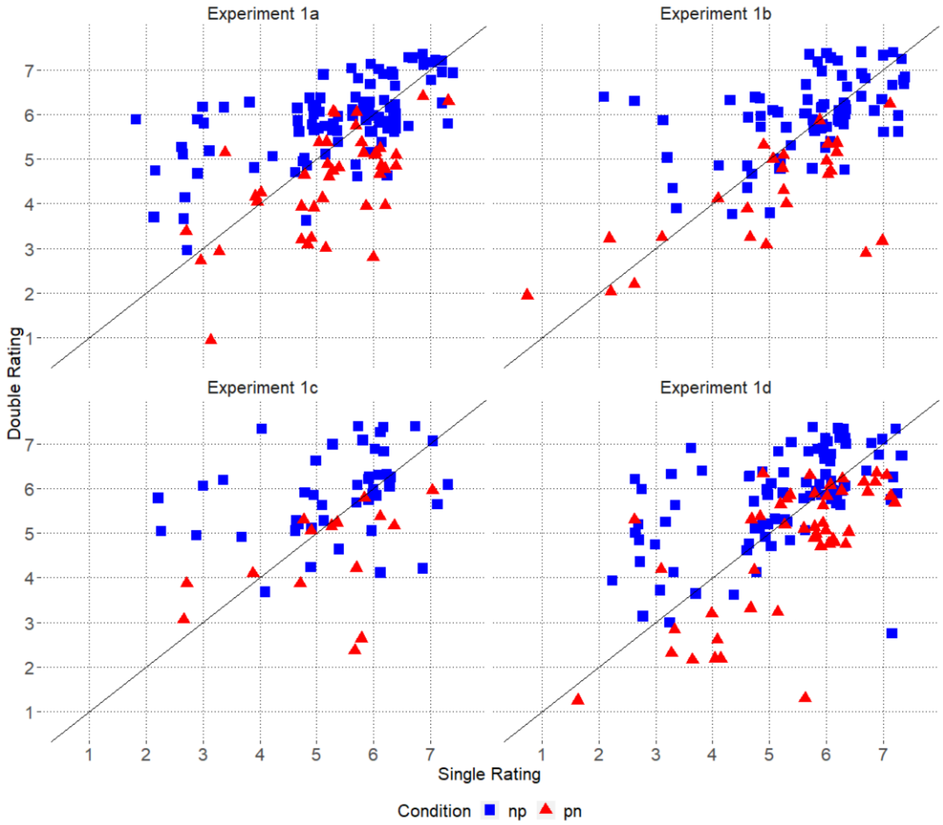

Figure 4. Results for Experiments 1a, 1b, 1c, 1d: Each point represents a participant's single and double ratings for the strategy question (which was the second question) in the NP and PN conditions. Points are jittered to avoid overlap and participants in the EQ condition are omitted for clarity. In the NP condition, if a participant behaves according to the prediction, we expect blue, square points above the black diagonal line and in the PN condition we expect red, triangle points below the diagonal.

These results illustrate how it is possible to identify a reliable EB effect, even when the stimuli/ questions are not controlled a priori, based on a data-driven classification into PN or NP conditions. However, this classification relies on participant ratings, which may be noisy. How robustly can the present approach identify an EB, given that it exists, but in the presence of increasing noise? We considered the means of the second question with (4.46) and without (5.15) the intermediate rating, in the PN condition, with a pooled standard deviation from these conditions. The sample size for this particular condition was 40. We then simulated groups of scores drawn from the *t*-distribution (as the population variance is unknown), based on a fixed pooled standard deviation, fixing one of the means, and offsetting the other mean from its original position across the range [-1,1]. The produced scores were subjected to uniform random noise as a percentage of the Likert scale (the max percentage of noise varied from 0 to 100%). For each combination of mean shift and max possible percentage noise we computed the *p* value of a paired samples *t*-test, analogous to the one computed for testing the EB. As seen in Figure 5, non-significant p-values are observed only when either the means are closer together or with high levels of noise. Note, even though a simple moving average smoothing function was applied, the surface remains bumpy because of the way this simulation was carried out. Overall, an expectation of noisy data does not greatly undermine the prediction of an EB..

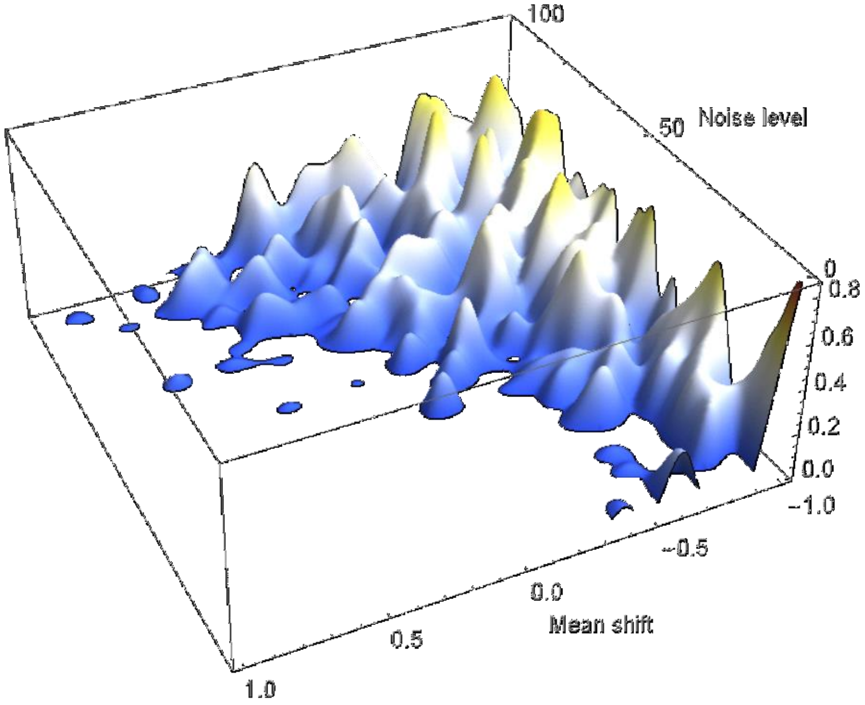

Figure 5. An analysis of when a paired samples *t*-test between the ratings for the second stimulus with and without an intermediate rating fails significance, assuming data analogous to what was observed in the PN condition of Experiment 1a.

3. Model fits

Each response corresponded to a triplet of judgments, {FSDR, SSSR, SSDR}. All 868 responses were collated together; there were 240, 193, 140, 295 participants in Experiments 1a, 1b, 1c, and 1d respectively. Both the QT model and Hogarth and Einhorn's (1992) model distinguish between PN (FSDR>SSDR) and NP processes (FSDR<SSDR). However, for both models there should be an adjustment in predictions as the distinction between a PN process and an NP one becomes less pronounced. We therefore decided to randomly assign to the PN vs. the NP condition data points for which FSDR=SSDR. This approach is reasonable both theoretically and because, as it turned out, there were a large number of data points for which FSDR=SSDR. Specifically, the number of data points which conformed to FSDR>SSDR, FSDR<SSDR, and FSDR=SSDR were respectively 130, 307, 431, for a total of 868, as above. Note that overall we observed a majority of positive responses, e.g., for 647 data points FSDR was greater than the midpoint of the ratings scale. This could relate to the fact that organizations which are performing better are more likely to send their employees to a business school program. Ratings on a 1-7 scale were linearly transformed onto a 0-1 scale, as outlined in Section 1.3.

The application of the QP model follows Section 1.3. Hogarth and Einhorn's (1992) model has been predominantly applied to question order effect, not to the modelling of constructive influences. In Appendix 1 we outline in detail how the distinction between SbS and

EoS processes in Hogarth and Einhorn's (1992) model can be exploited to also predict an EB, under particular parameter settings. We summarize Appendix 1 here.

We define the following variables: S_k $(0 \leq S_k \leq 1)$ is the impression of participants after considering k statements. $s(x_k)$ is the subjective evaluation of the kth advert. $s(x_1, \ldots, x_k)$ is the combined impact of all the statements, statement 1 to k. Because in the present case we are employing a unipolar scale, we can assume $0 \leq s(x_k) \leq 1$ (Hogarth & Einhorn, 1992, p.11). R is the reference point against which the impact of the kth statement is assessed. w_k $(0 \leq w_k \leq 1)$ is an adjustment weight in relation to how the kth statement impacts on the belief state. The main equation of the model (Equation (1) in Hogarth and Einhorn's, 1992, paper) is

$S_k = S_{k-1} + w_k[s(x_k) - R]$……………………………………………………..…..Equation 4

This equation dictates how the new belief state will depend on the previous state, the impression from the new information that was received relative to a reference point, and an adjustment weight. All these details can depend on the format of the evaluation process. Moreover, somewhat extending Hogarth and Einhorn's (1992) original ideas, we can associate the SbS process with the double rating condition and the EoS one with the single rating condition. With some algebra, it can be shown that the set of equations allowing modelling of the EB is, for the PN case:

$SSSR = S_0 + aS_0[s(x_P, x_N) - S_0]$, from the model equation regarding the EoS process…………………………………………………………………………………Equation 5a

$FSDR = (1 - \beta(1 - S_0))S_0 + \beta(1 - S_0)s(x_P)$, which is the first judgment in the SbS process…………………………………………………………………………………Equation 6a

$SSDR = (1 - aS_P)S_P + aS_P s(x_N)$, which is the second judgment in the SbS process…………………………………………………………………………………Equation 7a

The a, β parameters are sensitivity weights capturing the relative impact of positive vs. negative information.

The equations for the NP case are:

$SSSR = S_0 + \beta(1 - S_0)[s(x_N, x_P) - S_0]$, ...Equation 5b

$FSDR = (1 - aS_0)S_0 + aS_0 s(x_N)$..Equation 6b

$SSDR = (1 - \beta(1 - S_N))S_N + \beta(1 - S_N)s(x_P)$.......................................Equation 7b

To reduce the number of parameters, we assume that, in the NP case $s(x_N, x_P) = s(x_P)$ and $S_N = FSDR$ and in the PN case $s(x_P, x_N) = s(x_N)$ and $S_P = FSDR$. In both cases, we can further assume $s(x_P) = 1 - s(x_N)$. Additionally, we can assume that $a = \beta = 1$ (Appendix 1). Given this specification, Hogarth and Einhorn's (1992) model can be fitted to the data in a way closely analogous to that of the QT model. Equations 5a, 6a, 7a (or 5b, 6b, 7b) have two free parameters, $s(x_P)$ and S_0. Notice that $s(x_P)$ is analogous to the *n* parameter in the QT model, since both parameters concern the impact of the second stimulus on the belief/ mental state. Also, S_0 is analogous to the *rating* parameter in the QT model, since both parameters concern the initial belief/ mental state, prior to encountering the first stimulus. Then, we can use the FSDR and SSSR equations to compute the $s(x_P)$, and S_0 parameters, since there are two equations and two unknowns. In practice, solving these equations required sum of squares optimization, because it was sometimes impossible to solve them analytically given the parameter bounds. Given values for the $s(x_P)$ and S_0 parameters, a prediction for SSDR follows. As for the quantum model, we consider this version of Hogarth and Einhorn's (1992) model parameter-free, since we employ part of the data to set the parameters and part of the data to test the resulting prediction.

Models were evaluated with Maximum Likelihood Estimation (MLE), for assessing predicted against observed probabilities. We computed $G^2 = 2\sum_{all\ trials}\left(o_i \ln\frac{o_i}{e_i} + (1-o_i)\ln\frac{1-o_i}{1-e_i}\right)$, where the summation extends across all instances of SSDR judgments in the experiment (e.g., as applied in Broekaert et al., in press). That is, model fit is based on the correspondence between predicted and observed SSDR values. Then, models were compared on the basis of the Bayesian Information Criterion, $BIC(model) = G^2(model) + \ln(N)\cdot p$, where *N* is the number of observations and *p* the number of model parameters. Setting $p = 0$, we have that $BIC(model) = G^2(model)$. To avoid indeterminate results in the Mathematica script implementing the G^2 function, probabilities were restricted to a range of [0.0001, 0.9999]. BIC allows us to compare non-nested models, as is the case with the quantum and Hogarth and Einhorn's (1992) models. We fitted and compared the models separately for the PN and NP conditions.

For the PN condition after randomly assigning data points for which FSDR=SSDR we had N=346. For the quantum model $BIC = 85$. For Hogarth and Einhorn's (1992) model $BIC = 66$. For the NP condition we had N=522. For the quantum model $BIC = 129$. For Hogarth and Einhorn's (1992) model $BIC = 135$. Readers might wonder whether the qualitative conclusions change if we exclude the data points for which FSDR=SSSR. This was mostly not the case. For the PN condition, N=130, for the quantum model $BIC = 64$, and for Hogarth and Einhorn's (1992) model $BIC = 52$. For the NP condition, N=307, for the quantum model $BIC = 104$, and for Hogarth and Einhorn's (1992) model $BIC = 122$. Scatterplots for observed and fitted probabilities are shown in Figures 6a, 6b and 7a, 7b; in these graphs, bubble size is determined by number of overlapping points. Indicatively, we note that in Figure 6a the number of overlapping points varied from 1 to 91.

We briefly consider parameter distributions for the quantum model in Figures 8a, 8b, for each of the two conditions. The *rating* angle would be expected to be in the $\left(0,\frac{\pi}{2}\right)$ range. In many cases, the rating value for the PN condition was in the $\left(0,\frac{\pi}{4}\right)$ range and for the NP ones in the $\left(\frac{\pi}{4},\frac{\pi}{2}\right)$ range. The distribution for the *n* values is less well behaved in that, especially in the NP condition, many values are larger than what we would expect. Recall that the quantum model is based on trigonometric functions. Therefore, there are multiple angles for which the resulting probabilities are equivalent. The distribution of fitted parameters for Hogarth and Einhorn's (1992) model is shown in Figures 9a, 9b. It is not immediately obvious whether this distributional information can inform the workings of the model, but at the very least we have confirmation that optimization respected parameter bounds.

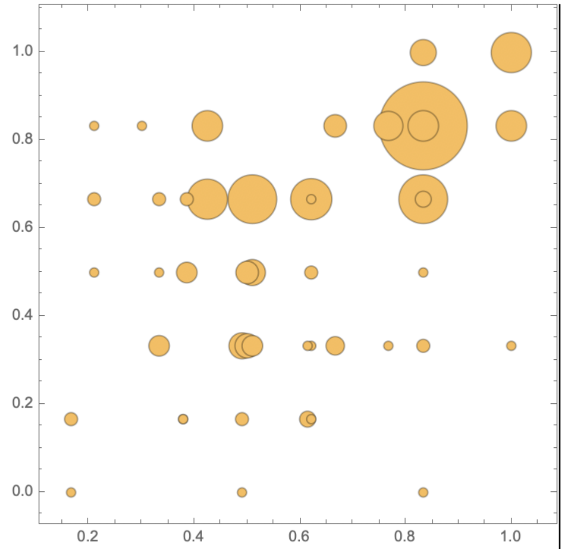

Figure 6a. Scatterplot for observed (vertical) vs. fitted probabilities for the quantum model, for the PN condition.

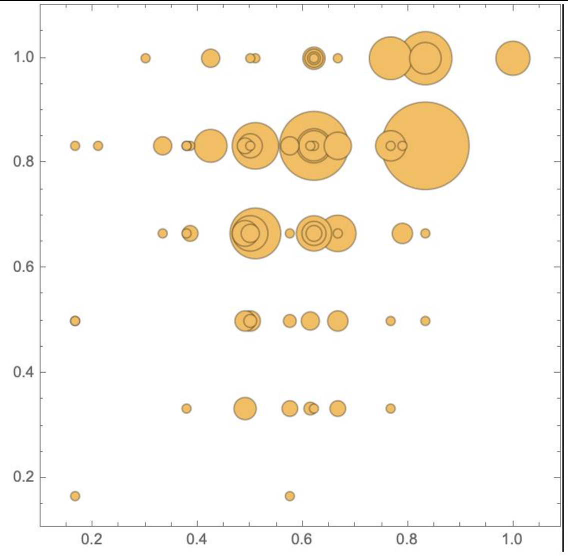

Figure 6b. Scatterplot for observed (vertical) vs. fitted probabilities for the quantum model, for the NP condition.

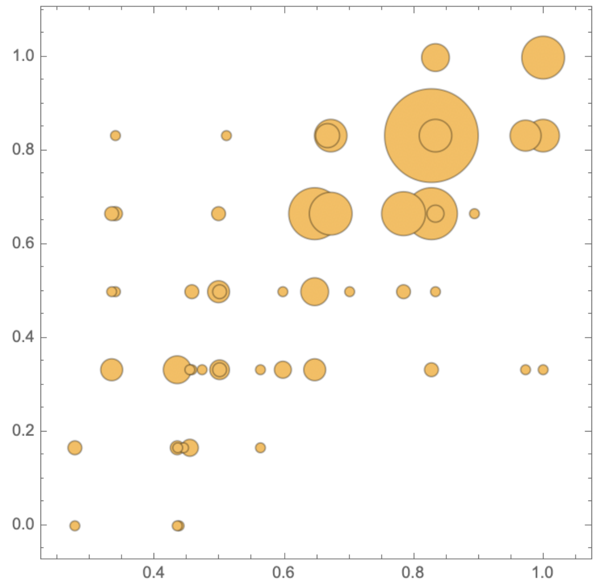

Figure 7a. Scatterplot for observed (vertical) vs. fitted probabilities for Hogarth and Einhorn's (1992) model, for the PN condition.

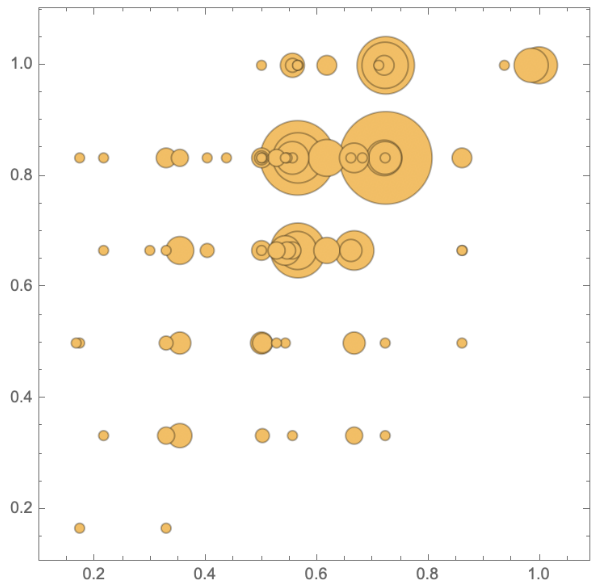

Figure 7b. Scatterplot for observed (vertical) vs. fitted probabilities for Hogarth and Einhorn's (1992) model, for the NP condition.

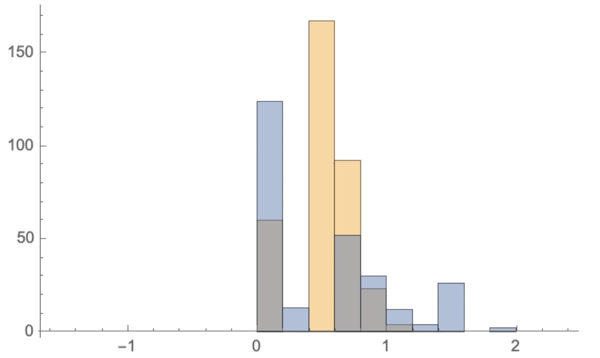

Figure 8a. The distribution of *n* (blue bars) and *rating* (yellow bars) values for the PN condition.

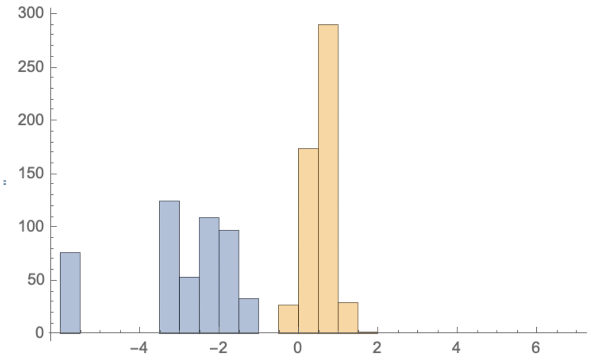

Figure 8b. The distribution of *n* (blue bars) and *rating* (yellow bars) values for the NP condition.

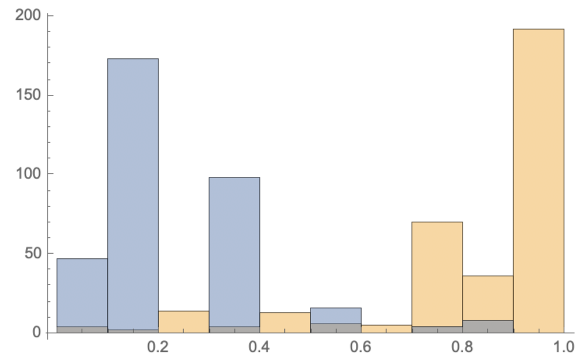

Figure 9a. The distribution of $s(x_P)$ (blue bars) and *S0* (yellow bars) values for the PN condition.

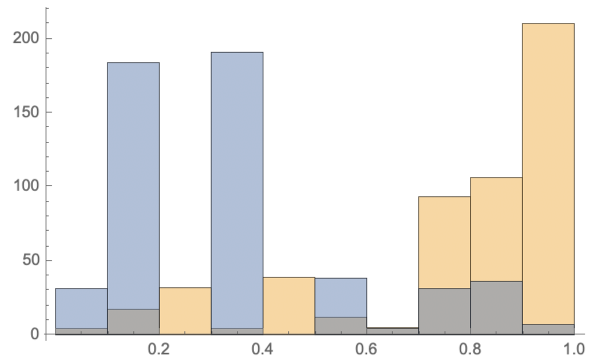

Figure 9b. The distribution of $s(x_P)$ (blue bars) and *S0* (yellow bars) values for the NP condition.

The central tenet of both models is that there should be an interrelatedness between the FSDR, SSSR, and SSDR values. This can be tested directly by randomizing one set of values and re-fitting the models. Is there still an association between empirically observed SSDR values and the ones predicted by the models? We randomized the SSSR ratings (uniform distribution in the [0,1] range), keeping the FSDR ratings intact. For the quantum model, this means that the *rating* parameters would be as before, but the *n* parameters would be extracted using randomized data. For Hogarth and Einhorn's (1992) model, both model parameters were fit to the pairs of FSDR and SSSR from scratch. For both models, randomizing SSSR ratings produced worse fit. For the PN condition, for the quantum model $BIC = 287$ and for Hogarth and Einhorn's model (1992) $BIC = 291$. For the NP condition, the corresponding values were $BIC = 478$ and $BIC = 234$. Clearly, the exact values from these control simulations will vary from simulation to simulation – but in all cases, one observes a healthy increase in fit values, as a result of randomizing SSSR ratings.

4. Discussion

It is well-established, and indeed intuitive, that previous judgments can change the mindset or perspective for subsequent ones (e.g., Bless & Schwarz, 2010; Schwarz, 2007). Some of the relevant effects have substantial practical importance, as in the case of question order effects (Bergus et al., 1998; Moore et al., 2002). The more subtle issue is whether judgments can alter the mental state in a way that impacts on subsequent information. It might seem that any judgment can alter the mental state, simply because of the information-gathering potential of the judgment. However, what is at stake is whether the behavioral impact of just processing a piece of information vs. of making a corresponding judgment are different. This is hardly a straightforward question. The baseline perspective from the predominant framework for probabilistic inference, CT, is that judgments reveal preexisting (albeit unknown) information. In this sense, there is limited room for putative constructive influences from judgments.

We know from previous research (e.g. Damasio, 1994; Zajonc, 1980) that people can accurately and rapidly form an affective evaluation. However, most studies conclude this because after showing the stimulus, the participant is asked to explicitly evaluate it. Perhaps it is the case that without asking someone for an explicit rating of the first stimulus, their feelings about it remain more ambiguous and they are therefore in a different state when they see a subsequent stimulus. There have been some proposals postulating belief changes commensurate with judgments (e.g., Festinger, 1957; Gloeckner et al., 2009) and some supporting empirical results (e.g., Sharot et al., 2010). Additionally, some researchers have argued that, for example, stating preferences does not reflect the ‘reading out’ of pre-formed beliefs, attitudes or values but rather is a constructive exercise whereby we improvise and create our evaluation in the moment (e.g. Chater, 2018; Dennett, 1993; Slovic, 1995).

In this work we explored two decision models which incorporate constructive influences, QT and Hogarth and Einhorn's (1992) belief-adjustment model. We used data from an extension to an empirical paradigm originally proposed to study constructive influences: In White et al.'s (2014, 2015) paradigm the main component was pairs of sequentially presented stimuli. With controlled stimuli, so that in each pair one stimulus would be positively valenced and the other negatively valenced, or vice versa, they showed that a judgment vs. simple observation of the first stimulus would entail a more intense evaluation for the second stimulus. The empirical objective of the present work was to consider whether this EB that White et al. (2014) reported could be observed with realistic stimuli in an applied context fostering more thoughtful judgments.

We conducted a test of the EB with respondents being managers and professionals answering questions about their own organization. Exactly the same two questions in exactly the same order were responded to differently by the same participant, depending on whether there was an intermediate judgment (a judgment to the first question) or not. The intermediate judgment led to a more contrasting judgment for the second question, with this contrast ranging from 1.29 units to 1.86 units, which represents, respectively 0.86 to 1.66 standard deviations. We draw attention to the fact that this is a substantial amount; the size of the effect further undermines a perception of questionnaires as revealing pre-existing attitudes or knowledge in a bias-free manner. Our results are in in line with other demonstrations of questionnaire biases, involving participants responding to questions relevant to their expertise (e.g., Bergus et al., 1998) or when there is strong motivation to respond in an objective way (e.g., McKenzie, Lee, & Chen, 2002; Pennington & Hastie, 1986).

The investigation of QT was motivated by the fact that a fundamental aspect of QT is the way the (mental) state has to change as a result of judgments or measurements. As a result, QT cognitive models provide a very constrained notion of how intermediate judgments should impact on subsequent ones (Trueblood & Busemeyer, 2011; Wang & Busemeyer, 2013). Such models have been applied previously to situations indicating constructive influences (Kvam et al., 2015; Yearsley & Pothos, 2016). We think that the EB paradigm enables a more direct test of the psychological plausibility of the constructive influences embodied in QT models, because of the simplicity of the paradigm. The QT model of the EB is that there is a qualitative difference between explicitly rating and not explicitly rating the first stimulus. The former changes the cognitive state in a way that is different from the latter and can be considered a constructive influence. The QT model we developed is based on simple assumptions regarding the representations of the two pieces of information and how the mental state can change.

One advantage of the QT model is that a qualitative prediction for the constructive influence corresponding to the EB can be made a priori. White et al. (2014) developed the original EB paradigm based on a simplified version of Figure 2. Because of the geometric nature of QT representations, it is often the case that complex mathematical intuitions can be expressed relatively simply. Another advantage of the QT model is that the prediction for the constructive influence is specific. The constructive influence has to be of a certain kind and in a certain direction. Notably, as a result of a judgment, QT requires the state to identify itself with the outcome of the judgment.

There are some disadvantages of the QT model for the EB paradigm as formulated at present. First, the assumed representation assumes a distinction between just positive vs. negative evaluations for the organizational statements. By contrast, participants have to rate these statements on a multi-point Likert scale. This means that when the first evaluation for the first stimulus is too ambiguous, the QT model is forced to predict a constructive influence which is too large. A future extension of the model should incorporate representations more closely matched to the assumed internal scale of evaluation. We note that this is not a straightforward issue. For example, is there a 'native' internal evaluation scale? Is it the case that an externally imposed evaluation scale is just replicated internally etc.? Second, the projection mechanism in QT can allow for 'error'. In this work we employed the more standard approach for projection, involving projection operators. Instead, one could employ Positive Operator Valued Measures (POVMs; Yearsley & Busemeyer, 2016). POVMs are just like projection operators, but for which projection is not errorless. That is, the answer to a question might be 'yes', but there is a probability that the state will be projected to the subspace for 'no'. The use of POVMs in the present case is potentially important, because it informs the way a state changes as a result of a judgment. Finally, we modeled the influence of introducing the second stimulus on the mental state with time independent unitary dynamics. In the two-dimensional, real vector space we employed for representations, unitary dynamics can be thought of as simple rotations. For example, in introducing a N stimulus, the mental state rotates towards the N ray. The problem is that when the rotation is too large, it can overshoot the N ray. The correct dynamics to employ would be so-called open system dynamics, which have the advantage that rotation can be set to asymptotically converge to a particular state (e.g., Nielsen & Chuang, 2000). However, this is a more technically complex approach and corresponding applications are far less common in psychology (Asano et al., 2011a, 2011b). Despite all these possible avenues for extending the QT model, we think that it is important that the model we evaluated in the present work was close to the original ideas of White et al. (2014).

The detailed consideration of Hogarth and Einhorn's (1992) model for the EB paradigm was partly motivated by the fact that this is a very influential model for the study of questionnaire biases. Even though most applications of the model concern questionnaire order effects, the model's distinction of evaluation processes into SbS and EoS provides a natural framework for examining the EB paradigm as well. Moreover, a key assumption in Hogarth and Einhorn's (1992) model is that the mental state changes at each step of an evaluation process, as a result of intermediate judgments; that is, the model incorporates constructive influences. The difference between Hogarth and Einhorn's (1992) model and the QT model is that in the former the constructive influence is not set, but rather is allowed to vary as a function of the model's parameters. A disadvantage of Hogarth and Einhorn's (1992) model is that it has many parameters. In order to apply the model to an empirical situation as simple as the EB paradigm, several assumptions had to be introduced regarding which parameters can be eliminated and how the remaining parameters should be constrained. It would be desirable for Hogarth and Einhorn's (1992) theory to be developed more, so that some of these parameter decisions can be less post hoc. It is worth noting that Hogarth and Einhorn's (1992) model can be formulated for the EB paradigm in a way that no constructive influences are predicted at all (White et al., 2014). Relatedly, it is hard to anticipate the EB bias from Hogarth and Einhorn's (1992) model.

Looking at the QT and Hogarth and Einhorn's (1992) models together, it is worth stressing that the tests for the two models were set up in a strict way. Per triplet of judgments {FSDR, SSSR, SSDR}, we employed two of the judgments to fix the model parameters, and then let the models make a parameter-free prediction for the third judgment. Both models performed reasonably well, considering the predictive challenge. For the PN subset of the data, Hogarth and Einhorn's (1992) model was better than the quantum model and for the NP subset of the data the two models were more equivalent. At face value, this conclusion means that constructive influences are more flexible than what is assumed by QT. However, there are two qualifying considerations which advise caution regarding the strength of this conclusion. First, Hogarth and Einhorn's (1992) model is a more flexible model. Even though we employed BICs for model evaluation, which penalize for model complexity, there are more sophisticated techniques for doing so (Lee & Wagenmakers, 2013). Second, there are some fairly clear, albeit technically complex, directions for improving the QT model, as discussed above. The present analyses are important in helping appreciate the conditions of applicability of some common technical assumptions in QT cognitive models.

Relatively speaking, the reasonable performance of the models indicates that the EB paradigm does reveal constructive influences from earlier judgments onto later ones. It might seem desirable to augment the model-based explanations from the QT and Hogarth and Einhorn's (1992) models with claims concerning the way constructive influences might go hand in hand with changes in memory or attention (cf. Sharot et al., 2010). However, neither model currently incorporates any corresponding assumptions. Moreover, the empirical results themselves offer limited opportunity for interpretations relating to memory or attention processes. A possible exception concerns research suggesting that a prior stimulus can lead to differential effects on subsequent judgements, depending on the degree of awareness that the participant has regarding the stimulus (e.g. Bless & Schwarz, 2010). Whether a simple judgment vs. observation for a stimulus leads to increased awareness for the stimulus in subsequent judgements remains an open question. However, note that there is ample evidence that simple observation of a stimulus already produces plenty of information (e.g., Damasio, 1994; Zajonc, 1980).

In conclusion, the present results offer reassurance regarding the reality of constructive influences in judgment, even under realistic and (relatively speaking) high-stakes circumstances. The application of two formal decision models consistent with constructive influences provides some insight into the underlying psychological mechanisms. We highlight the QT model, as the more novel approach: QT embodies an inherent constructive influence from measurements, judgments, or evaluations and has enabled a priori predictions regarding the direction of the EB. However, the study of constructive influences with formal methods is fairly novel, and we identified several directions along which the two models can be improved. Regarding the QT model, there are more sophisticated tools for the various mechanisms which are required for the modelling of the EB; and regarding Hogarth and Einhorn's (1992) model more work is needed to identify ways to restrict its parametric flexibility.

Acknowledgements

EMP was supported by Leverhulme Trust grant RPG-2015-311 and ONRG grant N62909-19-1-2000. We would like to thank Michael Lee and an anonymous reviewer for their many helpful comments.

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology*, 53(5), 314-348. doi:10.1016/j.jmp.2009.04.005

Ariely, D., & Norton, M.I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences*, 12, 13-16. doi: 10.1016/j.tics.2007.10.008

Asano, M., Ohya, M., Tanaka, Y., Basieva, I., & Khrennikov, A. (2011a). Quantum-like model of brain's functioning: decision making from decoherence. *Journal of Theoretical Biology*, 281, 56-64. doi: 10.1016/j.jtbi.2011.04.022

Asano, M., Ohya, M., Tanaka, Y., Khrennikov, A., & Basieva, I. (2011b). On application of Gorini-Kossakowski-Sudarshan-Lindblad equation in cognitive psychology. *Open Systems & Information Dynamics*, 18, 55-69. doi: 10.1142/S1230161211000042

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology*, 41, 258-290.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic evaluation activation effect. *Journal of Personality and Social Psychology*, 62, 893–912. doi: 10.1037/0022-3514.62.6.893

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and order information. *Medical Decision Making*, 18, 412-417. doi: 10.1177/0272989X9801800409

Brehm, J.W. (1956). Post-decision changes in the desirability of choice alternatives. *Journal of Abnormal and Social Psychology*, 52, 384–389.

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. Vol. 42, pp. 319-373). San Diego, CA, US: Academic Press. doi: 10.1016/S0065-2601(10)42006-7

Broekaert, J. B., Busemeyer, J. R., & Pothos, E. M. (in press). The disjunction effect in two-stage simulated gambles. An experimental study and comparison of a heuristic logistic, Markov and quantum-like model. *Cognitive Psychology*.

Busemeyer, J. R., & Bruza, P. (2011). *Quantum Models of Cognition and Decision Making*. Cambridge, UK: Cambridge University Press.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–42.

Chater, N. (2018). *The Mind is Flat: The Illusion of Mental Depth and the Improvised Mind*. London: Penguin.

Clore, G. L., & Colcombe, S. (2003). Affective priming: Findings and theories. In J. Musch, & K.C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 345–380). Mahwah, NJ: Erlbaum.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.

Dennett, D. (1993). *Consciousness Explained*. London: Penguin.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, 13(6), 513-519. doi: 10.1111/1467-9280.00490

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of evaluations. *Journal of Personality and Social Psychology*, 50, 229–238. doi: 10.1037/0022-3514.50.2.229

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford Univ. Press, Stanford.

Gloeckner, A., Betsch, T., & Schindler, N. (2009). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, 23, 439 – 462.

Goldstein, D.G. & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review*, 109, 75–90.

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in probabilistic language of thought. In Eds. E. Margolis & S. Laurence, *New Directions in the Study of Concepts*, pp.623-653. MIT Press: Cambridge, MA.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition*, 17, 35–47. doi: 10.3758/BF03199555

Haven, E. and Khrennikov, A. (2013). *Quantum Social Science*. Cambridge University Press: Cambridge, UK.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology*, 24, 1–55. doi: 10.1016/0010-0285(92)90002-J

Hughes, R.I.G. (1989). *The Structure and Interpretation of Quantum Mechanics*. Cambridge, MA: Harvard University Press.

JASP Team (2016). JASP (Version 0.8.0.0)[Computer software]

Kahneman, D. (2001). *Thinking fast and slow*. Penguin: London, UK.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making*, 5(3), 187-200. doi: 10.1002/bdm.3960050304

Khrennikov, A., Basieva, I., Pothos, E. M., & Yamato, I. (2018). Quantum probability in decision making from quantum information representation of neuronal states. *Scientific Reports*, 8, 16225.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: quantum characteristics of evidence accumulation. PNAS, 112, 10645-10650.

LeDoux, J. E. (1996). *The Emotional Brain : The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

Lee, M. D. & Wagenmakers, E.J. (2013). Bayesian Cognitive Modeling: A Practical Course. Cambridge University Press.

Lewandowsky, S. & Smith, P. (1983). The effect of increasing the memorability of category instances on estimates of category size. *Memory & Cognition*, 11, 347-350

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science*, 16, 190–195.

Lewandowsky, S., & Kirsner, K. (2000). Knowledge partitioning: Context dependent use of expertise. *Memory & Cognition*, 28, 295–305.

Lewandowsky, S., Kalish, M., & Ngang, S. K. (2002). Simplified learning in complex situations: knowledge partitioning in function learning. *Journal of Experimental Psychology: General*, 131, 163-193.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13, 106-131.

Lieder, F., & Griffiths, T. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 1-85.

Litt, A., Eliasmith, C., Kroon, F. W., Weinstein, S., & Thagard, P. (2006). Is the brain a Quantum computer? *Cognitive Science*, 30, 593-603.

Martin, L.L. (1986). Set/Reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, 51(3), 493-504. doi: 10.1037/0022-3514.51.3.493

Martin, L. L., & Shirk, S. (2007). Set/reset and self-regulation: Do contrast processes play a role in the breakdown of self-control. In D. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (pp. 205–225). New York, NY: Psychology Press.

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, 15, 1–18.

Moore, D. W. (2002). Measuring new types of question-order effects. *Public Opinion Quarterly*, 66(1), 80-91. doi: 10.1086/338631

Mussweiler, T., & Neumann, R. (2000). Sources ofmental contamination: Comparing the effects of self-generated versus externally provided primes. *Journal of Experimental Social Psychology*, 36, 194–206.

Nielsen, M. A., & Chuang, I. L. (2000). *Quantum computation and quantum information*. Cambridge University Press.

Oaksford, M. & Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101, 608-631.

Pennington, N. & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242-258.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral & Brain Sciences*, 36, 255–327. doi: 10.1017/S0140525X12001525

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8, 364–382.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, 72, 54–107.

Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 497–502.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L.L. Martin & A. Tesser (Eds.), *The Construction of Social Judgments* (pp. 227-245). Hillsdale, NJ: Erlbaum.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, 25, 638–656. doi: 10.1521/soco.2007.25.5.638

Schwarz, N., Sanna, L.J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, 39, 127–161.

Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference?: Evidence from blind choice. *Psychological Science*, 21(9) 1231–1235. doi: 10.1177/0956797610379235

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25, 231-280.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*. 119: 3–22.

Slovic, P. (1995). The construction of preference. *American Psychologist*, 50(5), 364-371.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, 112, 881-911.

Tenenbaum, J.B, Kemp, C., Griffiths, T.L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, 331, 1279-1285.

Torre, J. B. & Lieberman, M. D.(2018). Putting feelings into words: affect labeling as implicit emotion regulation (2018). *Emotion Review*, 10, 116-124.

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, 35(8), 1518-1552. doi: 10.1111/j.1551-6709.2011.01197.x

Trueblood, J. S., Yearsley, J. M., & Pothos, E. M. (2017). A quantum probability framework for human probabilistic inference. *Journal of Experimental Psychology: General*, 146, 1307-1341.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuctive fallacy in probability judgment. *Psychological Review*, 90, 293-315.

Yearsley, J. M. & Pothos, E. M. (2016). Zeno's paradox in decision making. *Proceedings of the Royal Society B*, 283, 20160291.

Yearsley, J. M. & Busemeyer, J. R (2016). Quantum cognition and decision theories: a tutorial. Journal of Mathematical Psychology, 74, 99-116.

Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4). doi: 10.1111/tops.12040

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1407756111

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: the constructive role of articulating impressions. *Cognition*, 133(1), 48-64. doi: 10.1016/j.cognition.2014.05.015

White, L. C., Barqué-Duran, A., & Pothos, E. M. (2015). An investigation of a quantum probability model for the constructive effect of affective evaluation. *Philosophical Transactions A*, 374: 2015014. doi: 10.1098/rsta.2015.0142

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151−175. doi: 10.1037//0003-066X.35.2.151

Appendix 1. Hogarth and Einhorn's (1992) belief-adjustment model

We provide a detailed discussion of how to apply Hogarth and Einhorn's model (1992) to the description of constructive influences and the present data. Small parts of this appendix are included in the main text. In Hogarth and Einhorn's (1992) model, there is a distinction between step by step (SbS) and end of sequence (EoS) evaluation processes. In principle, the distinction between SbS and EoS fits nicely with the EB paradigm, since SbS can be associated with FSDR and then SSDR and EoS with SSSR. However, there are several different approaches to formulate Hogarth and Einhorn's (1992) model for the present situation and some of these predict no difference depending on the presence of an intermediate rating or not (White et al., 2014). We outline a more elaborate formulation here which allows a prediction for an EB.

We define the following variables: S_k $(0 \le S_k \le 1)$ is the impression of participants after considering k statements. $s(x_k)$ is the subjective evaluation of the kth statement. $s(x_1, \ldots, x_k)$ is the combined impact of all the statements, statement 1 to k. Because in the present case we are employing a unipolar scale, we can assume $0 \le s(x_k) \le 1$ (Hogarth & Einhorn, 1992, p.11). R is the reference point against which the impact of the kth statement is assessed. w_k $(0 \le w_k \le 1)$ is an adjustment weight in relation to how the kth statement impacts on the belief state.

The main equation of the model (Equation 1 in Hogarth and Einhorn's, 1992, paper) is

$$S_k = S_{k-1} + w_k[s(x_k) - R]$$

Hogarth and Einhorn (1992) develop the specification of the SbS process more so than of the EoS one. We describe the SbS process first. For an SbS process, depending on what R is there are two different forms of the model. According to Hogarth and Einhorn, we have an adding version, when R=0 and $S_k = S_{k-1} + w_k s(x_k)$, or an averaging version, when $R = S_{k-1}$ and $S_k = (1 - w_k)S_{k-1} + w_k s(x_k)$. Hogarth and Einhorn (1992) suggest that the averaging version applies when we have an estimation task, as opposed to an evaluation one, and when unipolar scales are employed, as opposed to bipolar ones. Therefore, according to both criteria, the SbS process relevant to the present experiments is $S_k = (1 - w_k)S_{k-1} + w_k s(x_k)$. Some authors adopt the adding vs. the averaging version based on considerations of goodness of fit (e.g., Trueblood & Busemeyer, 2012), but we prefer to follow the approach from Hogarth and Einhorn's (1992) theoretical assumptions.

The final part of the SbS process concerns the way new information updates the existing state. Notably, when $s(x_k) \leq R$, we have that $w_k = aS_{k-1}$ and when $s(x_k) > R$, $w_k = \beta(1 - S_{k-1})$ (these are equations 6a, 6b in Hogarth and Einhorn's paper). Recall, since we have an estimation process, then $R = S_{k-1}$. The $a,\ \beta$ parameters concern sensitivity to negative vs. positive information and their bounds are $0 \leq \{a,\ \beta\} \leq 1$. Some authors set them to 1, e.g., Trueblood and Busemeyer (2012, Equation 36). Moreover, in the present case, allowing the sensitivity parameters to be included in the model as free parameters raises risks of overparameterization. Given previous research (Trueblood & Busemeyer, 2012) and such practical considerations, we believe it is more appropriate to set the sensitivity parameters to 1. Our presentation below retains the sensitivity parameters in some cases, just for completeness of exposition. Putting everything together, for SbS processes, belief adjustment is governed by the pair of equations:

$$S_k = \begin{cases} (1 - aS_{k-1})S_{k-1} + aS_{k-1}s(x_k), & s(x_k) \leq S_{k-1} \\ (1 - \beta(1 - S_{k-1}))S_{k-1} + \beta(1 - S_{k-1})s(x_k), & s(x_k) > S_{k-1} \end{cases}$$

In order to apply the model to the EB paradigm, we need consider that there are two stimuli with either positive or negative valence (or approximately so). We assume that participants start from a neutral valence state S_0. The SbS process applies to the double rating condition. In the PN condition, we have $S_P = (1 - \beta(1 - S_0))S_0 + \beta(1 - S_0)s(x_P)$. Note, if

$\beta = 1$, this would be simplified to $S_P = S_0^2 + (1 - S_0)s(x_P)$. Here we have applied the part of the SbS process corresponding to $s(x_k) > S_{k-1}$, since we can trivially assume that $s(x_P) > S_0$, that is, in the PN condition, the first judgment will be more positive than the neutral baseline. Given that after the first judgment the current state is S_P, the state following the second stimulus and judgment should be:

$$S_{PN} = (1 - aS_P)S_P + aS_P s(x_N)$$

In this case we have applied the part of the SbS process corresponding to $s(x_k) \leq S_{k-1}$, since trivially $s(x_N) \leq S_P$. To see the logic of this final result, simplify for $a = 1$ so that $S_{PN} = (1 - S_P)S_P + S_P s(x_N)$. Also note that S_P will be close to 1, therefore the first term for S_{PN}, that is $(1 - S_P)S_P$, will be close to 0. So, $S_{PN} \sim S_P s(x_N)$, that is, it will have a low value, since $s(x_N)$ itself is low. Notice that by having S_P in the second equation (for S_{PN}), we explicitly assume linkage between the first and the second judgment. We could instead have S_0, since for Hogarth and Einhorn's (1992) model linkage between pieces of evidence is assumed because all information is combined towards an eventual evaluation. However, we think that having S_0 instead of S_P in the equation for S_{PN} would prejudice against Hogarth and Einhorn's model too much.

Applying a similar logic, for the NP condition we have:

$$S_N = (1 - aS_0)S_0 + aS_0 s(x_N)$$

In this case we have applied the part of the process corresponding to $s(x_k) \leq S_{k-1}$. Given that our current state is S_N, the state following the second stimulus and judgment should be

$$S_{NP} = (1 - \beta(1 - S_N))S_N + \beta(1 - S_N)s(x_P)$$

Regarding the EoS process, Hogarth and Einhorn offer the nearly identical Equations (5) and (6) in their paper, which are, respectively:

$$S_k = S_0 + w_k[s(x_1, \dots, x_k) - R]$$

$$S_k = s(x_1) + w_k[s(x_2, \dots, x_k) - R]$$

They note (p.12) that “$s(x_1, \dots, x_k)$ is some function, possibly weighted average, of the individual subjective evaluations (or scale values) of the items of evidence that follow the anchor”. We can adopt the same assumption as for SbS processes, that if the evaluation process is an estimation one then $R = S_{k-1}$. For an EoS process, we would have $R = S_0$. Regarding the adjustment weight w_k, note that the task is not one of integrating information towards a single evaluation, rather it is a sequential evaluation task. Therefore, we can safely assume that the $s(x_1, \dots, x_k)$ function is largely biased towards the last piece of evidence, so that $s(PN) < S_0$ and $s(NP) > S_0$. Based on Hogarth and Einhorn’s (1992) discussion for the SbS process, we can suggest that when $s(x_1, \dots, x_k) \le S_0$, we have $w_k = aS_0$ and when $s(x_1, \dots, x_k) > S_0$, $w_k = \beta(1 - S_0)$, remembering that in an EoS process we have a single updating step. Overall, for an EoS process we have:

$$S_k = \begin{cases} S_k = S_0 + aS_0[s(x_1, \dots, x_k) - S_0], & s(x_1, \dots, x_k) \le S_0 \\ S_k = S_0 + \beta(1 - S_0)[s(x_1, \dots, x_k) - S_0], & s(x_1, \dots, x_k) > S_0 \end{cases}$$

With the above tools in hand, for the PN condition $s(x_P, x_N) \le S_0$ and so we have

$$S_{PN} = S_0 + aS_0[s(x_P, x_N) - S_0]$$

For the NP condition, it is also trivial to assume that $s(x_N, x_P) > S_0$ and so we obtain

$$S_{NP} = S_0 + \beta(1 - S_0)[s(x_N, x_P) - S_0]$$

Tables 1A and 1B summarize the equations which are relevant for modeling result from the EB paradigm.

Table 1A. Hogarth and Einhorn’s (1992) model for describing results from the EB paradigm.

SbS	EoS
$S_{PN} = (1 - aS_P)S_P + aS_P s(x_N)$, $S_P = (1 - \beta(1 - S_0))S_0 + \beta(1 - S_0)s(x_P)$	$S_{PN} = S_0 + aS_0[s(x_P, x_N) - S_0]$
$S_{NP} = (1 - \beta(1 - S_N))S_N + \beta(1 - S_N)s(x_P)$, $S_N = (1 - aS_0)S_0 + aS_0 s(x_N)$	$S_{NP} = S_0 + \beta(1 - S_0)[s(x_N, x_P) - S_0]$

Table 1B. Hogarth and Einhorn's (1992) model for describing results from the EB paradigm, with the simplification of setting the sensitivity parameters to 1.

SbS	EoS
$S_{PN} = (1 - S_P)S_P + S_P s(x_N)$, $S_P = S_0^2 + (1 - S_0)s(x_P)$	$S_{PN} = S_0 + S_0[s(x_P, x_N) - S_0]$
$S_{NP} = S_N^2 + (1 - S_N)s(x_P)$, $S_N = (1 - S_0)S_0 + S_0 s(x_N)$	$S_{NP} = S_0 + (1 - S_0)[s(x_N, x_P) - S_0]$

We provide an examination of Hogarth and Einhorn's (1992) model for when the sensitivity parameters are set to 1, as an illustration of the capacity of the model to accommodate the EB. The EB is the finding that the rating for the second stimulus is more intense (more positive or more negative in the double rating condition, compared to the single rating one). So, in the PN case, the Evaluation bias is the finding that:

$$(1 - S_P)S_P + S_P s(x_N) < S_0 + S_0[s(x_P, x_N) - S_0] \Leftrightarrow$$

$$\big(1 - S_P + s(x_N)\big)S_P < S_0 + S_0[s(x_P, x_N) - S_0]$$

The above can be rewritten as:

$$\big(1 - S_0^2 - s(x_P) + S_0 s(x_P) + s(x_N)\big)\big(S_0^2 + (1 - S_0)s(x_P)\big) < S_0 + S_0[s(x_P, x_N) - S_0]$$

This function has many parameters, S_0, $s(x_P)$, $s(x_N)$, $s(x_P, x_N)$ and it cannot be immediately simplified. As for the fits reported in the main text of this paper, we set $s(x_P) = 1 - s(x_N)$ and $s(x_P, x_N) = s(x_N)$, essentially reducing the free parameters to $s(x_P)$ and S_0. This gives a condition for when the EB occurs as:

$$\left(1 - S_0^2 - s(x_P) + S_0 s(x_P) + s(x_N)\right)\left(S_0^2 + (1 - S_0)s(x_P)\right) < S_0 + S_0[s(x_N) - S_0]$$

We can now explore the above function regarding its capacity to produce the EB. We set ranges for the remaining free parameters as follows: $s(x_P) \in [0.5, 1]$, since $s(x_P)$ is assumed to be positive; $S_0 \in [0.3, 0.7]$, since we assume that the initial impression is mostly neutral. Then, it is immediately clear that the anchoring and adjustment model can produce the EB, for various configurations of its parameters (Figure 1S).

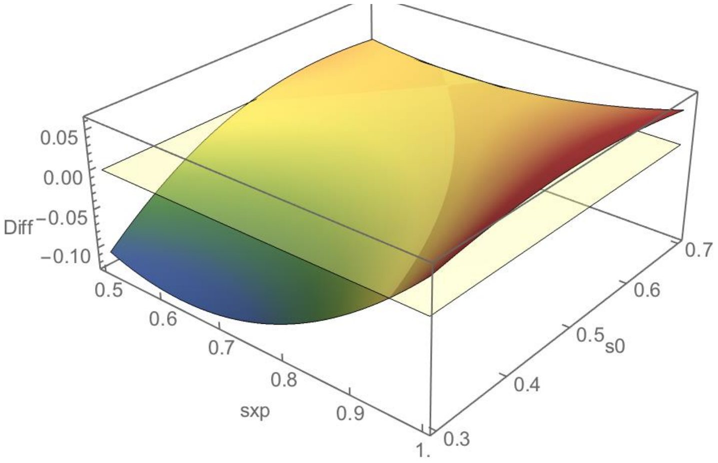

Figure 1S. Illustrating Hogarth and Einhorn’s (1992) model: We plot the EB (vertical axis), defined as SSSR-SSDR, where SSDR is the second judgment in an SbS process and SSSR is the only judgment in an EoS process, in a PN condition. Positive values correspond to an EB in the expected direction. The plane in semi-transparent yellow shows the 0 point.

If we ignore the sensitivity parameters, the version of Hogarth and Einhorn’s (1992) model we outlined above can be directly fitted to empirical data, in exactly the same way as the quantum model. That is, per triplet of judgments (SSSR, FSDR, SSDR), we can use two judgments to determine the two parameters of the model and then examine empirical predictions against the output of the third equation. Specifically, for the PN condition we have:

$SSSR = S_0 + S_0[s(x_N) - S_0]$, from the model equation regarding the EoS process.

$FSDR = S_0^2 + (1 - S_0)s(x_P)$, which is the first judgment in the SbS process.

$SSDR = (1 - S_P)S_P + S_P s(x_N)$, which is the second judgment in the SbS process.

Recall that $s(x_P, x_N) = s(x_N)$; $S_P = FSDR$; $s(x_P) = 1 - s(x_N)$.

For the NP condition we have:

$SSSR = S_0 + (1 - S_0)[s(x_P) - S_0]$, from the model equation regarding the EoS process.

$FSDR = (1 - S_0)S_0 + S_0 s(x_N)$, which is the first judgment in the SbS process.

$SSDR = S_N^2 + (1 - S_N)s(x_P)$, which is the second judgment in the SbS process.

In this case, recall that $s(x_P, x_N) = s(x_P)$.

If we include the sensitivity parameters, the equations become, for the PN condition:

$SSSR = S_0 + aS_0[s(x_N) - S_0]$, from the model equation regarding the EoS process.

$FSDR = (1 - \beta(1 - S_0))S_0 + \beta(1 - S_0)s(x_P)$, which is the first judgment in the SbS process.

$SSDR = (1 - aS_P)S_P + aS_P s(x_N)$, which is the second judgment in the SbS process.

For the NP condition we have:

$SSSR = S_0 + \beta(1 - S_0)[s(x_N, x_P) - S_0]$, from the model equation regarding the EoS process.

$FSDR = (1 - aS_0)S_0 + aS_0 s(x_N)$, which is the first judgment in the SbS process.

$SSDR = (1 - \beta(1 - S_N))S_N + \beta(1 - S_N)s(x_P)$, which is the second judgment in the SbS process.

References

Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating: the belief-adjustment model. Cognitive Psychology, 24, 1-55.

Trueblood & Busemeyer, 2012, A quantum probability model of causal reasoning, Frontiers in Psychology, 10.3389/fpsyg.2012.00138