PHILOSOPHICAL TRANSACTIONS A

Phil. Trans. R. Soc. A.
doi:10.1098/not yet assigned

An investigation of a quantum probability model for the constructive effect of affective evaluation

Lee C. White[1*†], Albert Barqué-Duran[2] & Emmanuel M. Pothos[2]

1 Department of Psychology, Swansea University, Singleton Park, Swansea, SA2 8PP, UK; 2 Department of Psychology, City University London, Northampton Square, London, EC1V 0HB, UK

Keywords: quantum probability; affective evaluation; judgement and decision-making; trustworthiness

Summary

The idea that choices can have a constructive effect has received a great deal of empirical support. The act of choosing appears to influence subsequent preferences for the options available. Recent research has proposed a cognitive model based on quantum probability, which suggests that whether or not a participant provides an affective evaluation for a positively or negatively valenced stimulus can also be constructive and so e.g. influence the affective evaluation of a second oppositely valenced stimulus. However, there are some outstanding methodological questions in relation to this previous research. This paper reports the results of three experiments designed to resolve these questions. Experiment 1, using a binary response format, provides partial support for the interaction predicted by the quantum probability model and Experiment 2, which controls for the length of time participants have to respond, fully supports the quantum probability model. Finally, Experiment 3 sought to determine whether the key effect can generalize beyond affective judgements about visual stimuli. Using judgements about the trustworthiness of well-known people, the predictions of the quantum probability model were confirmed. Together these three experiments provide further support for the quantum probability model of the constructive effect of simple evaluations.

Main Text

1. Introduction

There is a great deal of support for the constructive effects of choice, the phenomenon whereby the process of choosing actually influences the subsequent decision [e.g., 1–6]. For example, in Brehm's [7] original experiment, female shoppers were asked to evaluate the desirability of eight appliances. Then, as a reward for taking part in the study, they were offered two appliances, which they had rated equally desirable. After a short interval they were asked to re-evaluate the two appliances and it was observed that the rating of the chosen item was higher than the rejected item. In this and other demonstrations of the constructive effect of choice, the act of choosing appears to influence subsequent preferences for the original alternatives.

Recently, the circumstances in which a judgement may be seen as constructive have been extended by White, Pothos and Busemeyer [8,9]. They describe a cognitive model based on Quantum Probability (QP) and empirical evidence, which suggests constructive effects for simple affective evaluations, so that simply articulating how one feels about a positively or negatively valenced stimulus also leads to constructive effects. Note, we call QP the rules for how to assign probabilities to events from quantum mechanics, without any of the physics. This work, while promising, raises some key questions, notably regarding the robustness of the finding and the extent to which it generalized to other kinds of stimuli. In this paper, we briefly review the White et al. [8,9] research, including the QP principles that underlie the model, before describing the novel empirical directions that are the present focus.

1.1. The constructive effects of affective evaluation

In White et al.'s [8,9] experiments, fictitious adverts for insurance and mobile phones were created which had positive or

*Author for correspondence (l.c.white.517813@swansea.ac.uk).
†Present address: Department of Psychology, Swansea University, Singleton Park, Swansea, SA2 8PP, UK

negative content. The valence of images was either confirmed in a pilot study or images were selected from the Geneva Affective Picture Database [GAPED; 10], a database which contains images whose valence has been externally validated. In a within subjects design, participants were asked to view two images which were displayed sequentially in either a positive and then negative order (PN condition) or vice versa (NP condition; see Figure 1). In the double rating condition participants were asked to give a simple affective rating for the first image in the sequence and were then again asked for a rating for the second image. In the single rating condition, they saw the first image but provided no rating, instead moving on to view and rate the second image. The experiment was designed to address the following question: would there be a difference in ratings for the second image, between the single and double rating conditions?

<insert figure 1>

The results of three experiments, reported by White et al. [9], showed that there was a difference in ratings. Whether or not someone articulated an affective evaluation for the first image influenced how participants rated the second image. Specifically, when participants saw images in the PN condition, the ratings for the second negative image in the single rating condition were significantly more positive than the ratings of the same image in the double rating condition. In the NP condition, the ratings of the second positive image in the single rating condition were significantly less positive than ratings of the same positive image in the double rating condition. Thus, in both conditions, it appears that the intermediate rating increases the affective contrast between the two images.

White et al. [9] argued that this result could not be explained by other approaches, such as order effects or anchoring. The design of the experiments controlled for order effects, because the contrasting pairs of images involved the same pair of images presented in identical order (but, with and without the intermediate rating). The authors also reviewed Hogarth and Einhorn's [11] belief-adjustment model, which is in principle applicable in situations concerning the impact of intermediate judgments, and found that it predicted no difference between the single and double rating conditions. The possibility of anchoring effects was also examined by White et al. [9] who found no evidence that a prior rating was influencing the second rating, across all three experiments. Instead, the authors argued that a cognitive model, based on QP, could predict the empirical results they observed.

1.2. A QP model for the constructive effects of affective evaluation

Several researchers have recently been employing QP in cognitive modelling, in applications ranging across decision making [12,13,14], memory [15], concept combination [16], and other areas [17,18]; [for overviews see 19, 20, 21, 22]. QP is a formal framework for assigning probabilities to events, which originated from attempts to provide an explanation for paradoxical findings in physics, defying classical interpretations [22,23]. The well-known quantum mechanics theory is QP plus all the required assumptions to apply QP to physical systems. Classical probability (CP) and QP are based on different axioms. QP has some unique features, with no equivalents in CP, such as incompatibility, entanglement and superposition. It has been argued that QP may be useful in cognitive modelling because phenomena analogous to those observed in physics are also present in human decision making. Of relevance presently, CP models in decision theory naturally assume that, as the cognitive state changes from moment to moment, at any specific moment it is considered to be in a definite state (even if this state is unknown). Alternatively, QP models allow the cognitive system, at each moment, to be in a superposition state regarding a question (or a stimulus), which reflects ontic indefiniteness for the question outcomes (or feelings about the stimulus) – that is, the question outcomes do not exist, prior to a measurement. Superposition is a technical term in QP and indicates a special kind of uncertainty, such that the cognitive system has the potential for any of the possible decisions at each moment, but which one is selected cannot be determined until the system is measured (e.g. a judgment or affective evaluation is made). According to models based on standard CP theory, the measurement taken of a system reflects the state of the system immediately prior to the measurement. However, in QP theory, taking a measurement of a system can create, rather than record, a property of the system [25], which means that the subsequent state of the cognitive system is constructed from the interaction between the superposition state and the measurement taken [26]. These two fundamental principles of QP theory, superposition and the requirement that a measurement creates rather than records a property of the system, offer a natural and straightforward way to model constructive processes in judgement and decision making[1].

---

[1] It is possible to devise a CP based model to account for White et al.'s results by including in the model a mechanism for how decisions/ measurements could alter the relevant representations. But arguably, the QP approach is simpler because a

The model devised by White et al. [9] uses these ideas to predict a difference in the ratings of a positively or negatively valenced visual stimulus, depending on whether a previous, (broadly speaking) unrelated oppositely valenced stimulus was rated or not. The model leads us to the following insight into the psychological processes that underpin the observed effect. The participant's initial cognitive state is set by the first image in the sequence. Following the intermediate affective evaluation of the first image, the cognitive state is changed to being of either positive or negative affect. This change is represented in the model by a collapse of the state vector onto either a positive or negative affect ray (representing positive or negative affect; there are some simplifying assumptions here, but which are not relevant in this discussion). This is like an abstraction process, whereby some of the information about the first image is forgotten and attention is focused on information related to its affective properties. It is also the critical constructive step in the model: the intermediate rating changes the mental state in a certain way. This means that having made the intermediate rating, when the second oppositely valenced image is presented, it is evaluated from the perspective of a different cognitive state, than it would have been without an intermediate rating. As the second image is opposite in valence to the first, when the cognitive state is a pure affective one, there is a greater contrast in the impression made by the second image. Without the intermediate rating, the differences between the images concern aspects of their affective quality, but also differences between the images that are not related to affect, so the affective contrast between the first and second image is less pronounced. It is in this way that the QP model prediction arises, that the intermediate rating increases the affective contrast between the two stimuli. Further detail on the model can be found in Appendix A, where we provide a simple illustration of the model that shows how the key prediction of the QP model emerges.

Note, one can easily envisage cognitive models based on memory or attention which could postulate that the intermediate rating (somehow) increases attention (or memory) for the first image which, in turn, generates a feeling of greater affective contrast when the second, oppositely valenced, image is presented. Additionally, one can specify fairly easily a cognitive process on a vector space, based on rotation and an additional assumption that the mental state changes as a result of ratings, that can describe the empirical result. So, in what way is a QP approach to be preferred over such alternatives? The reason why we have favoured QP theory as an appropriate framework for understanding our key result is that, in QP theory, the way a rating (or judgment, decision etc.) affects the (mental) state is an integral, fundamental part of the theory. The collapse postulate is a key, structural aspect of QP theory and there is no room for altering it or applying QP theory without it. By contrast, approaches such as the ones outlined above, e.g., based on attention or memory, can include a constructive role for intermediate judgments, or not, with equal ease. We hasten to add that understanding the impact of intermediate judgments in terms of memory and attention is likely to be very valuable too. But, such approaches are primarily descriptive (because, as noted, constructive influences are not required), whereas, by contrast, the presence of constructive influences is a basic requirement (and so an informative test) of any cognitive model based on QP theory. It is also perhaps interesting to note that the original discovery of the effect [8,9] was based on an a priori application of QP model.

1.3. Methodological questions

We have identified two outstanding methodological questions with respect to White et al.'s [8,9] findings. The first concerns the specification of the QP model. Given that the experiments employed a nine point rating scale with anchors "1: very unhappy to 9: very happy", ideally judgements should be represented by a nine-dimensional vector space rather than the simplified two-dimensional vector space, used in White et al.'s QP model. A demonstration of the same result, when participants are required to make a simple binary choice between being either happy or unhappy in response to the images, would provide further support for White et al.'s QP model and this is the focus of Experiment 1 in this paper. That is, an implication of the White et al. [9] QP model is that the intermediate rating would increase affective contrast between the two images (in a measurable way), even when the images are rated in a binary (positive, negative) way.

Note, this is an ambitious prediction. We are predicting that, for example, in the NP condition, there will be participants who indicate that the second advert makes them feel happy in the double rating condition, and also indicate that the same advert makes them feel unhappy in the single rating condition (all key comparisons are within participants!). Overall, we predict that for the NP condition, the probability that second adverts in the double rating condition elicit a happy response will be greater than in the single rating condition. Analogously, for the PN condition, we predict that the probability that second adverts in the double rating condition elicit a happy response will be lower than in the single rating condition.

---

constructive effect of measurement is a fundamental part of the formalism, whenever the underlying state is a superposition one.

A second methodological question concerns the amount of time that participants had to process the images in the different conditions in the three experiments reported in White et al. In the double rating condition they saw the first image for five seconds and then had no limit on the amount of time they could take before providing their response. But in the single rating condition, they just saw the first image for five seconds, before being presented with the second image. So a difference in ratings might arise from the fact that people process the first image in the double rating condition for longer, perhaps increasing the likelihood of more deliberative or strategic processing. Such processing possibly implies that the image would leave a stronger impression and have greater saliency or become more accessible, as a point of reference, when considering the second image.

The rating of the second image in our experiments may be affected by a process of affective priming of the first image. Affective priming is the finding that processing of an affective stimulus can be faster and more accurate when preceded by stimulus of the same valence as opposed to an oppositely valenced prior stimulus. Does affective priming depend on processing time? Actually, there is a large body of research on affective priming, which suggests that the evaluation of a stimulus' valence does not depend on the amount of processing time. Affective priming research indicates that judgements are formed automatically, independently of other cognitive processes (even without fully processing the features of the stimulus) and can be generated in response to novel stimuli [27,28,29,30,31,32,33]. The evidence indicates that affective priming occurs most reliably when the stimulus onset asynchrony (SOA) is below 300 milliseconds. SOAs of above 500 milliseconds can result in strategic responses to stimuli from participants [30,34,35].

A related point is this: research also shows that an affectively charged prime (such as the first image in our experiments) facilitates the processing of a similarly valenced target, when compared with oppositely valenced targets [27,30]. These effects have been demonstrated with a variety of affectively charged primes, including words [27,30] and images of animals, objects and people [34; for a review see 35].

Such explanations of priming effects are similar to the selective accessibility account of anchoring. Many researchers explain priming effects using a spreading activation account of cognitive processing, to argue that the prime activates similarly valenced general concepts [30] or similarly valenced memories [36]. This activation would then translate into increased processing speed, because the activated information is more accessible. There is also evidence that affective priming can influence the evaluations of unrelated social and non-social stimuli, so that the affective evaluation of targets shows greater consistency with the valence of previously subliminally presented primes [37].

In summary, the research on affective priming suggests that the influence that the first image, in our experiments, has on the second image, is actually not dependent on whether the first image is processed for a long time. Instead, affective priming research indicates that the affective content can be processed relatively quickly and, in spite of the actual speed in which it is processed, can still have an influence on subsequent judgements. In our experiment, it seems reasonable to suppose that, in either condition, initial exposure will lead participants to rapidly form an affective impression of the image. But it is also possible that, as participants have longer than 500 milliseconds to view the first image, affective priming is not relevant, as the longer time scale provides them with ample time to process the image more deliberately. In order to rule out the length of time that participants have to process the first image as an explanation for the key effect, we carried out Experiment 2 in this study, which controls the amount of time people have to process the first image and make their ratings. The predictions are the same as those in Experiment's 1 – 3 in White et al. [9]; in the PN condition, the second image will be rated more negatively in the double rating condition than in the single rating condition and vice versa for the NP condition.

1.4. Boundary conditions

All three experiments conducted by White et al. [9] involved affective evaluations and visual stimuli. Perhaps there is something specific about this type of judgement and stimuli, which make them more susceptible to constructive effects? For example, the apparent impression we have that we can entertain positive and negative emotions concurrently [38] perhaps makes emotional stimuli more ambiguous and, therefore, easier to perceive in different ways. It seems clear that the QP prediction for the constructive influence of stimulus evaluations will not always apply, e.g., if a participant is shown the picture of a 'hammer' and is asked to name it, one hardly expects a constructive process. But, a critical aspect of this research is establishing whether the QP prediction can generalize to at least some other kinds of stimuli and judgments, which will in turn inform the boundary conditions in the applicability of the model.

In Experiment 3, we test whether the QP model applies to judgements of the trustworthiness of celebrities and well-known people. Making social judgements about people, and in particular judging their trustworthiness, has been described as essential for survival [39,40]. Research has shown that people make judgements about the trustworthiness of others automatically. They can judge the trustworthiness of a stranger's face, with as little as 50-100ms exposure to their image [41,42]. Trustworthiness can be assessed using only facial images. Studies have found that even when faces are unfamiliar to participants, there is a large degree of consensus between participants about the trustworthiness of those faces [43], even for strangers [39,44].

This research suggests that people should easily be able to make judgements about the trustworthiness of celebrities, given that making such a judgement is a basic human ability. This means that we can exert some control over the stimuli, setting up comparisons that will provide the low and high trustworthiness contrast that we need, to examine the prediction of the QP model, that an intermediate evaluation amplifies the contrast between the preceding and a subsequent stimulus. Would a problem arise if participants are unfamiliar with the celebrities we employed as stimuli? This is unlikely to be a confounding consideration. Even if participants rate a celebrity on the basis of their face alone (and, recall, judgments of trustworthiness are considered automatic and can be driven by just the face), within participant judgements of the same stimulus are likely to be consistent. In any case, to avoid any complications from differences in familiarity for the celebrities, we prepared and presented facial images of celebrities in the same manner, as in experiments, where participants are asked to judge the trustworthiness of strangers [39,45].

The use of celebrities as stimuli is not novel in QP cognitive research. A focus of QP modelling has been Moore's [46] result. Moore, using Gallup poll data, found that the American Vice President Al Gore was rated as being less honest and trustworthy, if the previous question was about the honesty and trustworthiness of President Bill Clinton. This order effect can be described using a QP model [14,47]. The QP explanation involves the idea that the first stimulus and the participant's response both provide a context, against which the judgement about the second stimulus is made.

In the current experiment, we are not interested in order effects, but in whether the absence of an intermediate trustworthiness judgement about the first celebrity influences the trustworthiness rating of the second celebrity. The prediction from the QP model is identical to that in Experiments 1 and 2. We expect that whether or not someone provides an intermediate rating of trustworthiness, regarding the first celebrity, will influence their rating of the second celebrity. So, for the PN condition, where a more trustworthy celebrity is seen first, before viewing a less trustworthy celebrity, an intermediate rating of trustworthiness for the first celebrity will result in a less trustworthy rating for the second celebrity, than if the first celebrity was not rated. The prediction for the NP condition, is reversed, in that the intermediate judgement will make the second celebrity appear more trustworthy. In both cases, the intermediate evaluation is predicted to increase the difference in the perception of trustworthiness for the two celebrities.

2. Experiment 1: the influence of an intermediate evaluation using a binary judgement

2.1. Participants and design

Forty City University London students participated in the experiment for course credit (31 women, average age 22.03 years). We employed a within-subjects design with two independent variables: advert order (PN, NP, neutral) and rating (single, double). The inclusion of a neutral condition for advert order (to mean one neutral stimulus was presented after another neutral one) was the only difference between this experiment and White et al.'s [9] original experiments. It was thought that these stimuli might serve to accentuate the positivity or negativity of the other stimuli.

2.2. Stimuli

The same positive and negative images from White et al.'s [9] Experiment 2 were used but rather than using the same filler adverts as in the previous experiment, we created a new set of adverts for a camera which involved neutral images. These neutral images were drawn from GAPED. The neutral stimuli were evaluated in the same way as the experimental stimuli (i.e. single and double rated). Stimulus materials were presented using Superlab.

2.3. Procedure

The procedure was the same as that employed in White et al. [9] (see Figure 1) with only the rating of the adverts being different. Participants were told that they would see several adverts and that for each advert, when asked, they should answer the question 'how does this advert make you feel?', by pressing the appropriate key to indicate one of two possible choices, "Z: Happy or M: Unhappy" (keys were appropriately labelled). Trials were organized into two blocks. One block contained the six single rating PN smartphone adverts, six double rating PN insurance adverts, six single rating NP insurance adverts, six double rating NP smartphone adverts, six single rating neutral camera adverts and six double rating neutral camera adverts. The other block contained the same adverts, but switching the requirement for single vs. double rating. Block order was counterbalanced between participants and trial order within blocks was randomized.

2.4. Results

As for White et al.'s [9] previous experiments, as the valence of the images had been established in the pilot study, we excluded four participants whose ratings for the first rated images in the double rating condition were over one standard deviation below the mean for positive adverts (M=0.87, SD=0.23) and one standard deviation above the mean for negative adverts (M=0.13, SD=0.21).

Happy responses were coded “1” and unhappy responses were coded “0”. We conducted a three (advert order: PN, NP, neutral) × two (rating: single, double) repeated measures ANOVA on participant ratings for the second adverts. There was a main effect of advert order (F(2,70)=93.23, p<.001), but not of rating (F(1,35)=0.13, n.s.). Importantly, the advert order × rating interaction was significant (F(2,70)=4.74, p=.012). Paired samples t-tests showed that in the NP condition, the positive advert was more likely to be rated positively, when there was an intermediate rating (M=0.93, SD=0.10), than without an intermediate rating (M=0.86, SD=0.25; t(35)=-2.18, p=.035; d=0.37). For the PN condition, the second negative advert was more likely to be rated negatively, when there was an intermediate rating (M=0.18, SD=0.22), than when there was no intermediate rating (M=0.23, SD=0.25) but not significantly so (t(35)=1.41, p=.17; d=0.22). In the neutral condition, there was no significant difference between single rated (M=0.37, SD=0.25) and double rated (M=0.35, SD=0.24) second neutral adverts (t(35)=1.19, n.s.). With the exception of the non-significant trend (but in the right direction) for the PN order, these results replicate White et al. [8,9].

3. Experiment 2: Controlling for the amount of time to process the first stimulus

3.1. Participants, design and stimuli

Twenty-five City University London mostly undergraduate students participated in the experiment for course credit (15 women, average age 24.84 years). We employed a within-subjects design with two independent variables: advert order (PN, NP) and rating (single, double). The same stimuli as used in White et al. [9] Experiment 2 were used in this experiment.

3.2. Procedure

<insert figure 2>

The timings for the presentation and rating of all adverts were controlled (see Figure 2). Based on an analysis of the reaction times for rating adverts in White et al.’s [9] Experiment 2 (M=3259 milliseconds, SD=2412 milliseconds), in the current experiment, participants were given 5000 milliseconds to view the first image in the double rating condition, followed by 3300 milliseconds to rate it. If participants took longer than 3300 milliseconds to rate the image, they were presented with a message informing them that they had been too slow and they proceeded to the next image, without rating the first. In the single rating condition, they were given a total of 8300 milliseconds to view the first image. The same timings were used when participants rated the second image in both single and double rating conditions. In all other respects, the procedure, including ordering of trials, block order and counter balancing was identical to that used in in White et al. [9] Experiment 2.

3.3. Results

Trials in the double rating condition in which participants failed to respond in time to the first image were eliminated from analysis. One participant was too slow on 17 out of 72 trials (23.5%), and so the calculation of average ratings for PN and NP single rated adverts was not possible. This participant was not included in further analyses. As for previous experiments, as the valence of the images had been established in a pilot study, we excluded one participant whose ratings for the first rated images in the double rating condition were over one standard deviation below the mean for positive adverts (M=6.48, SD=1.20) and one standard deviation above the mean for negative adverts (M=3.36, SD=1.19).

We conducted a two (advert order: PN, NP) × two (rating: single, double) repeated measures ANOVA on the ratings for the second adverts. There was a main effect of advert order (F(1,22)=69.51, p<.001), but not of rating (F(1,22)=3.22, n.s). The advert order × rating interaction was significant (F(1,22)=12.51, p=.002). Paired samples t-tests showed that in the NP condition, the positive advert was rated more positively when there was an intermediate rating (M=6.76, SD=1.25) than without an intermediate rating (M=6.07, SD=1.46; t(22)=3.77, p=.001; d=0.79). For the PN condition, the second negative advert was rated more negatively when there was an intermediate rating (M=3.34, SD=1.25) than when there was no intermediate rating (M=3.77, SD=1.16; t(22)=-2.59, p=.017; d=0.55). These results exactly replicate White et al. [8,9], showing that length of exposure or processing time is not a viable explanation for the key QP prediction.

4. Experiment 3: The influence of an intermediate evaluation on judgements of celebrity trustworthiness

4.1. Stimuli

To ensure that celebrity pairs would be familiar to participants and that they were differentiated in terms of their perceived trustworthiness, we conducted a pilot study to evaluate each celebrity’s trustworthiness. Further details on the pilot can be found in Appendix B and the results are shown in Table 1. The results indicated that five celebrity pairs in particular were not very well-recognised as compared with the other celebrity pairs, since the number of don't knows for these pairs was over 1 standard deviations above the mean (M=2, SD=3). These pairs were eliminated leaving 22 pairs. We decided to retain the remainder of celebrities, in spite of the small degree of difference in trustworthiness between some pairs, in order to ensure that we had a sufficient number of stimuli in the study.

<insert table 1>

Within each pair there was a celebrity perceived to be less trustworthy (N) than the other celebrity (P). In this experiment we are not interested in whether or not participants can recognise a celebrity, only in how trustworthy they judge the celebrities to be, based on whatever it is they know about them, or just their impression of their faces, [cf. 39,44]. Recall, as already noted, it is possible that participants might not recognize a celebrity, but still provide a rating of trustworthiness (based on facial information), rather than answering do not know. Partly so as to mitigate variability in such responses, we broadly matched images on colour, clothing or background, and emotional expression. Moreover, following procedure in related experiments [39,45], the images were standardised by cropping to the celebrity's head and shoulders, converting to grayscale and scaling to the same size. We also included the name of the celebrity, under their image, to aid recognition.

We constructed a second set of stimuli, which was identical to the first, except that the order of presentation of celebrity pairs was switched. For example, in one set of stimuli, Angelina Jolie was shown first followed by Brad Pitt, as a celebrity pair in the NP condition (Angelina Jolie was rated as being less trustworthy than Brad Pitt in the pilot). In the second set of stimuli, Brad Pitt was shown first followed by Angelina Jolie, as a celebrity pair in the PN condition. Stimulus materials were presented using Superlab.

4.2. Participants and design

Given the novelty of the task and the uncertainty about consistency in ratings of trustworthiness of celebrities we recruited more participants than in previous experiments. Eighty-one mostly undergraduate students from Swansea University and City University London participated in the experiment for course credit (69 women, average age 20.15 years). We employed a within-subjects design with two independent variables: order of celebrity trustworthiness (PN, NP) and rating (single, double). We use the same notational convention as in previous experiments to represent high and low levels of trustworthiness. So P represents higher trustworthiness and N represents lower trustworthiness.

4.3. Procedure

Participants were randomly assigned to view one of the two sets of stimuli (the sets only differed in the order of faces in each pair), as in other experiments.

Participants were then told that they would be shown various well-known people and that they would be asked to evaluate their trustworthiness. They rated each celebrity's trustworthiness on a 9 point scale, with anchors "1: very untrustworthy to 9: very trustworthy". They were also given the option of pressing "D" (corresponding to "don't know") if they did not know the celebrity at all.

Each trial involved the presentation of a celebrity followed by a request for rating (double rating condition) or not (single rating condition), followed by the second celebrity image and a final request for a rating. Trials were organized into two blocks (within participants). One block contained five single rating PN celebrity pairs, six double rating PN celebrity pairs, six single rating NP pairs and five double rating NP pairs. The other block contained the same pairs, but switching the requirement for single vs. double rating (i.e., participants rated pairs twice, once in the single rating condition, once in the double rating one). Trial order within blocks was randomized.

4.4. Results

Of the eighty-one participants who took part in the experiment, seven answered "don't know" to more than 50% of the trials, which meant there was insufficient data to analyse their responses. These seven were eliminated from further analysis[2].

As in previous studies, we checked the ratings to ensure they were in line with the ratings for trustworthiness that had been established in the pilot study. Two celebrity pairs, Angelina Jolie & Brad Pitt and David Beckham & Victoria Beckham were not rated as they had been in the pilot. Angelina Jolie should have been rated less trustworthy than Brad Pitt, but the reverse was observed in the main experiment. Similarly for David Beckham, who should have been rated less trustworthy than Victoria Beckham but was rated as more trustworthy[3]. As these pairs had been explicitly chosen because they were perceived in a way suitable for the condition they were in, they were eliminated from further analysis.

We then followed the same procedure as was used previously to check the ratings of the first celebrities in the double rating condition. For the first set of stimuli, 11 participants showed ratings that were either over one standard deviation below the mean for trustworthy celebrities (M=6.10, SD=1.21) or above the mean for less trustworthy celebrities (M=4.10, SD=1.02). For the second set of stimuli, another 11 participants showed ratings that were either over one standard deviation below the mean for trustworthy celebrities (M=5.5, SD=1.22) or above the mean for less trustworthy celebrities (M=4.40, SD=1.25). These 22 participants were eliminated from the analysis, leaving 52 participants.

We conducted a two (order of celebrity trustworthiness: PN, NP) × two (rating: single, double) repeated measures ANOVA on the ratings for the second celebrities. There was a main effect of order ($F(1,51)=81.19$, $p<.001$), but not of rating ($F(1,51)=0.03$, n.s.). The order × rating interaction was significant ($F(1,51)=7.11$, $p=.01$). Paired samples t-tests showed that, with the intermediate rating, the second celebrity was rated less trustworthy in the PN condition, compared to without the intermediate rating (M=4.36, SD=0.98 vs. M=4.54, SD=0.94; $t(51)=-2.23$, $p=.029$; $d=0.3$) and the trustworthy celebrity was rated more trustworthy in the NP condition (M=6.02, SD=0.90 vs. M=5.85, SD=1.05; $t(51)=2.23$, $p=.029$; $d=0.3$). In other words, the intermediate ratings increased the difference in trustworthiness between the two persons, a result which exactly replicates the findings of White et al. [8,9], with judgments of trustworthiness, instead of affective evaluation.

5. Discussion

QP offers a relatively simple mechanism by which the constructive effects (of a specific kind) of making a judgement can be modelled. A model based on QP principles was proposed by White et al. [8,9], to explain the results observed in their experiments, namely that whether or not someone articulates an affective evaluation for a positively or negatively valenced stimulus, can influence how an oppositely valenced stimulus is rated. With the present experiments, we address some methodological limitations in the experiments of White et al. [8,9] and, moreover, we extend their results with judgments of a completely different kind.

Specifically, Experiment 1 was an ambitious attempt to replicate the results of the previous experiments, using a binary response format for participants to indicate whether the stimuli made them feel happy or unhappy. The results of the experiment partially supported the interaction predicted by the quantum model (one of the predicted differences was significant, for the other there was a non-significant trend in the expected direction). In the NP condition, the intermediate rating increased the probability of the second positive advert being rated positively. In the PN condition, the probability of a positive rating for the second negative advert was lower following an intermediate rating than without, but the difference was not significant. No differences were observed between single and double rated neutral adverts.

[2] There were a number of foreign students taking part in the study, which might explain why some did not know the particularly UK-centric set of celebrities.
[3] This attests to the vagaries of public opinion regarding people in the media spotlight. The pilot was conducted in September 2012 and experimental data collected over 2012 and 2013. We can only assume that during that time these particular celebrities had demonstrated behaviour that led the public to perceive them as being more or less trustworthy than they were thought to be in 2012. Alternatively, the data for the pilot was collected in Swansea, UK and experimental data was collected in Swansea and London. Perhaps there is something different about the populations of students in these two locations that lead them to perceive these two celebrity pairs differently. Regardless, the source of this discrepancy is not relevant to the objectives of the study, as long as the materials presented were perceived in the assumed way by participants.

In Experiment 2, we also observed the same interaction, as predicted by the quantum model. We suggested that one possible reason for the difference in ratings of the second image between single and double rating conditions was the greater length of time that participants had to process the first image in the double rating condition, compared with the single rating condition. The logic here is that more processing time could mean that participants develop a mental representation that has greater saliency, which in turn may increase the contrast of the original stimulus with the second oppositely valenced stimulus, in relation to the same comparison, in the single rating condition. The results in this experiment suggest that this is an unlikely explanation, because, as was shown, when we equalised the length of time that participants had to process the first image (regardless of whether a rating was made or not), the same key interaction was observed.

For Experiment 3, the hypothesis regarding the effect of an intermediate judgement of trustworthiness, for both the PN and NP conditions, was confirmed. When a more trustworthy celebrity was rated first, the trustworthiness of the second celebrity was lower, than without the intermediate rating. Likewise, when a less trustworthy celebrity was rated first, the trustworthiness of the second celebrity was higher, than without the intermediate rating.

Overall, the results of all three experiments provide further support for White et al.'s [8,9] QP model. Furthermore, the results of Experiment 3 suggest that the QP model is not limited in applicability to affective evaluations of visual stimuli and that there are other domains in which this effect can be observed. Note an additional point regarding the results of Experiment 3. In White et al.'s [9] previous research, the images selected were chosen because they were unrelated, even though they were still in the same category of advertisements. One could argue that the stimuli used in Experiment 3, in the current paper, were more related (e.g., husband and wife). However, the predictions of the QP model work only as long as the stimuli are such that a definite opinion about one (intuitively at least) changes our perspective about the other, that is, that they are incompatible (in the QP technical sense). If this assumption about incompatibility holds, then resolving the trustworthiness of one celebrity should change our perspective regarding the trustworthiness of the other. Conversely, compatibility implies that the question about the trustworthiness of both celebrities could be resolved concurrently, thus preventing any putative constructive effects. Had we used, for example, Ant & Dec[4] as celebrity pairs, then a judgement of Ant's trustworthiness would not be expected to interfere with a judgment on Dec's trustworthiness.

Experiment 3 showed that the constructive effect predicted by the QP model could be extended to different stimuli and judgements. Regarding future extensions, our present work points to a need to consider more carefully issues concerning relatedness and compatibility, which are indeed ongoing questions in the QP research programme. The results of Experiment 3 do suggest that using known and unknown faces, as well as pairs of individuals that can be assumed to be compatible and incompatible, might be a useful approach to help examine more finely grained predictions from the QP model.

[4] Ant & Dec (Anthony McPartlin and Declan Donnelly) are a pair of British entertainers who only ever work with each other and are generally seen by the public together at all times. Indeed, many members of the British public are unable to say who is Ant and who is Dec. Their interchangeability and high frequency of co-occurrence suggest that an appropriate representation for Ant and Dec would be a compatible one.

Additional Information

Acknowledgments
We would like to thank James Yearsley for many helpful discussions.

Funding Statement
EMP was supported by Leverhulme Trust grant RPG-2013-004 and Air Force Office of Scientific Research (AFOSR), Air Force Material Command, USAF, grant FA 8655-13-1-3044. The U.S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

Data Accessibility
The datasets supporting this article have been uploaded as part of the Supplementary Material.

Competing Interests
We have no competing interests.

Authors' Contributions
L.C.W and E.M.P designed research; L.C.W and A.B.D performed the research; L.C.W and E.M.P analysed data; L.C.W, A.B.D and E.M.P wrote the paper.

References

1. Ariely D, Norton MI. 2008 How actions create – not just reveal – preferences. *Trends Cogn Sci* **12**, 13-16. (doi:10.1016/j.tics.2007.10.008)
2. Kahneman D, Snell J. 1992 Predicting a changing taste: Do people know what they will like? *J Behav Decis Making* **5**, 187-200. (doi:10.1002/bdm.3960050304)
3. Payne JW, Bettman JR, Johnson J. 1993 *The Adaptive Decision Maker*. Cambridge, UK: Cambridge University Press.
4. Sharot T, Velasquez CM, Dolan RJ. 2010 Do decisions shape preference? : Evidence from blind choice. *Psychol Sci* **21**, 1231–1235. (doi:10.1177/0956797610379235)
5. Sherman SJ. 1980 On the self-erasing nature of errors of prediction. *J Pers Soc Psychol* **39**, 211-221. (doi:10.1037/0022-3514.39.2.211)
6. Slovic P. 1995 The construction of preference. Am Psychol **50**, 364-371. (doi:10.1037//0003-066X.50.5.364)
7. Brehm JW. 1956 Post-decision changes in the desirability of choice alternatives. J Abnorm Soc Psych **52**, 384-389. (doi:10.1037/h0041006)
8. White LC, Pothos EM, Busemeyer JR. 2013 A quantum probability perspective on the nature of psychological uncertainty. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (eds M Knauff, M Pauen, N Sebanz, I Wachsmuth), pp. 1599–1604. Austin TX: Cognitive Science Society.
9. White LC, Pothos EM, Busemeyer, JR. 2014 Sometimes it does hurt to ask: the constructive role of articulating impressions. *Cognition* **133**, 48-64. (doi:10.1016/j.cognition.2014.05.015)
10. Dan-Glauser ES, Scherer KR. 2011 The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavioral Research* **43**, 468-477. (doi:10.3758/s13428-011-0064-1)
11. Hogarth RM, Einhorn HJ. 1992 Order effects in belief updating: the belief-adjustment model. *Cognitive Psychol* **24**, 1-55. (doi:10.1016/0010-0285(92)90002-J)
12. Busemeyer JR, Pothos E, Franco R, Trueblood, JS. 2011 A quantum theoretical explanation for probability judgment 'errors'. *Psychol Rev* **118**, 193-218. (doi:10.1037/a0022542)
13. Trueblood JS, Busemeyer JR. 2011 A quantum probability account of order effects in inference. *Cognitive Sci* **35**, 1518-1552. (doi:10.1111/j.1551-6709.2011.01197.x)
14. Wang Z, Solloway T, Shiffrin RM, Busemeyer JR. 2014 Context effects produced by question orders reveal quantum nature of human judgments. *P Natl Acad Sci USA*. (doi:10.1073/pnas.1407756111)
15. Bruza PD, Kitto K, Nelson D, McEvoy CL. 2009 Is there something quantum-like about the human mental lexicon? *J Math Psychol* **53**, 362-377. (doi:10.1007/978-3-642-04417-5_1)
16. Aerts D, Aerts S. 1995 Applications of quantum statistics in psychological studies of decision processes. *Found Sci* **1**, 85–97. (doi:10.1007/BF00208726)
17. Atmanspacher H, Filk T. 2010 A proposed test of temporal nonlocality in bistable perception. *J Math Psychol* **54**, 314–21. (doi:10.1016/j.jmp.2009.12.001)
18. Pothos EM, Busemeyer JR, Trueblood JS. 2013 A quantum geometric model of similarity. *Psychol Rev* **120**, 679–696. (doi:10.1037/a0033142)
19. Busemeyer JR, Bruza P. 2011 *Quantum Models of Cognition and Decision Making*. Cambridge, UK: Cambridge University Press.
20. Haven E, Khrennikov AY. 2013 *Quantum Social Science*. Cambridge: Cambridge University Press.
21. Pothos EM, Busemeyer JR. 2013 Can quantum probability provide a new direction for cognitive modeling? *Behav Brain Sci* **36**, 255-327. (doi:10.1017/S0140525X12001525)
22. Wang ZJ, Busemeyer JR, Atmanspacher H, Pothos EM. 2013 The potential for using quantum theory to build models of cognition. *Top Cog Sci* **5**, 672-688. (doi:10.1111/tops.12043)
23. Hughes RIG. 1989 *The Structure and Interpretation of Quantum Mechanics*. Cambridge, MA: Harvard University Press.
24. Isham CJ. 1989 *Lectures on quantum theory*. Singapore: World Scientific.
25. Peres A. 1998 *Quantum Theory: Concepts and Methods*. Kluwer Academic.
26. Bohr N. 1958 *Atomic Physics and Human Knowledge*. New York: Wiley.
27. Bargh JA, Chaiken S, Govender R, Pratto F. 1992 The generality of the automatic evaluation activation effect. *J Pers Soc Psychol* **62**, 893–912. (doi:10.1037/0022-3514.62.6.893)
28. Damasio AR. 1994 *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
29. Duckworth KL, Bargh JA, Garcia M, Chaiken, S. 2002 The automatic evaluation of novel stimuli. *Psychol Sci* **13**, 513-519. (doi:10.1111/1467-9280.00490)

30. Fazio RH, Sanbonmatsu DM, Powell MC, Kardes FR. 1986 On the automatic activation of evaluations. *J Pers Soc Psychol* **50**, 229–238. (doi:10.1037/0022-3514.50.2.229)
31. Greenwald AG, Klinger MR, Liu TJ. 1989 Unconscious processing of dichoptically masked words. *Mem Cognition* **17**, 35–47. (doi:10.3758/BF03199555)
32. LeDoux JE. 1996 *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
33. Zajonc RB. 1980 Feeling and thinking: Preferences need no inferences. *Am Psychol* **35**, 151−175. (doi:10.1037//0003-066X.35.2.151)
34. Hermans D, De Houwer J, Eelen P. 1994 The affective priming effect: Automatic activation of evaluative information in memory. *Cognition Emotion* **8**, 515–533. (doi:10.1080/02699939408408957)
35. Klauer KC, Rossnagel C, Musch J. 1997 List-context effects in evaluative priming. *J Exp Psychol Learn* **23**, 246–255. (doi:10.1037//0278-7393.23.1.246)
36. Bargh JA, Chaiken S, Raymond P, Hymes C. 1996 The automatic evaluation effect: Unconditional automatic evaluation activation with a pronunciation task. *J Exp Soc Psychol* **32**, 104–128. (doi:10.1006/jesp.1996.0005)
37. Ferguson MJ, Bargh JA, Nayak DA. 2005 After-affects: How automatic evaluations influence the interpretation of subsequent, unrelated stimuli. *J Exp Soc Psychol* **41**, 182–191. (doi:10.1016/j.jesp.2004.05.008)
38. Brehm JW, Miron AM. 2006 Can the simultaneous experience of opposing emotions really occur? *Motiv Emotion* **30**, 13-30. (doi:10.1007/s11031-006-9007-z)
39. Rule NO, Krendl AC, Ivcevic Z, Ambady N. 2013 Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Attitudes and Social Cognition* **104**, 409-426. (doi:10.1037/a0031050)
40. Tanner RJ, Maeng A. 2012 Celebrity facial cues influence trust and preference. *J Consum Research* **39**, 769-783. (doi:10.1086/665412)
41. Todorov A, Pakrashi M, Oosterhof NN. 2009 Evaluating faces on trustworthiness after minimal time exposure. *Soc Cognition* **27**, 813–33. (doi:10.1521/soco.2009.27.6.813)
42. Willis J, Todorov A. 2006 First impressions: Making up your mind after 100 milliseconds exposure to a Face. *Psychol Sci* **17**, 592–98. (doi:10.1111/j.1467-9280.2006.01750.x)
43. Engell AD, Haxby JV, Todorov A. 2007 Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J Cognitive Neurosci* **19**, 1508–1519. (doi:10.1162/jocn.2007.19.9.1508)
44. Porter S, England L, Juodis M, Brinke LT, Wilson K. 2008 Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. Can J Beh Sci **40**, 171-177. (doi:10.1037/0008-400X.40.3.171)
45. Rule NO, Ambady N. 2008 The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychol Sci* **19**, 109-111. (doi:10.1111/j.1467-9280.2008.02054.x)
46. Moore DW. 2002 Measuring new types of question-order effects. *Public Opin Quart* **66**, 80-91. (doi:10.1086/338631)
47. Wang Z, Busemeyer JR. 2013 A quantum question order model supported by empirical tests of an a priori and precise prediction. *Top Cog Sci* **5**. (doi:10.1111/tops.12040)

Tables

Table 1. Pilot study celebrity trustworthiness ratings

High Trustworthy Celebrities		Low Trustworthy Celebrities			
Name	M	Name	M	DK	Difference
Al Gore*	4.25	Bill Clinton*	4.24	9	0.01
Bill Gates	6.71	Steve Jobs	6.67	2	0.04
Ed Milliband*	4.27	David Milliband*	4.21	6	0.05
William Hague	3.14	George Osborne	3.07	4	0.07
Brad Pitt	5.82	Angelina Jolie	5.65	0	0.18
Victoria Beckham	5.53	David Beckham	5.35	0	0.18
Gordon Brown*	3.50	Ed Balls*	3.22	8	0.28
Catherine Zeta Jones	6.06	Michael Douglas	5.59	0	0.47
Zara Phillips	6.00	Mike Tindall	5.29	0	0.71
John Lennon*	5.75	Yoko Ono*	5.00	8	0.75
Beyonce	7.29	Jay Z	6.53	0	0.76
Stephen Merchant	6.43	Ricky Gervais	5.50	3	0.93
Katie Holmes	5.81	Tom Cruise	4.81	1	1.00
Vince Cable	3.92	Nick Clegg	2.88	5	1.03
Condolezza Rice*	4.11	George Bush*	2.94	8	1.17
Prince Charles	5.47	Camilla Parker Bowles	4.29	0	1.18
Dawn French	7.00	Lenny Henry	5.65	0	1.35
Tess Daley	6.56	Vernon Kay	5.13	1	1.44
Gary Barlow	6.82	Robbie Williams	5.12	0	1.71
Paul McCartney	5.59	Heather Mills	3.87	2	1.72

Charlotte Church	5.41	Gavin Henson	3.53	0	1.88
Barack Obama	6.76	Hilary Clinton	4.65	0	2.12
Boris Johnson	5.41	David Cameron	3.24	0	2.18
Coleen Rooney	5.65	Wayne Rooney	3.35	0	2.29
Katy Perry	6.12	Russell Brand	3.41	0	2.71
Peter Andre	5.94	Katie Price	2.88	0	3.06
Cheryl Cole	6.06	Ashley Cole	2.29	0	3.76

Notes: M=Mean rating on a 9 point scale (1 is "Very untrustworthy" and 9 is "Very trustworthy"). DK=number of times that someone responded 'don't know' to one or both of a pair. Difference =difference in Means. *Indicates celebrity pair that was not used in the main experiment. Celebrity pairs are matched by row, so the high and low trustworthiness classification is to be interpreted only within individual rows.

Figure and table captions

Figure 1. Sample adverts and procedure for the NP condition used in Experiment 2 (Source: White et al., (2014). Sometimes it does hurt to ask: the constructive role of articulating impressions. Cognition, 133, 48-64. Adapted with permission)

Figure 2. Procedure for Experiment 2: sample advert used in NP condition and procedure for presentation of single and double rated adverts.

Table 1. Pilot study celebrity trustworthiness ratings