Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2013). Choice Blindness and Preference Change: You Will Like This Paper Better If You (Believe You) Chose to Read It! *Journal of Behavioral Decision Making*. DOI: 10.1002/bdm.1807

Abstract: Choice blindness is the finding that participants both often fail to notice mismatches between their decisions and the outcome of their choice and, in addition, endorse the opposite of their chosen alternative. But do these preference reversals also carry over to future choices and ratings? To investigate this question, we gave participants the task of choosing which of a pair of faces they found most attractive. Unknown to them, we sometimes used a card trick to exchange one face for the other. Both decision theory and common sense strongly suggest that most people would easily notice such a radical change in the outcome of a choice. But that was not the case: no more than a third of the exchanges were detected by the participants. We also included a second round of choices using the same face pairs, and two stages of post-choice attractiveness ratings of the faces. This way we were able to measure preference strength both as choice consistency and by looking at measures of rating differences between chosen and rejected options. We found that the initially rejected faces were chosen more frequently in the second choice, and the perceived attractiveness of these faces was increased even in uncoupled individual ratings at the end of the experiment. This result is discussed in relation to Chen and Risen's recent criticism of the Free Choice Paradigm, as it shows that choices can affect future preferences.

Keywords: Choice blindness, decision making, preference change, preference change through choice, the free choice debate

Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1807

Choice Blindness and Preference Change: You Will Like This Paper Better If You (Believe You) Chose to Read It!

PETTER JOHANSSON[1,2]*, LARS HALL[1], BETTY TÄRNING[1], SVERKER SIKSTRÖM[3] and NICK CHATER[4]

[1]*Lund University Cognitive Science, Lund University, Sweden*

[2]*Swedish Collegium for Advanced Study, Uppsala University, Sweden*

[3]*Department of Psychology, Lund University, Sweden*

[4]*Behavioural Science Group, Warwick Business School, University of Warwick, UK*

ABSTRACT

Choice blindness is the finding that participants both often fail to notice mismatches between their decisions and the outcome of their choice and, in addition, endorse the opposite of their chosen alternative. But do these preference reversals also carry over to future choices and ratings? To investigate this question, we gave participants the task of choosing which of a pair of faces they found most attractive. Unknown to them, we sometimes used a card trick to exchange one face for the other. Both decision theory and common sense strongly suggest that most people would easily notice such a radical change in the outcome of a choice. But that was not the case: no more than a third of the exchanges were detected by the participants. We also included a second round of choices using the same face pairs, and two stages of post-choice attractiveness ratings of the faces. This way we were able to measure preference strength both as choice consistency and by looking at measures of rating differences between chosen and rejected options. We found that the initially rejected faces were chosen more frequently in the second choice, and the perceived attractiveness of these faces was increased even in uncoupled individual ratings at the end of the experiment. This result is discussed in relation to Chen and Risen's recent criticism of the Free Choice Paradigm, as it shows that choices can affect future preferences. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS choice blindness; decision making; preference change; preference change through choice; the free choice debate

INTRODUCTION

An influential tradition in psychology has held that choices influence preferences; we come to prefer what we have chosen more, whereas the rejected alternative is liked even less (Brehm, 1956). Using what is known as the free-choice paradigm (FCP), this effect has been demonstrated for a wide range of choices (e.g. Coppin, Delplanque, Cayeux, Porcherot, & Sander, 2010; Gerard & White, 1983; Sharot, de Martino, & Dolan, 2009; Shultz, Leveille, & Lepper, 1999) and for populations as different as amnesic patients (Lieberman, Ochsner, Gilbert, & Schacter, 2001), young children (Egan, Santos, & Bloom, 2007) and Capuchin monkeys (Egan, Santos, & Bloom, 2010).

Recently, this line of research has come under attack. Chen (2008) formally argued that all versions of FCP fall prey to a set of similar methodological errors. Given the design of the FCP paradigm, it is impossible to tell if there is a true effect of choice on preference or if the result found is a statistical artefact. Chen and Risen (2010) demonstrated this point in an elegant empirical study. They set up a typical FCP experiment, in which the participants first rate a number of paintings, then choose between two of them that are close in rank and then rate all the paintings again (i.e. a rating–choice–rating (RCR) procedure). This gave the expected result that ratings of the chosen paintings were increased whereas the ratings of the rejected items were reduced, a so-called spreading of the alternatives, which is the standard measure showing that choices influence preferences. However, in a second condition of the experiment, the two rating sessions were performed after each other, followed by a choice between two options that were close in rank in the first rating (i.e. a rating–rating–choice (RRC) procedure). But despite that the choice now came at the end, and thus in no way could have influenced the second rating, the participants showed the same spread in rating between the two alternatives as in the first condition: the second rating of the later-to-be-chosen object tended to increase, whereas the later-to-be rejected item was devalued. How can this be? All that is needed to explain this result, according to Chen and Risen (2010), is the assumption that participants have 'true' underlying stable preferences but that these preferences might not be perfectly measured by any single instance of rating or choice. Given this assumption, if rating spread is found in RCR, this may be not a *consequence* of the choice but simply a result of using the choice as a way to divide the participants and then look for a preference change in line with the choice. Thus, if the choice tells us that a participant 'actually' prefers A to B, it also makes it more likely that if there is a noise-induced difference between the first and second ratings, this difference would tend to be in the same direction as the choice (see also Izuma & Murayama, 2013, for simulations illustrating this result, and Coppin, Delplanque, & Sander, submitted, for an overview of the debate).

Importantly, Risen and Chen (2010) offered a number of constructive suggestions on how to fix the problems of the FCP and to properly test whether choices have an impact on preferences. According to their arguments, to control for the statistical artefacts inherent in the standard FCP, researchers must either (i) ensure that all participants make the same

*Correspondence to: Petter Johansson, Lund University Cognitive Science, Lund University, 222 22 Lund, Sweden. E-mail: petter.johansson@lucs.lu.se

choice, (ii) control for the information revealed by the choice, (iii) remove the information from the choice or (iv) manipulate the choices that people make. Here, we would like to focus on the fourth and final suggestion:

> A [final] approach that researchers can use is to manipulate the choices that people make and then calculate spread for each participant based on their randomly assigned condition /…/ there are two central problems that need to be solved. First, researchers must effectively manipulate choice. Second, researchers must avoid directly manipulating preferences. (Risen & Chen, 2010, p. 1159)

We believe that our *choice blindness* (CB) methodology is a good example of choice manipulation that meets the criteria and solves the problems identified by Chen and Risen (2010).

Choice blindness

The CB methodology was originally inspired by techniques from the domain of close-up card-magic, which permits a surreptitious manipulation of the relationship between choice and outcome that the participants experience. In Johansson, Hall, Sikström, and Olsson (2005), participants were shown pairs of pictures of female faces and were given the task of choosing which face in each pair they found more attractive. In addition, immediately after their choice, they were asked to verbally describe the reasons for choosing the way they did. Unknown to the participants, on certain trials, a double-card ploy was used to covertly exchange one face for the other. Thus, on these trials, the outcome of the choice became the opposite of what they intended. Both decision theory and common sense strongly suggest that everyone would notice such a radical change in the outcome of a choice. But on the great majority of trials, participants failed to notice the mismatch between choice and outcome, while still being prepared to offer introspective reasons for why they chose the way they thought they had.

We have demonstrated CB for attractiveness of abstract artistic patterns and for male and female faces, both when presented in a live interaction, as described earlier (Johansson et al., 2005; Johansson, Hall, Sikström, Tärning, & Lind, 2006), and in a computerized paradigm (Johansson, Hall, & Sikström, 2008). CB has also been demonstrated in different modalities. Steenfeldt-Kristensen and Thornton (2013) have established CB for tactile choices, and we have extended it to the linguistic domain for purely spoken decisions (Lind, Hall, Breidegard, Balkenius, & Johansson, submitted). We have also shown the effect of CB for the taste of jam and the smell of tea in an ecologically valid supermarket setting (Hall, Johansson, Tärning, Sikström, & Deutgen, 2010). Merkelbach, Jelicic, and Pieters (2010) have applied the phenomenon of CB to the problem of malingering in the clinical domain, and recently, we have established that CB has clinical relevance as a diagnostic instrument to study obsessive-compulsive disorder (Aardema et al., submitted).

Most relevant for the current discussion is the preference reversal at the core of the CB paradigm. Following the assumptions of Chen and Risen (2010), choice and endorsement of an alternative in a non-manipulated (NM) trial are a paradigmatic case of a stated preference. Consequently, CB trials where participants endorse and argue for the originally rejected alternative must be seen as a preference reversal. This reversal can sometimes take fairly dramatic form, such as when CB can be found for moral judgments involving hotly debated topics in the current political debate (Hall, Johansson, & Strandberg, 2012), or when it can be shown to strongly influence voting intentions just a week before a national election (Hall et al., 2013).

Even though the core CB phenomenon represents a simple demonstration of preference reversals, the question remains whether these reversals are ephemeral or lasting, and how the dynamics of preference change plays out over multiple choice and rating points (see Coppin et al., 2012; Sharot, Fleming, Koster, Yu, & Dolan, 2012, for two recent studies concerning the stability of preference modulation over time). It also needs to be clearly demonstrated how CB as a methodology can be used to overcome the problems highlighted by Chen and Risen. For the RCR and RRC, Chen and Risen (2010) argued that a single point preference elicitation (whether by rating or choice, or otherwise) tends to be less informative than the whole series of instances. For CB, does this mean that participants will revert to their originally revealed preference, or will they continue to prefer the option they endorsed in the manipulated (M) trials?

To investigate these questions, we set up a CB experiment with a similar repeated structure as in the typical FCP. Using the same card-trick methodology as in Johansson et al. (2005), we let participants choose between two faces, and for some trials, we reversed their choices. But we also included a second round of choices using the same face pairs, as well as two stages of post-choice attractiveness ratings of the faces (choice–rating–choice–rating). This way, we are able to measure preference strength both as choice consistency, that is, to what extent the participants prefer the same face the first time and the second time they are presented with the pairs, and as a difference measure between the chosen and rejected items.

EXPERIMENT 1: CHOICE BLINDNESS AND PREFERENCE CONSISTENCY

Method

Participants

Forty mainly student participants (24 women) took part in experiment 1 (mean age 24.3 years, $SD = 4.7$), each receiving a cinema ticket as payment for participation.

Material

Fifteen pairs of grey-scale pictures of female faces from a student population were used as stimuli. The pairs were constructed by the experimenters, and an attempt was made to keep physical similarity constant at an intermediate level (i.e. clearly different, but not drastically so). All pictures were rated by 17 independent raters for attractiveness on a scale from 1 to 10 (1 = *not at all attractive* and 10 = *very attractive*). Six of the 15 pairs were chosen as target pairs for the study, with a difference in attractiveness between the face pairs ranging from moderate ($M = .23$) to large ($M = 1.7$) (Figure 1). All pictures were printed and glued on red cardboard, size 7.5 × 9.3 cm. An

DOI: 10.1002/bdm

additional set of copies of the six target pairs were printed and glued onto black laminate plate.

Procedure

The participants were presented with 15 pairs of pictures of female faces and were asked to decide which in each pair they found more attractive. Each pair was shown for 4 seconds and was then put face down on the desk pad. The participants were instructed to point at the face they found the most attractive as soon as the cards were turned down. For the six target pairs, the participants were asked to explain their choice when they picked up the chosen card. For three of these six pairs, a manipulation was introduced, so that the participants received the opposite of their choice (Figure 2). All pairs were presented in a randomized order. The manipulations were randomly distributed over the last 10 presented pairs; however, two manipulated pairs were never presented in a row.

After the participants had explained why they preferred the chosen picture (i.e. the face they did chose or in M trials were led to believe they chose), they were asked to rate the attractiveness of the chosen face on a 10-point scale. When they had stated their rating, the participants were also given the non-chosen card to rate for attractiveness.

After the first round of 15 choices, all the pairs were presented a second time in a randomized order, and the participants were asked to choose the one in each pair they now preferred. In the second round, no manipulations took place, and the participants did not have to explain their choices. Finally, after the second round of choices, the participants were asked to rate all the pictures again. This time, the pictures were not presented in pairs, but one by one in a randomized serial order.

At the end of the experiment, the participants were asked what they thought about the experiment in general and if they thought anything had been strange with the design of the experiment. Finally, the participants were told about the manipulation and the true purpose of the experiment. If the participants at any point during the debriefing indicated that they had felt that something was strange with the pictures, they were asked to look through all pictures again and pick out the ones they thought had been manipulated.

Results

Detection rate

Out of 120 M trials, 11% of the manipulations were immediately detected, 9% of the trials were detected when the originally chosen image was presented during the first rating procedure and a final 13% of the trials were categorized as retrospectively detected if the participants in the debriefing claimed to have experienced something being strange during the experiment (see Johansson et al., 2005, 2006, for more details on detection criteria and debriefing procedure). This means that the large majority of the manipulations remained undetected.

Preference change

We analysed preference strength for the six target pairs by using a number of different measures and at several different

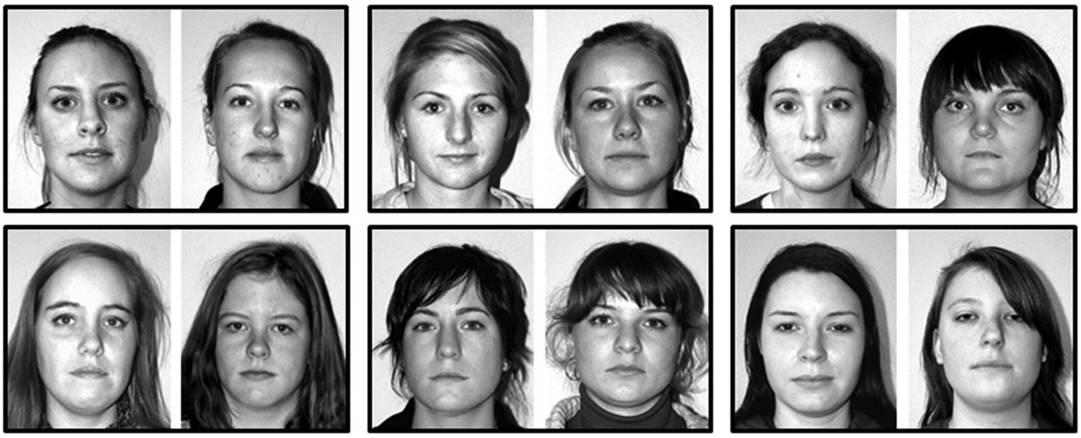

Figure 1. The six target pairs used in the study

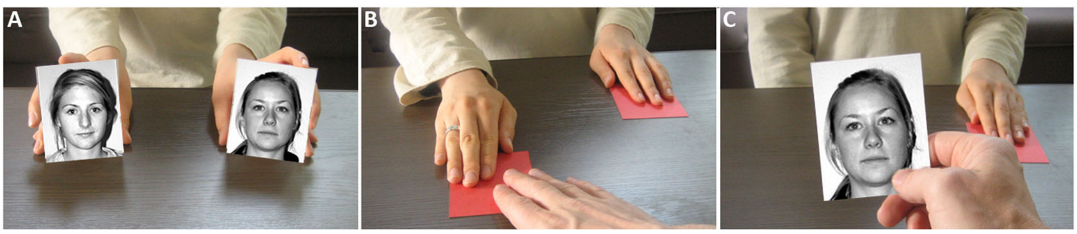

Figure 2. A snapshot of the choice procedure during a manipulation trial. (A) The participant is shown two pictures of female faces and is asked to choose which one he or she finds more attractive. Unknown to the participant, a second card depicting the opposite face is concealed behind the visible alternatives. (B) After 4 seconds, the pictures are turned down, and the participant has to indicate his or her choice by pointing at the preferred card. The experimenter then slides the hidden picture over to the participant and rakes the hidden black card down into her own lap. (C) The participant picks up the picture and is asked to explain why the chosen face was preferred. The participant then rates the face they hold in their hand for attractiveness, and then the rejected face (i.e. the original choice) is picked up and rated in the same fashion

DOI: 10.1002/bdm

time points during the experiment. To provide the most conservative measure of the preference effects, our main analysis was performed on all trials, including detected M trials (see Chen & Risen, 2010, on the general problems of trial exclusion in choice paradigms). As reported earlier, two thirds of the detections were made already during the exposure or first rating phase, which makes it likely that much smaller preference change effects would be seen in detected trials. Thus, to better understand the dynamics of the CB manipulation, we also compare and report the effects of non-detected and detected M trials separately.

Looking at the first attractiveness rating for the NM trials, the chosen faces are rated much higher than the rejected ones ($Mdiff = 1.49$, $SD = 1.1$). This relationship is reversed for the M trials; the mean rating difference between the originally chosen alternative and the originally rejected alternative is negative, indicating an overall change in relative preference in favour of the originally rejected photographs ($Mdiff = -.35$, $SD = 1.2$) (Figure 3). This difference in rating is significantly higher in the NM trials compared with the M ones, $t(238) = 12.46$, $p < .00001$, $d = 1.62$. In a direct comparison between the attractiveness ratings for the originally chosen faces in NM trials ($M = 5.8$, $SD = 1.4$) and M trials ($M = 5.1$, $SD = 1.6$) trials, the ratings were significantly higher in the NM trials, $t(238) = 3.40$, $p < .001$, $d = .44$. Similarly, the ratings for the rejected faces were significantly higher in the M trials ($M = 5.5$, $SD = 1.4$) when compared with the NM trials ($M = 4.3$, $SD = 1.3$), $t(238) = 6.63$, $p < .00001$, $d = .86$. This indicates that the manipulation leads to an increase in the perceived attractiveness for the originally rejected alternatives, as well as a decrease in the rated attractiveness of the originally preferred faces. If we compare the rating difference for non-detected ($Mdiff = -.5$, $SD = .86$) and detected M trials ($Mdiff = -.04$, $SD = 1.6$), we see that both are negative but that the difference is significantly larger for non-detected trials, $t(118) = 2.02$, $p < .05$, $d = .37$.

The second and main preference measure is consistency between choice 1 and choice 2, that is, to what extent the same face in each pair was preferred the second time the choice was made. The level of choice consistency between the first and second choices was close to ceiling for the NM trials (93.3%). However, for the M trials, this number was just 56.6%. Comparing choice consistency between NM and M trials, the difference was highly significant ($\chi^2(1,\ N = 240) = 43.02$, $p < .0001$, $V = .42$). Again, it is interesting to note that for the M trials, the choice consistency is lower for the non-detected trials (43.8%) compared with the detected trials (82.5%) ($\chi^2(1,\ N = 120) = 16.31$, $p < 0001$, $V = .37$).

The attractiveness rating performed at the end of the experiment shows a similar pattern, despite this being a non-paired serial rating of the faces. Comparing the rating difference between the chosen alternative and the rejected alternative, we again find a significant difference, with larger positive difference for NM ($Mdiff = 1.2$, $SD = 1.4$) compared with M trials ($Mdiff = .2$, $SD = 1.4$), $t(238) = 5.62$, $p < .0001$, $d = .73$ (Figure 1). Most of this effect is accounted for by the increased preference for the initially rejected faces: The rating of the originally chosen faces was not significantly lower in the M ($M = 5.5$, $SD = 1.5$) trials compared with the NM trials ($M = 5.7$, $SD = 1.5$), $t(238) = 1.49$, $p = .138$, $d = .19$, but the rejected face was again rated higher in M trials ($M = 5.2$, $SD = 1.6$) when compared with the NM trials ($M = 4.6$, $SD = 1.4$), $t(238) = 3.35$, $p < .001$, $d = .43$. When comparing the final rating difference of non-detected M trials ($Mdiff = .09$, $SD = 1.0$) and detected M trials ($Mdiff = .5$, $SD = 2.0$), we find the effect to be no longer significantly larger, $t(118) = 1.58$, $p = .12$, $d = .29$.[1] The result of the final rating is congruent with that of Sharot, Velasquez, and Dolan (2010), in which the investigators found a preference modulation for the chosen but not for the rejected alternative. The natural interpretation of this is that the belief in having made a choice has a stronger impact on the preference for the chosen object compared with a decrease in preference for the item believed to have been rejected.

The analysis earlier reveals how CB influences future preferences. Participants chose X, but end up endorsing Y, and often keep doing so in repeated choices and ratings. However, to clearly demonstrate how this experiment fulfills the methodological criteria suggested by Risen and Chen (2010) and detailed in the Introduction section of the current article, we need to look at the data from a different perspective.[2] Remember that in Risen and Chen's description on how to properly manipulate the participants' choices in order to show a preference change effect (i.e. the fourth option on their list), not only the choice should be controlled by the experimenter but the outcome should also be randomly assigned for each choice so the participants obtain a predetermined option regardless of what they prefer themselves. With such a random assignment, some participants will

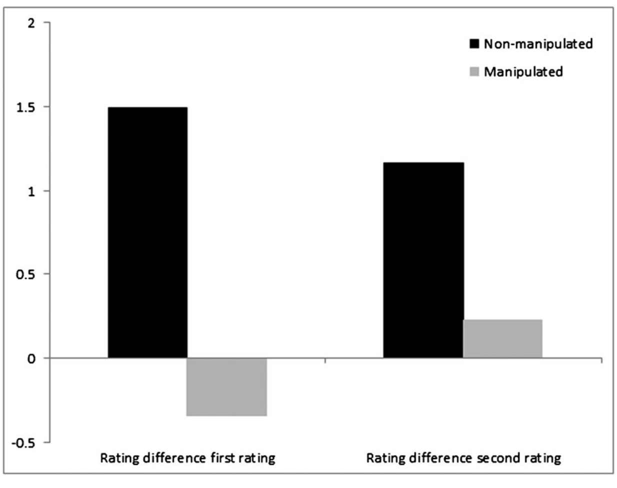

Figure 3. The difference in attractiveness rating between the chosen and rejected alternatives in non-manipulated and manipulated trials for the first and second ratings

[1]There were no gender differences in detection rate, $t(238) = 0$, $p = 1$, choice consistency, $t(238) = .61$, $p = .54$, or in the first rating difference, $t(238) = .18$, $p = .86$, or the second rating difference, $t(238) = 1.35$, $p = .18$. Overall, women rated the faces significantly higher than the men did, both for the chosen (rating 1: $t(238) = 3.59$, $p < .0001$; rating 2: $t(238) = 3.45$, $p < .0001$) and the rejected (rating 1: $t(238) = 4.55$, $p < .0001$; rating 2: $t(238) = 3.08$, $p < .005$) images, but as there was no difference in the rating difference between chosen and rejected faces, this gender difference does not influence the main results.

[2]We would like to thank an anonymous reviewer for suggesting this presentation of the data.

DOI: 10.1002/bdm

experience their 'true' choice, and some will be presented with a manipulated reversed outcome. We divided the dataset of the experiment on the basis of whether the participants were presented with the left face (L-group) or the right face (R-group) in each pair as having been their choice. Some of the participants in the L-group were thus presented with their true choice (i.e. the NM trials), whereas some participants were presented with the opposite of their choice (i.e. the M trials), and vice versa for the R-group. This post-hoc division is statistically equivalent to having made a randomized group assignment before the start of the experiment.

When comparing the second choice distributions of the two groups, as predicted, we find that the right face is strongly preferred in the R-group, and the left face is strongly preferred in the L-group ($\chi^2(1, N = 240) = 31.3$, $p < .00001$). Similarly, creating a difference score by subtracting the attractiveness rating of the right face from that of the left face, we find that the two groups differ significantly in both the first rating (L-group, $M = 1.3$, $SD = 1.3$; R-group, $M = -.6$, $SD = 1.1$), $t(238) = 12.7$, $p < .00001$, $d = 1.65$, and the second rating (L-group, $M = .9$, $SD = 1.5$; R-group, $M = -.1$, $SD = 1.4$), $t(238) = 5.6$, $p < .00001$, $d = .73$.

This comparison shows that it is possible to use the CB methodology to fulfil the criteria set out by Risen and Chen. Our result indicates that the manipulation strongly influenced subsequent choices and attractiveness ratings. The initially rejected face is chosen more frequently in a second choice, and the perceived attractiveness is increased even in ratings performed when the faces are presented outside the manipulated pairing with another face. The natural interpretation of this result is that, in the M trials, the participants come to prefer the face they were led to believe they liked.

A possible objection to this interpretation is that the change in ratings and choice consistency is not due to the participants' belief in having made the choice but rather a result of the increased presentation time of the initially rejected alternative. When the participants explain why they preferred the chosen alternative in an M trial, they also experience this originally rejected face for a longer period. Prior research has suggested that prolonged exposure may influence the perceived attractiveness of visually presented objects, the so-called mere exposure effect (Zajonc, 1968, 2001). To rule out this possibility, we ran a simple CB condition with a second round of choices but with the rating procedure removed. This was compared with a mere-exposure condition in which either the chosen or rejected picture was shown to the participants for an extended period, but without any manipulation, to see if the extra exposure would have an effect on the second choice.

EXPERIMENT 2: MERE EXPOSURE AS A POSSIBLE MECHANISM FOR THE CHOICE BLINDNESS EFFECT?

Method

Participants

Forty mainly student participants (24 women) took part in experiment 2 (mean age 24.0 years, $SD = 3.6$), each receiving a cinema ticket as payment for participation.

Material

The same 15 pairs of printed images of faces were used as in experiment 1, with the same six pairs used as target pairs.

Procedure

Experiment 2 consisted of two conditions, a simple CB condition ($N = 20$) and a mere-exposure condition ($N = 20$).

In the simple CB condition, the procedure was the same as in experiment 1 regarding presentation time and choice procedure but with the rating procedure of the images removed. For the six target pairs, the participants were either given the chosen face or, through manipulation, the rejected face and were asked to explain their choice. The chosen/rejected image was visible on average 8.9 seconds ($SD = 4.2$) during the motivation.[3]

In the mere-exposure condition, as in previous experiments, the participants were presented with the 15 pairs of faces, each pair shown for 4 seconds and then turned face down on the table until the participants had decided which one they thought the more attractive. For three of the six target pairs, the chosen image was given to the participant to hold and simply look at for 10 seconds; for the other three pairs, the non-chosen image was given to the participants to look at for 10 seconds but without any attempt to conceal the fact that it was the face not preferred (i.e. it was given to the participants directly from the hand that held the non-preferred face). This way, the participants viewed the chosen face and the rejected face for an equal amount of time but without being led to believe that they had chosen the non-preferred face.

In both conditions, after the first round of 15 choices, all the pairs were presented one more time in a randomized order, and the participants were asked to choose the one in each pair they now preferred.

Results

The detection rate for the M trials was similar to that in experiment 1, with 10% of the manipulations detected concurrently and another 16% retrospectively detected after the experiment.

Comparing the level of choice consistency between first and second choices in the CB condition, the participants were consistent in 83.3% of the NM trials and in 61.6% of the M trials. This amounts to a significant difference in level of choice consistency ($\chi^2(1, N = 120) = 7.06$, $p < .01$). In the mere-exposure condition, the participants were also 'exposed' to three of the chosen alternatives and to three of the rejected alternatives an equal amount of time but without being misled in relation to which alternative they preferred in each pair. If mere exposure was the sole mechanism at play in the CB experiment, there should be the same preference effect in this condition—the participants should be

[3]The viewing time was estimated from the length of the recording of the verbal report. In the mere exposure condition, we therefore added a second to an even 10 seconds.

DOI: 10.1002/bdm

significantly less consistent for the pairs in which they had had a prolonged interaction with the non-preferred alternative. However, there is no evidence that this is the case: in trials in which the participants were given the preferred image after the first choice, the choice consistency was 88.3%, whereas for trials in which the participants were given the rejected alternative, the level of choice consistency was 78.3%, a non-significant difference ($\chi^2(1, N = 120) = 2.16, p = .14$).

Comparing the choice consistency for the M trials in the simple CB condition (61.3%) with the trials in which the participants received the opposite of their choice in the mere-exposure condition (78.3%), we do find that the M trials are significantly less consistent ($\chi^2(1, N = 120) = 3.97, p < .05$). Given this result, mere exposure cannot be the mechanism behind the preference change found in the CB experiments reported here.

General discussion

Summarizing the results, we find that our CB methodology fulfills the criteria set out by Risen and Chen and that CB manipulations can strongly influence subsequent choices and attractiveness ratings in a setup similar to the classic FCB procedure. The initially rejected face is chosen more frequently in the repeated choice trials, and the perceived attractiveness of the initially rejected face is increased even in uncoupled individual ratings at the end of the experiment. This demonstrates that the participants come to prefer the face they were led to believe they liked and that the effects of CB are not only visible in snapshot measures (whether these are relatively inconsequential, as the judging of abstract patterns in Johansson et al., 2008, or concern choices of great personal and societal importance, as the voting decisions in Hall et al., 2013) but rather can manifest themselves over multiple choice and rating points, potentially with long-lasting consequences.

An objection to our finding would be that the preference change might be confounded by mere exposure of the alternatives (Zajonc, 1968, 2001), but experiment 2 showed that this is most unlikely to be the case. Another objection would be that the participants we classified as not having detected the manipulations actually did so, but refrained from telling us, then deduced the purpose of the experiment and altered their choices and ratings to please us. However, we find this objection very unlikely. One thing we have consistently found in our CB studies is that there are remarkably few differences between how the participants behave in NM and non-detected M trials, whether this is shown in linguistic behavior, such as emotionality, specificity and certainty in the verbal reports motivating their choices (Johansson et al., 2005, 2006), in expressed confidence in the choices made (Hall et al., 2010) or in lack of differentiation on standard compliance and social desirability scales (Aardema et al., submitted; Merkelbach et al., 2010). In a recent study, participants were given a computerized CB task while their eye movements and pupillary responses were recorded (Pärnamets, Hall, Strandberg, Balkenius, & Johansson, in preparation). First of all, the simple fact that the participants look at and fully attend to the manipulated images after presentation rules out inattention as a possible explanation for CB. In line with previous studies, we also found no differences in viewing patterns after presentation between NM and non-detected M trials, whereas we find distinct differences when compared with detected M trials. In addition, we found a significant increase in pupil dilation in the detected M trials compared with both non-detected M trials and NM trials and, at the same time, no differences between non-detected M trials and NM trials. Pupil dilation has been used as a robust measure of surprise and cognitive load (Porter, Troscianko, & Gilchrist, 2007; Preuschoff, 't Hart, & Einhäuser, 2011), and this new result thus lends strong support to the claim that the participants are not conscious of the manipulation. In the current study, the objection is further undermined by the fact that the preference change effects are considerably smaller for the detected trials. Together, this pattern of results clearly indicates that in trials classified as being non-detected, the participants are truly unaware of the manipulation made.

In line with the methodological improvements suggested by Risen and Chen (2010), three new versions of the FCP have recently been introduced (Coppin, Delplanque, Porcherot, Cayeux, & Sander, 2012; Egan, Bloom, & Santos, 2010; Sharot et al., 2010). The first two of these procedures try to control the impact of prior preferences by letting the participants make a choice between two options without knowing what the options are, a so-called blind choice procedure[4] (corresponding to option 3 in the introductory listing of Chen and Risen's suggested solutions).

For example, in Sharot et al. (2010), the participants first rate a long list of names of holiday destinations and are then asked to make choices between two equally ranked destinations that are claimed to be subliminally presented on the screen (that is, they are unseen, but participants are forced to 'choose' anyway). After the choice, the two masked alternatives are made visible and marked according to what the participants had indicated as their choice. Finally, the participants rate all the alternatives one more time. The original twist here is that only nonsense symbols are shown during the 'subliminal' presentation, so the participants have no information to base their forced choices on. Still, this procedure led to an increase in the rating of the chosen alternative (but no reduction in value for the 'rejected' alternative). This result was further extended in a follow-up study, showing that some of these changes in preference were present up to 3 years after the blind choice was made (Sharot et al., 2012).

Another attempt to show both short- and long-term preference effects of choice while still avoiding the objections by Chen and Risen was introduced by Coppin et al. (2012). In this experiment, the participants had to smell and then rate individual odours for pleasantness. Before the start of each trial, the participants had to make a choice whether to 'pay' for the upcoming trial, without knowing what smell they would receive. The participants were given money to pay for exactly half of the trials. The study found an increased pleasantness rating for odours preceded by the choice to pay,

[4]This description is confusingly similar to 'choice blindness' but is not otherwise related. Our term derives its name from the parent phenomenon of change blindness and was named in Johansson et al. (2005).

DOI: 10.1002/bdm

and reduced for odours with no cost, an effect found both immediately and a week after the experiment. As the choice was made before the presentation of the odours, it was also 'blind' with respect to the participants' underlying preferences.

These studies have gone some way towards re-establishing the notion that choices can influence future preferences. But the blind choice approach can be seen as something of a pyrrhic victory. Holden (2013) argued that by comparing objects participants are nearly indifferent between, and by removing the actual comparison of the objects from the choice, it is too far removed from real-world decision making to be considered a choice at all. It has also been noted that the effect sizes in the blind choice experiment are considerably lower than in the original FCP experiments, again to the point of questioning the real-world relevance of the findings (Izuma & Murayama, 2013). Comparing the average effect size of studies addressing the Chen and Risen criticism ($k = 4$, $M(d) = .26$; from the meta-analysis in Izuma & Murayama, 2013), with, for example, the effect size of the final non-paired rating difference score for the left and right choices in experiment 1 ($d = .73$), we find that the effect is considerably stronger in the current study.

It is also unclear how far the result of Coppin et al. (2012) speaks to the issue at hand. The participants chose whether to pay to smell an odour, and this influenced future preferences. This design avoids the criticism by Chen and Risen by making the choice 'blind', but the effect obtained is best explained in relation to the effort justification rather than the FCP. The participants' choices are not made *between* two options whose preferences are to be altered (or not); it is just a choice to spend some money on this trial or the next (without this in any way being related to what odour the participants will receive). See Izuma and Murayama (2013) for a similar discussion of Coppin et al. (2012).

If we apply the preceding discussion to the CB methodology, we argue that it evades the problems of the blind choice paradigm. In the current experiment, all participants were deeply familiar with the task of evaluating facial attractiveness, and only two simple options were compared before they made their choice. Most important, our primary outcome measure were not just nearly equal ratings somewhat pushed apart from their initially positions, but a choice completely reversed, and a subsequent choice consistency in the non-detected trials as low as 43.8%. The face pairs used were pre-rated and selected to differ in attractiveness (from moderately to strongly different), instead of matched to the point of indifference as in the original FCP, or the blind choice studies, but despite this, we still found a strong preference change effect.

However, as we see it, the comparatively easy part is to satisfy the requirements of Chen and Risen (2010) and to demonstrate effects of choice on future preferences. The difficult part is to evaluate the underlying assumptions about 'true' preferences that fuelled their original effort to re-describe the FCP. Couched in terms of standard decision theory (von Neumann & Morgenstern, 1947), and following the assumptions laid out by Chen and Risen (2010), CB becomes very paradoxical. The high consistency between the first and second choices (and ratings) in NM trials seems to indicate a strong and enduring true preference. However, this notion is contradicted by the dramatic reversal found in the M trials, where participants stated their preference by choosing, and then immediately created maximal potential difference by accepting the opposite alternative.

Some proponents of the concept of true underlying preferences might want to argue that CB reveals that participants never had any preference to begin with (notwithstanding the strong consistency of the NM trials), but this is a barbed bitter bullet to bite, as it renders a great many preference elicitations potentially meaningless (e.g. see Hall, Johansson & Strandberg, 2012). This becomes most apparent when we consider studies where CB has been shown with ecologically valid stimuli, and with decisions involving real-world consequences. Would proponents of true preferences be willing to argue that the 50% detection rate in Hall et al. (2012) demonstrates that half the Swedish population holds no articulated attitudes about the most visible moral issues in the current societal debate? For example, this is a move that would threaten to make the meticulous critique of FCP by Chen and Risen (2010) entirely superfluous, as most likely CB could be demonstrated for the different choice and rating points in the FCP. For example, the original FCP studies compared choices between the seventh and ninth ranked options for aesthetic evaluations of paintings, and considerable levels of CB have already been demonstrated for classical and modern paintings by Masuda, Seiko, and Watanabe (2010). If none of the apparent preferences in FCP count as real (choice-induced or not), then this constitutes a far more severe indictment of the field than the possibility of statistical errors in the way spread is calculated.

At the same time, the gist of the criticism made by Chen and Risen (2010) makes a great deal of sense. Often, decision research is severely impoverished in the elicitation of preferences across time and context (Chater, Johansson, & Hall, 2011; Johansson, Hall, & Chater, 2011). In everyday life, unless forced to, who would rely on a single statement or choice to fully explain the needs and desires of our fellow beings (Dennett, 1991a)? Evolutionarily speaking, it makes sense to focus both on patterns of stability (unsurprisingly, a famished person will nearly always have a preference for food) and on the ability for constant contextual re-evaluations that we have had to display (to decide whether to continue foraging for food, to sleep and conserve energy, to escalate a conflict, to stand down and so on; Davies, Krebs, & West, 2012).[5] But unless one assumes that people have perfect introspective access to their own preferences, a most contentious assumption absent from the list of Chen and Risen (2010), it follows from their own arguments that choices reveal information about preferences not only to us as observing researchers but also to the participants themselves (Carruthers, 2011; Dennett, 1991b; Johansson et al., 2005, 2006; Nisbett & Wilson, 1977).

[5]Decisions might also be made from habits, which might exhibit reward insensitivity after overtraining, or even as a reflexive action, operating outside the realm of instrumental actions (Morsella, 2009), but this would not typically be the case in the type of choices studied by social psychologists.

DOI: 10.1002/bdm

Most previous experiments on choice-induced preference change have interpreted the effect in relation to cognitive dissonance (e.g. Brehm, 1956; Festinger, 1957; Gerard & White, 1983; Shultz & Lepper, 1996), but in the current context, we argue that some form of self-perception theory is the more likely candidate (e.g. Ariely & Norton, 2008; Bem, 1967; Chater & Vlaev, 2011; Dennett, 1987). It is notoriously hard to differentiate between cognitive dissonance and self-perception models as they make almost identical predictions (Bem, 1967; Harmon-Jones & Mills, 1999). The main reason we favour self-perception theory is the previously mentioned results on CB and pupil dilation, as this indicates that the participants do not experience any 'dissonant' emotions of cognitions that would drive the change in preference. Even if further experiments are needed to firmly decide this issue, it is clear that the CB methodology is well suited as a tool to differentiate between these two models.

Similarly, in the current study, it is not possible to measure the relative impact of believing that a choice has been made, or believing that a choice has been made *and* having explained the choice. Asking why the choice was made is a natural way to make sure that the participants interact with the 'chosen' item after manipulation, but post-choice attention to the manipulated stimuli could in future studies be measured independently by, for example, eye tracking. But regardless of the exact nature of the underlying mechanism, the combined result of CB and blind choice indicates that it is not the choice per se that drives the preference change but rather the belief that a certain choice has been made. We can thus begin to separate the act of choosing from the belief in having made a certain choice.

Given the small body of work that has been carried out on CB, it is difficult to estimate its pervasiveness. If we assume some underlying preferences, and a certain amount of noise in the process of rating and choosing, but still think that choices may influence preference, we will find ourselves with naturally occurring CB-like situations, where participants get what they choose but not what they 'want'. If someone actually prefers A over B, but happens to choose B through random variation, they might then shift their preference towards B through self-inference. We also have regular CB situations, where people for some reason do *not* get what they choose, but fail to notice it, and then set themselves up for potential preferential loops, as in our experiments. Logically, we cannot know anything about the frequency of mismatches we do not notice (i.e. we might have unknowingly been short-changed thousands of times in our lives or walked for months in oddly paired socks), but taken together, these two alternatives highlight how every preference we measure might change as an outcome of us measuring it.

ACKNOWLEDGEMENTS

P. J. would like to thank the Bank of Sweden Tercentenary Foundation, and L. H. would like to thank the Swedish Research Council for funding this research. N.C. was supported by ERC grant 295917-RATIONALITY, the ESRC Network for Integrated Behavioural Science, the Leverhulme Trust and the Templeton Foundation.

REFERENCES

Aardema, F., Johansson, P., Hall, L., Paradisis, S-M. & Roberts, S. (submitted). *Choice Blindness and Confabulatory Introspection in Obsessive-Compulsive Disorder: A New Area of Clinical Investigation*. First author affiliation: University of Montreal, Fernand Seguin Research Center.

Ariely, D., & Norton, M. I. (2008). How actions create—not just reveal—preferences. *Trends in Cognitive Sciences*, *12*(1), 13–16.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal Psychology*, *52*(3), 384–389.

Carruthers, P. (2011). *The opacity of mind*. Oxford: Oxford University Press.

Chater, N., & Vlaev, I. (2011). The instability of value. In M. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making: Attention and performance XXIII* (pp. 81–100). Oxford: Oxford University Press.

Chater, N., Johansson, P., & Hall, L. (2011). The non-existence of risk attitude. *Frontiers in Psychology*, *2*, 1–3. DOI: 10.3389/fpsyg.2011.00303

Chen, M. K. (2008). Rationalization and cognitive dissonance: Do choices affect or reflect preferences? *Working paper*. New Haven, CT: Yale University.

Chen, M. K., & Risen, J. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, *99*(4), 573–594. DOI: 10.1037/a0020217

Coppin, G., Delplanque, S., Cayeux, I., Porcherot, C., & Sander, D. (2010). I'm no longer torn after choice: How explicit choices implicitly shape preferences of odors. *Psychological Science*, *21*(4), 489–493.

Coppin, G., Delplanque, S., Porcherot, C., Cayeux, I., & Sander, D. (2012). When flexibility is stable: Implicit long-term shaping of olfactory preferences. *PLoS ONE*, *7*(6), 1–13. e37857.

Coppin, G., Delplanque, S., & Sander, D. (submitted). The impact of decision-making processes on preferences in the free-choice paradigm: Classical research, recent methodological issues, and emergent conceptual insights. First author affiliation: *Department of Psychology*, *University of Geneva*.

Davies, N., Krebs, J., & West, S. (2012). *An introduction to behavioural ecology*. Chichester: John Wiley & Sons.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press. DOI: 10.2307/2026682

Dennett, D. C. (1991a). *Consciousness explained*. Boston: Little, Brown & Company.

Dennett, D. C. (1991b). Real patterns. *Journal of Philosophy*, *88*(1), 27–51.

Egan, L., Bloom, P., & Santos, L. (2010). Choice-induced preferences in the absence of choice: Evidence from a blind two-choice paradigm with young children and Capuchin monkeys. *Journal of Experimental Social Psychology*, *46*(1), 204–207.

Egan, L., Santos, L., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science*, *18*(11), 978–983.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Gerard, H. B., & White, G. L. (1983). Post-decisional reevaluation of choice alternatives. *Personality and Social Psychology Bulletin*, *9*(3), 365–369. DOI: 10.1177/0146167283093006

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE*, *7*(9), 1–8. e45457.

Hall, L., Johansson, P., Tärning, B., Sikström, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*(1), 54–61. DOI: 10.1016/j.cognition.2010.06.010

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., & Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes