When do we know our own choices?

Investigation of false feedback acceptance phenomena.

Mariya Kirichek

PhD thesis

Behavioural Science Group

Warwick Business School

May 2017

Author’s Note

Mariya Kirichek is supported by the European Research Council grant 295917.

Contents

i. Acknowledgements

First and foremost, I would like to thank my supervisor Nick Chater. Your vision of human behaviour has been an inspiration, and this thesis is just a fraction of what I learned from you over the past few years. Whilst my own confidence wavered, you always reminded me that completing my PhD could become a reality, and not just a dream. I would also like to thank my second supervisor Graham Loomes, for stepping in at short notice, and providing support at the critical idea generation stage of my research.

I am very grateful to Dr Petko Kusev and Alex Cooke at Kingston University, and Dr Paul van Schaik at Teesside University. The time you have dedicated to helping me develop ideas and interpret results has not only been crucial in producing the work presented here, but also in making the process enjoyable. Whilst working away from Warwick, Kingston became somewhat of an academic refuge – I cannot thank you enough for always making me feel welcome.

My grateful thanks also go to friends and colleagues within Warwick Business School. Linda Donovan and Angela Gibson, without you I cannot imagine ever managing to navigate the organisational maze, which seems to be an inevitable part of doing a PhD. I also owe a lot of my remaining sanity to my fellow PhD students, especially Jess Whittlestone and Avri Bilovich. You reminded me that I am not alone, and helped me find my feet when the ground beneath me seemed to be slipping away.

To my friends that have stood by me since long before the idea of doing a Doctorate even entered my mind, I am sorry for disappearing for weeks on end, and being somewhat gloomy at times. Especially Fran Manning, Lidia Bartoszewicz, Nick Berggren, and Andrew Payne – thank you for not giving up on me, and being there through every celebration and commiseration.

I owe my deepest gratitude to my family. Oksana, mum, thank you for listening to my endless complaining, making me take a break when needed, and for reading all my ramblings regardless of the time of night at which they appeared in your inbox. Oleg, dad, thank you for all the scientific discussions, and for believing that my research is the best in its field – as a result, I was able to take one step closer to becoming a real researcher.

Last but not least, my partner, David Johnson. You were there to witness it all, from the ecstatic look on my face with every new idea to the complete meltdown with every non-significant result. It couldn’t have been easy, but we made it. Thank you for everything, I could not have done it without you.

ii. Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree.

The work presented (including data generation and analysis) was carried out by the author, except where acknowledged by citation to an existing publication mentioned in the list of references.

iii. Abstract

It is widely assumed that people have direct access to knowledge about themselves, such as their preferences or propensities to behaviour. However, research has shown that people often accept bogus feedback about personal characteristics and decisions, even when it completely contradicts information they stated previously. This thesis investigates two examples of such feedback acceptance, the Barnum effect (Meehl, 1956) and choice blindness (Johansson, Hall, Sikström & Olsson, 2005), to explore the conditions required for false feedback to create a false perception of one's choices.

The Barnum effect refers to the tendency to accept false feedback about one's personal characteristics when it is thoughts to be derived from personality measures. The first paper explores whether undergoing the Barnum effect can influence people's perception of choices they would make in the future. We find that whilst the Barnum effect does occur, this does not alter people's self-reported propensity to behaviour.

Choice blindness demonstrates that, following a choice, people are often willing to accept the non-selected alternative as the indicated outcome, if this is suggested by feedback. The remaining work presented here investigates which parameters of choice and feedback can determine how likely it is that people will experience choice blindness. The results suggest that people's susceptibility to choice blindness varies with the number of alternatives presented in the choice task, the framing (positive or negative) of the task itself, and whether the option presented as false feedback was encountered as part of the choice, in a different context, or not encountered at all.

I conclude that the effects of false feedback on self-perception are situation dependent, however, difficult to eliminate completely, at least in some domains. The implications are discussed in light of past literature on false feedback effects, as well as related fields such as preferential choice, introspection, error detection and behavioural change.

iv. List of Original Papers

Paper I

Kirichek, M., & Chater, N., (in preparation). The Barnum effect and its consequences: can bogus feedback change behaviour.

Paper II

Kirichek, M., Cooke, A., Van Schaik, P., & Kusev, P. (submitted). Ternary choice blindness: increasing the number of choice alternatives enhances the detection of mismatch between intention and outcome.

Previously presented as a poster at Second International Meeting of Psychonomic Society, Granada, Spain.

Paper III

Kirichek, M., & Chater, N. (in preparation). Choice blindness for preferred versus non-preferred stimuli.

Paper IV

Kirichek, M., & Chater, N. (in preparation). Choice blindness for stimuli external to the choice.

Chapter 1.

Introduction

Consider a world where instead of choosing what you prefer, someone else informs you of your preferences. For example, you sign up to a dating website where you are allowed to examine the photographs of all the possible people you could choose to meet. You carefully consider the possible options and decide on one finalist. You then receive a confirmation of the person you selected to meet. However, the dating company has decided that it is better equipped to decide what you would prefer and send you a photograph of a person you did not choose. In this fictitious world, you cannot keep track your own choices, and do not notice that your original choice has been replaced with a different person entirely, simply accepting the information provided as your selected preference, and happily prepare for the date with your dream partner.

Most people would consider such a world very different from the one we live in: the inability to monitor our own choices and preferences is hard to imagine for a normal functioning human being. Although the scenario described above is fictitious, the phenomenon of choice blindness suggests that if placed in that situation you may react in exactly the way described. Choice blindness was first reported in 2005, when Johansson and colleagues (Johansson et al., 2005) found that when presented with the task of justifying a previously preferred alternative from two female faces, people fail to notice if the choice presented is not the one actually selected. Furthermore, the effect occurred even if the face presented was dissimilar to the one selected, and the justifications provided were found to refer to features only present in this non-chosen face.

The current thesis stemmed from my own initial doubts that we are inherently bad at knowing what we might have, or in fact have, preferred in the very recent past, leading to a search of an extraneous variable that might explain or eradicate the phenomenon of choice blindness. My approach to the research, however, evolved to treating choice blindness as a continuous measure of choice stability, to try and understand what factors can minimise the effect. Consider the earlier example of a dating agency sending you the photograph of the person you did not prefer. What would have happened if they sent you a photograph you did prefer, but asked you to explain why you rejected the person depicted? Or alternatively, if they simply sent you a photograph of someone you have never seen before, indicating that this was in-fact your selection? These are the kind of questions I try to address in my research, through measuring participants' ability to detect the erroneous nature of the outcome of their behaviour, or choice blindness.

I also extend my research of choice stability to a closely related but distinct phenomenon, the Barnum effect, testing whether altered personality feedback can influence subsequent preferential choice. Whilst the study of the Barnum effect may appear to be somewhat out of place, the range in research reflects the selection of the studies presented here from a wider body of experiments conducted over the course of the last three and a half years. Other research topics considered but not presented here include the effects of false feedback about a learnt categorisation task on subsequent reaction times exhibited if the task is repeated; effects of repeated choice blindness manipulation on risk preferences; and how the co-occurrence of seemingly irrelevant features in repeated choice can create subsequent biases for alternatives exhibiting such features. The four papers presented here were chosen partially because they address one specific domain, forming a coherent thesis, and because they were deemed to be of sufficient power and rigour to meet publishable standards.

The research presented in this thesis consists of four papers, with papers II-IV at its core dedicated to understanding which variables can affect choice blindness and thus choice monitoring. The papers are summarised in chronological order in the next section. Paper I explores the Barnum and its impact on proceeding decisions. Paper II examines how increasing the number of choice alternatives, from two to three, affects the proportion of people who detect a switch of their chosen alternative, as well as establishing how similarity and attractiveness of the presented alternatives may impact switch detection for ternary and binary choice. Paper III examines how positive versus negative question framing of the choice and justification tasks may impact the proportion of people who detect a switch of their choice. Paper IV examines whether choice blindness is confined to the alternatives presented in the choice or whether we can be “fooled” into believing that we have chosen an alternative that was not even available to us.

Although each study addresses a distinct question, the four papers largely share their domain and methodological approach, examining the circumstances under which people accept false feedback about their personality (Paper I) or past choices (Paper II-IV).

1.1 Summary of Papers

This thesis consists of four papers on the general topic of people's (in) ability to recognise false feedback, summarised here to support the readers' understanding of the introductory chapters. Paper I examines the Barnum effect, whilst papers II-IV focus on the choice blindness phenomenon. The papers are presented in chronological order, as in some instances later studies build on earlier work. More specifically papers III and IV build on stimuli established in paper II. Links with other work conducted will be noted where relevant.

1.1.1 Paper I – The Barnum effect and its consequences: can bogus feedback change behaviour.

The 'Barnum effect' refers to individuals' tendency to rate bogus descriptions of their own personality as highly accurate, because these descriptions have supposedly been tailored specifically to them (Meehl, 1956). Past research has shown that experiencing the Barnum effect can impact future behaviours. More specifically, Halperin and Snyder (1979) have shown that when participants accept feedback that suggests higher potential to change, they are more responsive to phobia therapy. Sakamoto, Miura, Sakamoto and Mori (2000) further show that feedback suggesting higher levels of extraversion results in participants being more interactive with strangers. Unfortunately, the research demonstrating the effect has failed to use appropriate control groups, or a gender-balanced sample, and utilises seldom-used individual difference measures. Paper I uses well-established personality and risk attitude measures to induce the Barnum effect, and investigate any subsequent impact this may have on self-reported propensity to volunteer for psychology experiments and make risky choices respectively. The research is comprised of two experiments. Both experiments examine participant behaviour following real, as well as altered feedback, in order to provide an appropriate baseline for accurate interpretation of the results. In addition, the consequences of the Barnum effect are examined using a mixed gender sample for the first time.

Experiment 1. Participants completed a Big Five personality assessment (Goldberg et al., 2006) followed by receiving personalised feedback on their performance. Whereas some participants received real feedback, for others the scores were altered in a direction pre-determined by a randomly allocated condition. Feedback was altered to suggest either traits associated with low likelihood to volunteer for a psychology experiment, or high likelihood to volunteer for a psychology experiment (the associated traits were established in a pilot study). The participants were asked to rate the accuracy of the presented feedback, to establish whether the Barnum effect had occurred. Lastly, the participants answered four questions on volunteering behaviour: rating their likelihood to volunteer for online, phone, face-to-face, or group, psychology experiments.

Experiment 2. Participants underwent a risk attitude questionnaire (Blais & Weber, 2006), and were provided with feedback on their ethical, financial, recreational and social risk attitudes. High scores indicated risk seeking attitudes, whereas low scores indicated risk averse attitudes in the specified domain. For some participants, their scores on the financial risk attitudes were either enhanced to suggest higher risk preference, or lowered to suggest risk aversion. Participants were asked to rate the accuracy of the feedback presented, and lastly completed ten preferential choice decisions between risky and certain lotteries.

Both experiments successfully induced the Barnum effect, to the extent that participants rated altered feedback as more accurate than chance. The effect was, however, asymmetrical for both experiments. In experiment one, people rated personality feedback as less accurate when it was altered in the direction associated with low likelihood to volunteer for experiments, compared to when it was altered in the direction associated with high likelihood to volunteer, or not altered at all. For experiment two, participants rated the feedback as less accurate when the risk feedback was altered to suggest increased financial risk preference, compared to decreased financial risk preference or real feedback. We propose that this may be the result of volunteering behaviour and low risk preference being regarded as socially desirable, resulting in feedback suggesting a propensity to such behaviours more likely to be accepted by the participants due to its positive nature (Macdonald & Standing, 2002). Regardless of the manner in which the feedback was altered, both experiments failed to induce any significant changes in the subsequent self-reported propensity to behaviour. Any differences in the responses are so small, that the data appears to be much more likely to be obtained under the null hypothesis compared to the hypothesis that believing false feedback can alter subsequent responses. It seems that people do tend to "accept" false feedback, but do not alter their understanding of themselves substantially enough to affect subsequent predictions about their behaviours.

Whilst experiencing the Barnum effect does not appear to alter people's choices, the related phenomenon of choice blindness shows that using false feedback about the choice itself can directly alter the perception of one's choices. Choice blindness refers to people's inability to detect the switch of a stimulus they selected as their preference in a choice task, with the non-selected alternative, during feedback (Johansson et al., 2005). Given the similarity between the two phenomena (both use false feedback to impact self-perception), a partial aim of paper I was to explore potential parallels between the Barnum effect and choice blindness, as the substantial literature base on the Barnum effect could have proven very useful for understanding choice blindness if the two processes were to be considered the same. Specifically, it was anticipated that participants could undergo the same process when faced with false information in both paradigms, thinking 'I must be the kind of person who would choose X', and adjusting choice perception, and subsequent choices accordingly. Whilst this process may occur for choice blindness (Johansson et al., 2005; Johansson, Hall, Tärning, Sikström & Chater, 2014), paper I suggests this is not the case for the Barnum effect, although the reason for this discrepancy remains unclear. To gain better understanding of the nature of false feedback effects the subsequent papers focus on choice blindness directly, drawing hypotheses from the broader literature on preferential choice, building on our knowledge of variables that impact choice stability (or consistency). One such variable is the size of the choice set from which the participants make their selection (e.g., Collins & Vossler, 2009) which is investigated in paper II.

1.1.2 Paper II – Ternary choice blindness: increasing the number of choice alternatives enhances the detection of mismatch between intention and outcome.

Choice blindness, or the failure to detect a mismatch between the intended outcome of a choice task and an erroneous outcome presented as feedback (Johansson et al., 2005), has become a well-established phenomenon. It has been demonstrated across various decision types and domains (e.g., personal finance – McLaughlin & Somerville, 2013; haptic stimuli – Steenfeldt-Kristensen & Thornton, 2013; eyewitness testimony – Sagana, Sauerland & Merckelbach, 2013), however, has been largely limited to binary choice. Past research has shown that increasing the number of choice alternatives can play an important role in preferential choice (e.g., Collins & Vossler, 2009). Accordingly, we explored how increasing the number of decision alternatives to three options affects the proportion of people who detect a switch between intended outcome and the one presented.

In line with the majority of choice blindness research, we used female faces to create the binary and ternary sets of stimuli to be used in the choice blindness paradigm. Since similarity has been shown to impact choice blindness in past research (e.g., Sagana et al., 2013; Steenfeldt-Kristensen & Thornton, 2013), the physical similarity and the similarity of perceived attractiveness of the sets was controlled for, using similarity and attractiveness ratings established in an earlier pilot experiment. For binary choice, the pairs were either physically similar or dissimilar, and of similar or different attractiveness to each other. For ternary choice, two of the alternatives were always similar physically and on attractiveness, whereas the third choice was either physically similar or dissimilar, and of similar or dissimilar attractiveness level to the other two faces. Additionally, the perceived relative dissimilarity of a face was expected to be enhanced when presented alongside two alternatives, compared to one (Ariely, 2008), as a result of grouping (Tversky, 1977) and salience (Taylor & Fiske, 1978) effects, suggesting an interaction between the number of alternatives and similarity should be anticipated.

Participants underwent a computerised, one shot decision variation of the choice blindness task online. The task consisted of making a preferential choice between two or three female faces, providing a confidence rating for the selection, and lastly explaining why a (non-chosen) face was preferred. The number of options, similarity and relative attractiveness of the facial stimuli were randomly allocated for each participant. Choice blindness was assessed using the justification provided, as well as by asking whether the participant noticed anything unusual post-task.

Detection of the switch in outcome was higher for ternary compared to binary choice, but only when the face presented in the switch was visually dissimilar and of lesser attractiveness than the chosen alternative. The trend is in line with previous literature that suggests ternary choice leads to decisions that are more stable than binary choice. It is likely however, that in the current experiment the observed effect is a result of enhancing perceived differences between stimuli through altered salience, as opposed to a more generic mechanism previously proposed, such as the increase in the number of alternatives corresponding to the increase in likelihood that a suitable choice will be found (Caussade, de Dios Ortúzar, Rizzi & Hensher, 2005).

Given the emerging relationship of choice blindness with the broader choice literature, the next experiment continued to investigate how factors known to impact choice stability interact with choice blindness. Specifically, paper III focuses on framing. Past research has demonstrated that when a task is framed negatively (asks participants to select the worst vs. best alternatives) participants exhibit higher consistency between choices and previously stated preferences (Kogut, 2011). Accordingly, we tested whether negative framing would also increase the proportion of people that detect a mismatch between intended and presented outcomes, as would be expected for choices with better stability.

1.1.3 Paper III – Choice blindness for preferred versus non-preferred stimuli.

Procedural invariance, or the notion that the way in which the same question is asked should not affect the answer, is one of the fundamental assumptions of rational choice theory, yet past research has demonstrated that choosing versus rejecting alternatives can impact consistency of choices with previously stated priorities (e.g., Shafir, 1993; Kogut, 2011). In turn, the current paper examined how framing can impact the choice blindness paradigm. The classic choice blindness paradigm consists of asking participants to select their preferred option from two presented alternatives (the choice task), followed by displaying the option inconsistent with their choice and requesting a justification for the decision made (justification task). Not noticing the mismatch between the chosen and presented options is then classified as choice blindness. The current experiment manipulated the framing of the choice task, as well as the subsequent justification task in terms of participants' preferred or the least preferred alternative. This formed four distinct variations of the task: the preferred or least preferred framing of the choice task, combined with either preferred or least preferred framing of the justification task. In line with past research, faces were used as stimuli. It was hypothesised that the participants who were required to select their least preferred face and explain why they did not prefer the (preferred) face would be most likely to detect a switch of their intended choice, as expected if negative framing is to lead to higher choice stability.

Participants were randomly allocated to one of the four task variations. All participants chose either their most or least preferred alternative out of two faces and rated their confidence in the decision made. They were then presented with a face and asked to explain why they preferred or did not prefer the presented image. The face shown during this justification task, was always incongruent with the justification instructions. For example, if asked to explain why they preferred the presented face, they were presented with the non-preferred alternative. At the end of the experiment, participants were asked to describe anything unusual they noticed during the task. This response was used to assess whether the participant had experienced choice blindness, alongside the analysis of justifications provided within the task.

The findings suggest that participants are more likely to detect a mismatch when both the choice and justification aspects of the choice blindness paradigm are framed positively. In other words, when participants are asked to select their preferred alternative and explain why they chose the presented (non-preferred) alternative, as in the original choice blindness paradigm, participants are more likely to detect that their initial response does not match the outcome presented. Participants who encountered the negative framing of either the choice or justification task, or both, did not differ in their likelihood of detection.

The finding that positive framing leads to increased detection of a switched choice is somewhat surprising, as it suggests positively framed tasks lead to more stable preferences: the opposite to the pattern hypothesised based on past research by Kogut (2011). The paper puts forward a number of explanations for this discrepancy, from the possibility that mechanisms underlying choice blindness and choice consistency are, in fact, distinct, to procedural difference that may have impacted the effect such as the number of alternatives, and task format.

The findings presented in paper III, suggest that choice blindness might be more distinct from other choice consistency measures than initially anticipated. This led me to consider properties of the choice blindness paradigm, which other choice consistency measures could not capture due to procedural limitations. One question that we can address with choice blindness, but not choice consistency paradigms, is whether people are able to detect a mismatch between their chosen alternative, and an alternative that was not even a part of the task, or an ‘imposter’ choice. This is the question considered in paper IV of this thesis. By investigating whether people’s acceptance of an imposter choice as their stated preference is dependent on prior exposure to, or the level of interaction with, the imposter stimulus, we can enhance our understanding when and why choice blindness occurs.

1.1.4 Paper IV – Choice blindness for stimuli external to choice.

Whether choice blindness, or the failure to detect a mismatch between an intended choice and its outcome, is contingent on the stimulus presented being a part of the earlier undertaken task is unclear. In the current paper, we explored whether people accept an imposter choice regardless of having encountered it before; because they had seen it before; because they have evaluated it before; or because it was encountered as a part of comparative choice, even if that choice was separate from the one they are receiving feedback for.

All participants made a choice between two female faces, and following a distractor task were presented with a face and asked to justify why they preferred that face. However, the face was always replaced with the one they did not choose, or in fact, an imposter that was neither of the faces presented in the original choice task. Depending on condition, the image presented was either a face they had never seen before; a face they were shown prior to the choice task; a face they were earlier asked to evaluate; or a face they rejected in a separate choice task, completed prior to the task they are receiving the feedback for. Choice blindness was measured by assessing the justification for responses provided (concurrent detection), as well as by asking participants whether they detected anything unusual after the experiment (retrospective detection).

The results revealed that participants that were asked to explain why they selected an imposter face that they encountered in an earlier choice task exhibited a significantly lower level of detection, compared to participants in the other conditions. For conditions where the participants were asked to justify an imposter face that they had never seen before, or had encountered before but not as a part of a choice (just looked at the face in an array of options, or provided evaluative comments for the face as part of a distinct task), a very low proportion of participants failed to detect that their choice had been switched, with less than fifteen percent experiencing either concurrent or retrospective detection. The findings demonstrate that deliberating an alternative as a part of a choice process plays a crucial role in determining whether choice blindness will occur. This can be interpreted in two possible ways: first, that the similarity of the contexts in which the actual chosen alternative and imposter alternative are encountered is crucial to choice blindness, or second, that the actual process of making a choice is crucial to inducing choice blindness. Since, a variation of choice blindness can be achieved using a judgement task (e.g., Hall, Johansson & Strandberg, 2012), we conclude that the former explanation is more likely.

The papers presented here demonstrate that choice blindness is likely to be guided by different processes to the Barnum effect (paper I), whilst being closely intertwined with the cognitive elements involved in making a decision (paper IV). I further conclude, that it is possible to manipulate the likelihood with which participants experience choice blindness (paper II & III), however, reducing choice blindness remains a challenge since increasing the number of alternatives only improves detection with very specific alternative properties (paper II), and the original choice blindness paradigm appears to already be framed in a manner that maximises the possibility of detection (paper III).

Having presented the main findings of my work in this summary, over the next two chapters I will review the relevant literature pertaining to false feedback effects, specifically the Barnum effect, choice blindness, as well as broader areas of choice and self-perception. Chapter 2 will introduce false feedback effects, and their role within different areas of research, whereas Chapter 3 will outline factors that can impact false feedback effects and the possible explanations of the processes that give rise to such effects. The full papers will then be presented in Chapters 4-7, ending with a conclusion and summary in Chapter 8.

Literature Review

Chapter 2.

False Feedback Acceptance: Why it matters.

2.1 What are False Feedback Acceptance Phenomena?

Feedback is fundamental to the behaviour of all living creatures. We rely on feedback on the biological level, to let us know if it gets too cold, or if we get thirsty; and on the behavioural level, for reward or punishment, to determine whether a behaviour should be repeated (Skinner, 1938; Ayllon & Azrin, 1968). Similarly, feedback features in most forms of higher cognition such as learning and general monitoring. For example, we use feedback to acquire speech by comparing the auditory feedback of the words we produce with those we are trying to imitate (Perkell et al., 1997); or to keep safe, by constantly processing environmental feedback about unexpected changes, such as shifts in motion or colour (Abrams & Christ, 2003; Boot, Brockmole & Simons, 2005). In a manner of speaking, all of our senses provide us with a form of feedback about our interaction with the environment. But, what happens when a piece of feedback we receive is incorrect, contradicting other sources of information such as our memory, knowledge of ourselves, or even the laws of physics? Would we be able to identify such information as false and ignore its implications, or would we accept it, updating our beliefs accordingly? To investigate this, psychologists have established a range of experimental approaches to investigate when we tend to accept false feedback and the implications of doing so.

The study of false feedback acceptance can be broadly split into two categories, according to the type of feedback used: feedback about psycho-physical processes and feedback about the psychological aspects of the self, such as attitudes, memories and behavioural propensities. This thesis focuses on the latter, specifically exploring two types of paradigm: the Barnum effect (Meehl, 1956) and the choice blindness phenomenon (Johansson et al., 2005). Before proceeding with the detailed discussion of these effects, it is important to mention that as far as psycho-physical feedback is concerned, there seems to be a general consensus that altered feedback is widely accepted and is capable of changing our perception, and responses, accordingly. The most illustrative example of this is our ability to adapt to altered visual feedback, in the form of inverted perception (with the help of specially designed glasses). Research has shown that if everything we see is turned upside-down for a consecutive length of time, it takes as little as ten days for our vision to adjust completely, and perceive our surroundings as they would be without any inverting apparatus (e.g., Stratton, 1896; Erismann & Kohler, 1953). The research on false psycho-physical feedback has been widespread, ranging from examining how false feedback of our own actions can impact motor control (e.g., Fourneret & Jeannerod, 1998), to how altered auditory feedback of our heart-rate can affect perceived attractiveness (Valins, 1966), and has been advancing our understanding of human behaviour for over a century.

The examples described above focus on augmenting environmental, or physical feedback that can be measured in an objective manner, for instance, the degree to which visual input is rotated, which can be established with great precision. The nature of feedback used in the early research on false feedback acceptance with regard to one's psychological predispositions was very different, relying on vague and generalizable characteristics as the underlying properties that lead to acceptance of bogus, supposedly personalised, descriptions of the self (e.g., Forer, 1949; Sundberg, 1955).

The tendency to accept such inaccurate feedback about one's own personality was first reported by Forer in 1949. Forer administered a personality test to 39 of his students and pretended to score their tests to derive personalised feedback for each person. However, instead of scoring the tests, he gave the same feedback, copied from an astrology column of a newspaper, to all the students. The participants then rated how accurately they thought the feedback described them, resulting in an average score of 4.3 out of 5, suggesting high perceived accuracy of the statements provided. The findings led Forer to conclude that when provided with general feedback that could apply to anyone, people tend to ignore its vague and widely applicable nature, and rate it as a highly accurate description of themselves. Forer described this phenomenon as the ‘fallacy of personal validation’, whilst literature initially referred to it as the ‘Forer effect’. The term ‘Barnum effect’ was later popularised by Meehl (1956), so named after P.T. Barnum – a circus entertainer with the catch phrase ‘we have something for everybody’.

Since Forer’s (1949) experiment, the Barnum effect has continued to be successfully replicated. It has been demonstrated for feedback supposedly derived from a wide range of personality assessment tools, including astrology (e.g., Fichten & Sunerton, 1983; Glick, Gottesman & Jolton, 1989; Rosen, 1975), clinicians’ descriptions (e.g., Halperin, Snyder, Shenkel & Houston, 1976; Rosen, 1975; Snyder & Larson, 1972), and trait measures of personality (e.g., Furnham, 1989; Guastello & Rieke, 1990; Wyman & Vyse, 2008), for both person and computer generated assessments (e.g., Baillargeon & Danis, 1984; Guastello & Rieke, 1990; O’Dell, 1972; Snyder & Larson, 1972). Acceptance of bogus feedback also appeared to be influenced by the characteristics of the person assessing the accuracy of feedback (e.g., Furnham, 1989; Sundberg, 1955; Weinman, 1982), as well as the nature of the feedback itself (e.g., Johnson, Cain, Falke, Hayman & Perillo, 1985; Macdonald & Standing, 2002; for review see Dickson & Kelly, 1985; Furnham & Schofield, 1987; Snyder, Shenkel & Lowery, 1977). The risk of the Barnum effect was especially high if the generated feedback included vague (e.g., 'you enjoy a certain amount of change and variety in life'), double-headed (e.g., 'you are generally cheerful and optimistic but get depressed at times'), or favourable statements (e.g., 'you are forceful and well-liked by others'), or described common characteristics of the subject's group (e.g., 'you find that study is not always easy', Sundberg, 1955; Dickson & Kelly, 1985).

The findings raised widespread concerns across the field of psychology, as they questioned the validity of clinical and individual difference measures that were validated using subjective accuracy ratings (see Dickson & Kelly, 1985; Sundberg, 1955; Poškus, 2014 for discussion).

However, the interest in the Barnum effect has largely subsided over the last thirty years (Poškus, 2014), perhaps as a result of real personality measures becoming more established, and demonstrated to be less susceptible to the Barnum effect compared to the traditional vague and general 'Barnum statements' (Greene, Harris & Macon, 1979; Wyman & Vyse, 2008). For example, Wyman and Vyse (2008) found that when using feedback derived from the five-factor personality model (Costa & McCrae, 1985; Goldberg, 1990), arguably the most widely accepted approach to personality assessment, participants were able to identify real feedback with better accuracy than chance, if presented with real and false personality profiles side by side. When presented with astrology profiles, however, their accuracy rate was approximately fifty percent. Although the concern regarding common use of general statements in accepted instruments may have become less consequential, research slowly began to turn its attention to the occurrence of the Barnum effect in more specific feedback types, to understand whether false feedback would be accepted even if it did not follow the generic characteristics associated with the Barnum effect.

Research using trait personality measures showed promising results, suggesting that specific feedback is less prone to the Barnum effect (e.g., Andersen & Nordvik, 2002; Wyman & Vyse, 2008). For example, Andersen and Nordvik (2002) demonstrated that the accuracy ratings of altered personality profiles decreased, as the difference between their real profile, and that presented to them, increased. Furthermore, studies have consistently showed that negative feedback was rated less accurately compared to real feedback (Johnson et al., 1985; Macdonald & Standing, 2002; for exception see Dmitruk, Collins & Clinger, 1973), supporting the notion that some types of feedback are less susceptible to the Barnum effect than others. Whilst the format used in real personality feedback appeared to outperform traditional generic feedback associated with the Barnum effect, examination of the actual ratings revealed that false feedback is still consistently rated as more accurate than the mid-point rating of 'neither accurate, nor inaccurate' (Poškus, 2014; Wyman & Vyse, 2008), even when the feedback presented was a full inversion of the real personality profiles. This indicates that whilst the perceived accuracy of false feedback may reduce for specific and negative feedback types, to some extent participants are still prone to the Barnum effect.

As evident from the research examples discussed, the definition of the Barnum effect has evolved over time. What started out as a term referring to acceptance of feedback that is vague and general enough to apply to anyone (Forer, 1949), has been expanded to refer to the tendency to accept any personality feedback as true despite its validity (Poškus, 2014). With the updated definition in mind, it is apparent that the Barnum effect is still relevant to individual difference research today, as well as a range of related fields. Whilst some progress has been made in identifying when and why the Barnum effect occurs, and the potential subsequent effects it can have on beliefs about the self and behaviours (see Chapter 3), there is still a lot to discover about the cognitive mechanisms involved in bringing it about and how beliefs about the self may be updated as a result.

In 1974, Loftus and Palmer reported that people's tendency to accept false information extends beyond abstract, hard to define, personal characteristics to actual memories of encountered information. In this classic paper two experiments were carried out to investigate how information embedded into questions can alter participants' reports of car crash scenes previously shown to them in a movie. In experiment one, participants saw seven films of traffic accidents, following which they were asked to provide their account of the events and answer a number of questions. One of the questions was "About how fast were the cars going when they hit each other?", however depending on the experimental group an equal number of participants saw words *smashed*, *collided*, *bumped* and *contacted* in place of *hit.* Results demonstrated that the reported speed did indeed vary with the different verbs (40.8 mph for smashed, 39.3 for collided, 38.1 for bumped, 34.0 for contacted and 31.8 for hit). In experiment two, participants repeated the procedure however only with the hit and smashed question variations. A week later the subject came back and answered another set of questions, one of which was "Did you see any broken glass?". Fourty-seven percent of participants who were initially exposed to the smash variation of the question answered that there was broken glass, compared to 16% who saw the ‘hit’ variation and 14% that did not encounter the question at all. The researchers concluded that implicit information contained in memory assessment questions can influence people’s memories.

The ability to influence human memory through external information has become known as the misinformation effect capturing anything from minor memory alterations from contextual cues (e.g., Loftus, 1977), to creation of false memories (e.g., Roediger & McDermott, 1995) in over 200 experiments (see Loftus, 1997). It is thought that deterioration of memory over time (e.g., Loftus, 2005), as well as aroused emotional state (e.g., Van Dame and & Smets, 2014) play an important role in determining malleability of the memory in questions, by weakening the memory and decreasing the likelihood that the information is noticed. Whilst information contained in questions is not a direct example of false feedback, it can be described as such as it provides a source of external information that can in theory be compared to the internal beliefs held.

Johansson and colleagues (2005) introduced a new paradigm which demonstrated a more direct form of false feedback acceptance, furthermore without a significant time delay or the use of emotionally arousing stimuli. This paradigm, routed in the domain of preferential choice, became known as choice blindness. In the first paper to demonstrate choice blindness (Johansson et al., 2005) participants were required to select their preferred choice out of two pictures of female faces printed on pieces of card, over 15 distinct choice trials. However, on the 7th, 10th and 14th trial, after the participants made their choice, the experimenter used a sleight of the hand trick to replace the selected face with the alternative, and participants were asked to explain why they selected the presented face. On the majority of trials participants (estimated as 74-88%) failed to notice the switch and proceeded with providing a justification. After the justification was provided, the researchers engaged in a conversation with the participant to try and establish whether they really did not detect the switch or just failed to report it. Even in such discussions, the majority maintained that they did not notice anything unusual. Furthermore, when the researchers analysed the reasons participants provided for preferring the non-chosen alternative, the justifications often referenced features specific to the presented choice, suggesting that the effect was not a result of them failing to distinguish the two alternatives provided, or preferring the same features across the alternatives (Johansson, Hall, Sikström, Tärning & Lind, 2006).

It must be noted that there are substantial similarities between choice blindness and the misinformation effect. Both approaches provide participants with information that alters the representations of past memories and choice respectively, altering the perception of events experienced in the past. There are however also differences, in that misinformation effect does not necessarily provide contradicting information but what is better described as additional information, that often cannot be described as categorically false. As previously mentioned, the core elements in the misinformation effect such as time delay and emotion also appear to have limited scope in their application to choice blindness. Whilst I would like to highlight the importance of the misinformation effect in emergence of choice blindness it remains unclear how much underlying cognition is shared between this effect and choice blindness (see Sagana, 2015) and a full discussion of the effect is outside the scope of this thesis, which sets out to explore the effects of definitively contradictory feedback. Accordingly, the discussion will proceed with a focus solely on the Barnum effect and choice blindness as a way of exploring false feedback acceptance.

The choice blindness procedure of switching a chosen alternative, for a non-chosen one and asking participants to justify their choice has been applied to a range of domains since it was first reported, including faces (Johansson, Hall & Sikström, 2008; Johansson et al., 2006), abstract patterns (Johansson et al., 2008), food and drink (Hall, Johansson, Tärning, Sikström & Deutgen, 2010; Somerville & McGowan, 2016), ingredient labels (Cheung et al., 2015), personal finance (McLaughlin & Somerville, 2013), haptic choice (Steenfeldt-Kristensen & Thornton, 2013), school equipment and toys (Sauerland, Sagana, Otgaar & Broers, 2014), psychological symptoms (Merckelbach, Jelicic & Pieters, 2011), and witness testimony for incident details, faces and voices (Aardema et al., 2014; Cochran, Greenspan, Bogart

& Loftus, 2016; Sagana et al., 2013; Sauerland, Sagana & Otgaar, 2013). Some research followed the original procedure very closely. For example, in demonstrating choice blindness for the haptic, or the touch, modality Steenfeldt-Kristensen and Thornton (2013) asked participants to make preference decisions about pairs of common 3D objects that they could physically touch but not see, over 15 trials. Participants placed their hands into a specially constructed box in order to freely explore pairs of objects and verbally indicated a preference for one of them. Participants were asked to justify their choice and were allowed to haptically re-examine their preferred object. On three of these trials, a silent turntable was used to switch the preferred alternative, for the other alternative examined, before re-examination. The switch was detected on 46 percent of the trials, indicating choice blindness does occur for haptic choice.

Other studies adjusted the methodology to suit their needs (Hall et al., 2010; Sagana et al., 2013). For example, Sagana and colleagues (Sagana et al., 2013) adapted the choice blindness paradigm to investigate false feedback acceptance in a field study of eyewitness recognition of people they had seen earlier, linking back to the misinformation effect (Loftus & Palmer, 1974) briefly discussed earlier. The researchers created a situation where passers-by encountered two individuals (under the pretence of tourists asking for directions), and after the interaction a third individual approached the person and asked if they would be willing to take part in the experiment. If consent was given participants were presented with two separate line-ups, consisting of six photographs, and asked to identify the first and the second individual encountered earlier (one from each line-up). After a distractor task, participants were then sequentially presented with the photographs they selected in each line-up and asked to motivate their decision, however, the choice made in the second line-up was always switched for a different face presented in that line-up. Despite the objective nature of the decision (recognition memory) and the field methodology, Sagana et al., 2013 successfully demonstrated choice blindness, with approximately 40% of the participants failing to detect the switch in faces.

Since it was first introduced, the paradigm used to measure choice blindness has also been extended beyond discrete decisions, to scalar judgements (Hall et al., 2012; Hall et al., 2013; Merckelbach et al., 2011; Sagana, Sauerland & Merckelbach, 2014a). Two approaches have been taken to changing the responses of the participants to achieve reversal in apparent preference. The first involves switching the statement the participants rated, but keeping the rating the same, and was used in demonstrating choice blindness with moral opinions (Hall et al., 2012). The experimenters asked participants to state their agreement with a statement on a pre-set scale and then used a magic trick to switch the statement, to its reverse (e.g*., 'large scale governmental surveillance … ought to be forbidden'* was changed to *'large scale governmental surveillance … ought to be permitted'*). Participants failed to detect 52.8% of the switches. The other approach involved keeping the statements the same, but changing the rating themselves (Hall et al., 2013; Merckelbach et al., 2011; Sagana et al., 2014a). For example, Sagana and colleagues (2014a) asked participants to provide sympathy ratings for female faces on a scale of 1 to 10. The authors then presented the ratings back having altered responses by three points for 3 out of the 20 rated stimuli, and asked the participants to motivate their choices. This approach also successfully demonstrated choice blindness, showing that people were unable to detect the change in their stated response on 40.5% of the trials.

As well as the procedure itself, the definition of it means for a person to be 'choice blind' has also varied across experiments. The first, and most commonly featured approach involved assessing the justifications provided by the participants for the switched choice (Johansson et al., 2005). The responses were assessed for any mention of a switch, erroneously selecting the wrong option, or any other indication that the option presented differs from the one they selected. This is termed 'Concurrent detection'. The second approach is to assess 'Retrospective Detection' or the indication that the participants noticed the switch post experiment. This type of detection has been measured by asking participants to indicate whether they noticed anything unusual after the experiment. Additionally, some researchers have also combined all forms of detection in their analysis, terming this 'overall detection'. Each detection type listed here can produce a slight variation in results, specifically with overall detection yielding higher levels of detection compared to other types.

Furthermore, there is always as subjective element in the analysis of verbal reports which is impossible to eliminate. For instance, if a participant gives negative evaluations such as 'that alternative was unattractive', without explicitly stating that their choice was switched, do we consider this to be an indication of detection or simply as a person's preference for things 'unattractive'. Of course, in face to face studies the experimenter could request a clarification of such vague responses, yet this may in turn alert the participants to the switch that they may have overlooked before, artificially enhancing the level of detection. As a result, the actual level of detection can only be estimated and is bound to change depending on methodology used. Nonetheless, there is abundant evidence to suggest that all types of detection demonstrate some level of choice blindness and we can still measure when the level of detection changes as long as a consistent measure is used within a study, and the manner in which verbal reports are analysed is clearly set out.

In light of such variety of methodology used, when I discuss the choice blindness paradigm in the current thesis I refer to a whole host of procedural approaches that have been used to demonstrate the lack of ability to detect a mismatch between previously stated intentions and presented outcome. Every approach, however, is underpinned by the switch of a stated response, for one that is different. Choice blindness is then measured by the proportion of trials on which the switch was detected.

Whilst the problems with procedural differences can be counteracted by using a consistent method of detection and comparing the different methodologies used, it does raise a theoretical question as to what it truly means for a trial to be detected. A mechanistic model would assume we create a representation of our choice, and compare it to the feedback presented, with detection of an error if the choice is sufficiently dissimilar to the one made. The higher levels of detection observed for retrospective compared to concurrent detection however suggests that this is unlikely to accurately represent the true process taking place, as the process is clearly not an all or nothing event. Perhaps representing detection as a continuous variable that needs to reach a certain threshold for a ‘detected response’, such as likelihood of the information being incorrect (see Chapter 3.3), could account for such difference. For example, if we fail to detect a switch concurrently however judge the likelihood of the information being incorrect only slightly below the detection threshold, follow up questions about the accuracy of such information may raise suspicion and push the judgement above the threshold. In this instance, we may anticipate accurate feedback to be retrospectively judged as false on some trials due to the suspicion factor alone. Unfortunately, we do not have such data available and this will need to be researched in the future, but even with such knowledge it is difficult to understand what true detection means. For instance, if people fail to report detecting the switch but experience some level of suspicion can a trial be really classed as non-detected? In my own work, I take the view that in studying the weakness of human cognition it is important to capture all types of detection, as it is the deviation from rationality that poses the difficulty in describing the human mind. Exploring the trials classed as undetected using one measure but not the other remain of great interest however, and the underlying mechanisms and potential effects on subsequent memories and behaviour would comprise a fascinating topic for future research.

Despite the varied approaches and the wide range of domains in which choice blindness has been investigated, research has consistently found that people are prone to accepting false feedback about their choices regardless of methodology, although this has been found to vary with familiarity of the stimuli used (Somerville & McGowan, 2016), similarity of the items presented within the choice set (e.g., Steenfeldt-Kristensen & Thornton, 2013) and the time limit within which the choice needs to be made (e.g., Johansson et al., 2005).

As well as which factors may influence the level of detection exhibited, some progress has been made in understanding what underlies the phenomenon (Pärnamets, Hall & Johansson, 2015; Sagana et al., 2014a;

Somerville & McGowan, 2016), and the possible subsequent effects experiencing choice blindness may have on preferential choice and judgement ratings (Johansson et al., 2014; Merckelbach et al., 2011). These advancements will be discussed in Chapter 3.

Although the majority of research into the Barnum effect precedes choice blindness (see Poškus, 2014), there are a number of parallels that can be drawn between the Barnum and choice blindness phenomena. In a sense, both paradigms deceptively modify the outcome of choices in order to dissociate the actual and perceived behaviour, although this is achieved through manipulating aggregate scores as personality representations for the Barnum effect, whereas for choice blindness the individual outcome of a choice is manipulated. Similarly, the two types of invalid feedback acceptance played an important part in drawing attention to potential limitations of the methodology used in research, whilst Barnum effect cautioned the scientific community about using self-reported accuracy to validate individual difference measures, choice blindness has raised concerns about using participants' choices as indicators of stable preferences.

Overall, there is ample evidence to suggest that people often fail to identify false feedback about the self. This is evident in the literature on the Barnum effect and the choice blindness phenomenon. Whilst false feedback acceptance is well established, the factors that affect the likelihood of such acceptance, mechanisms that underlie it, and the possible consequences, require further research to establish a comprehensive understanding of such phenomena. This thesis aims to contribute to such understanding.

The following sections of this chapter aim to describe the research to which false feedback acceptance may be relevant. Specifically, I consider how our knowledge of false feedback acceptance relates to the study of introspection, or the access to knowledge about the self, and our ability to detect errors. In the last section of this chapter I consider the research on preferential choice, an area more specifically related to the choice blindness paradigm.

2.2 Introspection

One research area closely intertwined with acceptance of false feedback is that of introspection and metacognition, or the understanding of one's own thought processes. The very nature of accepting false feedback about the self demonstrates limitations of our introspective abilities. For example, when we readily accept a false personality profile of ourselves (e.g., Poškus, 2014; Wyman & Vyse, 2008), assuming the feedback is specific and is indeed inaccurate, we demonstrate an inability to compare the external information to our real personality traits, in turn failing to access information about the self that would be used in such comparison. Similarly, when we accept false feedback about our preferences (e.g., Johansson et al., 2005; Hall et al., 2010), we demonstrate a lack of ability to compare real preferences with the ones received during feedback, indicating that we fail to bring our real preference to mind.

Our inability to access information about the self is surprising, given that most people can not only express their attitudes and preferences, but also provide coherent reasons for such characteristics, as well as overtly expressed behaviours. Nisbett and Wilson (1977) explain such discrepancy by demonstrating that the reasons people provide are often inaccurate, and are nothing more than causal theories, or plausible explanations that are not necessarily representative of the real cognitive processes undertaken. In one of the experiments presented by Nisbett and Wilson (1977), for example, participants were presented with an identical array of night gowns and instructed to select the one they prefer the most. They were then asked to explain why they chose the way they did. Although the selection presented a strong right-side bias (right most garment was most likely to be chosen), participants failed to report order of the garments as a factor affecting their choice, yet provided other apparently coherent reasons for their selection. The study demonstrated that participants were not able to access the 'processes' underlying their choices, and failed to recognise this lack of access, instead providing alternative causal theories for their behaviours. The phenomenon of constructing inaccurate explanations has become known as ‘confabulation’. This example is just one of a string of research by Nisbett, Wilson and colleagues (Nisbett & Bellows, 1977; Nisbett & Wilson, 1977; Wilson & Nisbett, 1978) demonstrating behavioural effects that are not reflected in subsequent causal reports, or causal reports that do not reflect the processes revealed in behavioural measures.

The work by Nisbett and Wilson (1977) has come under some scrutiny after its initial publication (White, 1988; see also Wilson, 2002). Specifically, questions were raised as to whether the methodology was sufficient in capturing the cues participants used to make their decisions, as well as the sufficiency of definitions for concepts such as ‘mental process’ and ‘introspection’ itself (see White, 1988). Nonetheless, it is hard to argue with the observation that on a behavioural level, participants cannot accurately report what aspects of the environment influenced their choice. In addition, other areas of research have also documented our lack of ability to explain what guides behaviours. For example, subjects that perform behaviours outside of their control, such as under hypnosis or through magnetic stimulation of the brain, have been found to confabulate coherent reasons for their actions (Dywan, 1995; Brasil-Neto et al., 1992). Similarly, people have been found to confabulate internal explanations of their own behaviour (e.g., I must have liked that one more), when other visible external sources of explanation are not available (for discussion see Festinger, 1957; Bem, 1967).

The choice blindness research provides further evidence that people tend to provide explanations for their choices that cannot possibly be accurate. Johansson et al. (2006) demonstrated that participants provide coherent explanations for how they reached a choice that they did not, in fact, select, supporting the notion that there is a dissociation between the real cognitive process that took place to reach a decision and that suggested in the explanation provided. Furthermore, the researchers reported that in justifying the choice presented as feedback, participants tend to refer to the features specific to the presented alternative and not to the initially chosen one, indicating that people rely on external information to construct a plausible reason for their choice as opposed to inappropriately applying the real decision process that took place.

The accuracy of such introspective reports and the existence of 'insider' knowledge of the processes undergone have formed a crucial, and yet unresolved, debate in psychology and philosophy alike. Whilst the prevailing view appears to remain that we have transparent access to our cognitive processes (see Carruthers, 2011), experimental research provides a growing evidence base highlighting the inaccuracies of self-report descriptions of processes that guide behaviour (see Wilson, 2002). The demonstration of choice blindness has once again re-ignited interest in this field of research (Wilson & Bar-Anan, 2008), posing an additional observation that sometimes people do not only fail to report how (or why) they reached a choice that they did, but are not even sure of what that choice was. Raising the question of what do we actually know about ourselves?

It is rarely disputed that there are some situations in which we can access our attitudes, motivations and behavioural propensities. Such ability is demonstrated in our ability to reason out loud and monitor how closely we are following a plan (see White, 1988; Wilson, 2002). The reality appears to be that introspective access is situation and individual dependent, making it impossible to definitively say how much we know about our internal states and the processes underlying our behaviours. However, the variation in our self-knowledge appears to be systematic (consistently affected by characteristics of the situation, see section 3.1 for discussion) and therefore open to academic scrutiny. By studying environmental variables that affect false feedback acceptance we can establish when people are more likely to detect a mismatch between feedback and their own 'mental states' and thus measure when introspective access is stronger, or weaker.

2.3 Error Detection

Human behaviour is riddled with errors, and understanding how these errors can be minimised has been of great importance to a broad range of activities (see Rizzo, Ferrante & Bagnara, 1995). For example, human error is the biggest cause of accidents in the aviation and medical industries, as well as work related accidents more generally (Amalberti, 2013; Sarter &

Alexander, 2000; Ghaferi, Birkmeyer & Dimick, 2009). However, successful organisations have been found to differ in their ability to detect and neutralise errors before they lead to irreversible consequences, and not in the initial amount of errors made (e.g., Ghaferi et al., 2009). Yet the propensity to perceive false feedback as accurate appears to demonstrate that on the individual level our ability to detect a mismatch in expected and actual information about the self is very limited, even when that information is collected and then changed in a controlled environment, with minimal environmental distractors and limited range of possible outcomes. Although the study of false feedback acceptance has been largely limited to feedback on personality traits and preferential choice, it seems to question the fundamental cognitive mechanisms that have been identified as necessary for error detection; (i) a feedback mechanism with some monitoring function that compares what is expected with what has occurred, and (ii) the ability of the cognitive system to catch a discrepancy between expectations and occurrences (Norman, 1981; Rizzo et al., 1995).

In light of the limitations of introspective abilities discussed in the previous section (e.g., Nisbett & Wilson, 1977) questioning the existence of an efficient error monitoring system may seem like a natural progression. Yet our inability to detect a mismatch between our stated preference and outcome are more surprising in the context of low level psychological processes such as motor control (e.g., Bernstein, 1967; Adams, 1971; Schmidt & White, 1972; Schmidt, 1975; Scott, 2004). To illustrate, consider a study by Fourneret and Jeannerod (1998) which required participants to trace a straight line to a target approximately 20cm away. However, instead of being able to see the action of their arm directly, participants were presented with the feedback of their action on screen with a computerised cursor that distorted the movement. The study found that participants adjusted their behaviour according to the feedback, in other words, if the feedback deviation was to the right they moved their arm towards the left to compensate for this discrepancy and vice versa. The findings necessitate the ability to monitor the discrepancy between expected and actual outcomes, as otherwise it would be impossible to adjust the behaviour.

Although the study by Fourneret and Jeannerod (1998) reports the monitoring ability for real time processes that may appear to be distinct from the personality and choice literature, similarly successful error monitoring can also be observed in simple discrete decisions. For example, when choosing which button to press in response to a presented stimulus, participants successfully detect trials on which they select the wrong response even if no feedback is provided (Rabbitt, 1966). Indeed, error detection appears to be hard-wired within our cognition, with distinct neural processes dedicated to monitoring whether observed outcomes match our expectations (see Holroyd & Coles, 2002; Yeung, Botvinick & Cohen, 2004). More specifically electroencephalography (EEG) research has identified neural activity specific to error detection, termed the Error-Related Negativity (ERN; Gehring, 1992). The ERN is a sharp negative EEG signal that typically peaks 80-150 milliseconds after the motor response begins, and has been found in humans and non-human primates alike (e.g., Godlove et al., 2011) across a wide range of tasks (e.g., categorical discrimination – Gehring, Coles, Meyer & Donchin, 1995; flanker task – Jodo & Kayama, 1992; Go/No Go task – Ruchsow, Spitzer, Grön, Grothe & Kiefer, 2005; Stroop task – Masaki, Tanaka, Takasawa & Yamazaki, 2001). Additionally, the ERN can also be observed in response to negative feedback received after the task. For example, when participants were required to press a button when 1 second had elapsed following presentation of a warning stimulus, and received feedback as to whether they were in an appropriate accuracy range, if the feedback indicated that the response was not within the criterion (negative feedback) it elicited an ERN (Miltner, Braun & Coles, 1997). The ERN exhibited in response to negative feedback is often termed the feedback ERN (fERN), and has now been well documented in research (e.g., Holroyd, Hajcak & Larsen, 2006; Moser & Simons, 2009; San Martín, Manes, Hurtado, Isla & Ibañez, 2010).

Although to my knowledge neural activity of participants undergoing either the Barnum or the choice blindness paradigm is yet to be investigated, the very nature of wrong information being presented would, in theory, suggest that the same error-related neural activity should be anticipated. For the Barnum effect, it could be argued that error detection and therefore ERN are not present because the expectations of the outcome are unclear due to lack of transparency of how responses are translated to the actual personality profile, and therefore there is a lack of comparison point for the feedback seen. Indeed, measures which have a clear relationship between the response and personality generated are less susceptible to the Barnum effect (see Furnham & Schofield, 1987, for discussion). On the other hand, the wide range of domains in which the Barnum effect occurs suggests that at least some of the experiments should contain a perceivable mismatch between reality and feedback. Without an exploratory study of the neural activity, however, it is impossible to conclude whether no ERN is present, or whether it does occur but does not translate to a conscious response.

For choice blindness on the other hand, extrapolating findings from other choice tasks suggests that an ERN should be present. In fact, this may be the case on two accounts, first a direct mismatch of feedback to the action performed should result in ERN, second since presenting participants with the non-chosen alternative also entails presenting the less preferred outcome, or negative feedback, we can also anticipate a fERN. Without empirical research this is, of course, impossible to determine for sure, and the finding that the majority of people fail to notice when the feedback of their own choice is incorrect would suggest the opposite – since no error detection occurs it is unlikely that the associated neural activity does either. There is also another possibility, which is that the ERN does occur even for participants that fail to notice the error but it fails to reach the threshold necessary for the error to be detected in consciousness. It has indeed been reported that the ERN can be observed, albeit lower in strength, when participants make an error even when they are not consciously aware of this (e.g., Nieuwenhuis, Ridderinkhof, Blom, Band & Kok, 2001), suggesting that this may be a plausible hypothesis for the underlying neural activity in choice blindness.

Whether false feedback acceptance occurs because of an absence of any error-related neural activity, or because such activity fails to reach consciousness, it is nonetheless surprising that despite an established neural process in place to detect errors we often fail to do so. It is, of course, possible that the specificity of domains used in false feedback acceptance research may be responsible for the uniquely poor error detection rates exhibited by participants. However, as you will see in the following chapters the range of domains that are prone to false feedback acceptance is very broad (e.g., Hall et al., 2010; McLaughlin & Somerville, 2013; Richards & Merrens, 1971; Sagana et al., 2013; Wyman & Vyse, 2008), suggesting that it is implausible that every domain selected is uniquely prone to poor error detection. It is however, also very clear that false feedback is not perceived as accurate one hundred percent of the time, and that the proportion of trials on which people do accept false information about their characteristics and preferences varies systematically depending on the precise nature of the task at hand (e.g., Andersen & Nordvik, 2002; Poškus, 2014; Steenfeldt-Kristensen & Thornton, 2013; Somerville & McGowan, 2016). It is, therefore, crucial to investigate why error detection can be poor, and what factors are responsible for distinguishing between detected and undetected invalid feedback. This can not only help us understand how we operate on the cognitive and neural level, but also identify the best techniques that can be used to minimise error detection in real life settings.

So far, in discussing introspective abilities and error detection mechanisms, I have tried to discuss the characteristics of human cognition as a whole, extrapolating between domains and behaviour types to try and paint a comprehensive picture of how people process information. In the next section, I will narrow the field of discussion specifically to preferential choice, as this is the main domain choice blindness research set out to explore (e.g., Johansson et al., 2005, 2008; Hall et al., 2010). Accordingly, the next section excludes the contributions of Barnum effect literature from the discussion. It must be noted that since choice blindness has been demonstrated for recognition of previously seen stimuli (Sagana et al., 2013), individual characteristic ratings (Sagana et al., 2014a; Sauerland et al., 2013), and perception of psychological symptoms (Merckelbach et al., 2011), it is likely that preferential choice is simply a sub-section of choice and judgement more generally. Accordingly, much of the discussion can inform cognition beyond the confines of preferential choice alone.

2.4 Preferential Choice

Preferences are inherently subjective, varying from person to person depending on emotions, goals, past experiences and even biological differences. For most domains, it is, therefore, impossible to determine what constitutes a 'good' or rational choice. Consider for example the scenario outlined in the introduction (Chapter 1), where one is faced with the task of selecting a person they would like to date from a range of potential candidates. Whether they prefer a person who is tall or short, or blond or brunette, there is no objectively correct answer – only one that best matches the person's subjective criteria. However, whilst one choice in isolation is hard to evaluate, a set of preferences in aggregate is expected to follow a certain pattern, which has long been of interest to psychologists and economists alike. This pattern is, in turn, thought to provide a benchmark against which the quality of choices can be evaluated.

A term that best captures what we have come to expect from preferences is consistency (Rieskamp, Busemeyer & Mellers, 2006). Any sets of preferences held by a decision maker are widely assumed to be stable (consistent over time), and to have a stable order, where if alternative A is preferred to B, and B is preferred to C, A must also be preferred to C (consistent across alternatives – Houthakker, 1950; Arrow, 1959). It is further commonly accepted that subjects' choices reflect this underlying preference order (Samuelson, 1938). For example, if we choose item A over item B when both items have the same associated costs we are thought to prefer item A, and unless some learning process takes place to alter this preference, repeating the task should yield the same result. These presuppositions have formed the basis of many influential theories (e.g., Samuelson, 1938; von Neumann & Morgenstern, 1944), and are often applied in real world practice areas (e.g., marketing, consumer sales; see Jacoby, 2000).

The expectation that our preferences are broadly consistent is natural. After all, in other, more objective, forms of decision making inconsistent outcomes signal a problem. Even outside of the decision context, consistency is thought to play a fundamental part in attitude formation (e.g., Festinger, 1957) and human behaviour more broadly (e.g., Ouellette & Wood, 1998). The stable and ordered model of preferences has also proven useful in describing how a rational agent should act if they are to avoid being open to manipulation. Consider the 'money pump' example commonly quoted in economics to illustrate the importance of preference stability, where a person prefers option A over option B, option B over option C, yet option C over option A. In this example, we would expect a person to be willing to pay a small amount to upgrade option A to B, B to C and C to A, which would result in the end product being no different to that at the start, whilst the person would have made a loss. With a stable set of ordered preferences on the other hand, such manipulation should not be possible.

Whilst a stable, ordered set of preferences provide a good description of how a completely rational agent would think, how far they describe real human behaviour has come into question over the past few decades. Research has increasingly shown that people's choices are consistently affected by seemingly irrelevant aspects of the environment. Let us consider the violation of procedural invariance to illustrate this. Procedural invariance states that as far as we have a stable preference order, the method used to elicit preferences should have no impact on the person's apparent preferential order (Slovic & Lichtenstein, 1968). However, in one scenario studied by Shafir (1993), the participants were given a choice between two ice creams; ice-cream one is very tasty but very high in cholesterol and ice-cream two is just moderately tasty. When participants were presented with the task of choosing their preferred option 72% selected the tasty alternative, however, when they were required to give up one of the choices only 55% rejected the non-tasty alternative, whereas the ordered preference approach would predict equivalent proportions in the two conditions. Similarly, Tversky and Kahneman (1981) demonstrated that people are equally easily swayed by the way the alternatives themselves are presented. The most commonly cited example of this is that choices involving gains are often risk averse and choices involving losses are often risk seeking. Consider the scenario of preparing for an outbreak of a deadly Asian disease, which is expected to kill 600 people. Here the participants are given a choice of saving 200 people for certain versus a 1 in 3 chance of saving everyone and 2 in 3 chance that no one will be saved. In this scenario, the majority of people selected the certain option. However, when faced with the choice of 400 people dying for certain versus 1 in 3 chance that no one will die and 2 in 3 chance that everyone will die, only a minority selected the first alternative. Mathematically the two sets of choices are identical, but merely described differently, yet the elicited preferences for two courses of action appear to change depending on how the alternatives are presented or 'framed'.

These examples of framing are by no means unique in challenging the notion of stable, ordered and revealed preferences, and many factors which should be irrelevant from a rational choice perspective have now been identified that can alter the elicited preferences of an individual, including the order of presentation (e.g., Nisbett & Wilson, 1977), perceptual fluency (e.g., Reber, Winkielman & Schwarz, 1998) and the presence of a third, however, irrelevant, alternative (e.g., Huber, Payne & Puto, 1982) to name a few. Furthermore, even under a controlled environment when presented with a similar choice on two different occasions, people tend to change their minds around 25% of the time (e.g., Camerer, 1989; Hey, 2001; Loomes & Sugden, 1998).

Whilst problematic for rational choice theory, the study of choice inconsistencies and preference have allowed us to make considerable progress in describing real, human behaviours. The newly emerging descriptive models of choice have become more inclusive of variable and context dependent choices. The predominant themes to emerge have included recognising that decision are made under limited cognitive resources (bounded rationality – e.g., Simon, 1956), in a probabilistic rather than discrete manner (probabilistic choice – e.g., Block & Marschak, 1960; Rieskamp, 2008), and are at least to some extent constructed from the information in the surrounding environment (constructed preferences – e.g., Payne, Bettman & Johnson, 1992; Slovic, 1995). Nonetheless, there is still no single model that perfectly describes all aspects of choice. Although research has shown that the rational choice approach fails to accurately describe human behaviour, some argue that we are still to create a theory powerful and comprehensive enough to put in its place (Posner, 1993, p367). In order to formulate comprehensive models and test the proposed governing rules of preferential choice, we need to provide more data that will allow us to test these models against real behaviours, and identify any choices that may deviate from what we would expect.

Choice blindness provides one of the latest, and most striking, examples of violations of stable preference assumptions, what is more, is that it appears to do so within one choice. If we were to assume stable and accessible preferences, the occurrence of choice blindness seems very unlikely as we would be able to detect a mismatch between what we prefer and what we are presented with. Furthermore, if we fail to initially detect that the presented face is the wrong one, the action of providing a justification for choosing the wrong face should serve as an additional alert (i) as it would require us to attend to the stimulus at hand and (ii) as we would not be able to provide the justification, because we actually prefer the other alternative. Indeed, the implications of choice blindness present an important challenge that needs to be addressed if we are to progress in our understanding and description of human behaviour (Hall et al., 2010; Somerville & McGowan, 2016).

Choice blindness undoubtedly contributes to the bulk of research that has put the notion of a master list of preferences into question, it does not, however, negate the existence of preference altogether. Although the effect has been demonstrated across a wide range of domains (e.g., personal finance – McLaughlin & Somerville, 2013; haptic stimuli – Steenfeldt-Kristensen & Thornton, 2013; eyewitness testimony – Sagana et al., 2013), in some instances the levels of choice blindness have been found to be very low. For example, when presented with choices in a domain we are already familiar with less than a fifth of participants fail to report a mismatch between their choice and the feedback provided (Somerville & McGowan, 2016). This will be further discussed in Chapter 3 of this thesis, however, situations where choice blindness would seem extremely unlikely are not hard to imagine – such as choosing between being tortured and receiving a huge sum of money for example. Of course, this is speculative, but even strong proponents of the constructed preference approach agree that some preference are very stable, even from birth (Lichtenstein & Slovic, 2006), making mistaking a large reward for detrimental punishment unlikely (although I expect there would be one or two exceptions).

One significant contribution provided by the choice blindness paradigm is a new tool that can help us better understand when, and hopefully more broadly why, we can detect a mismatch in exhibited behaviour and feedback available in our environment. As Payne et al. (1992) suggest, 'common sense dictates that consistent decisions are good decision', and since it has been established with reasonable certainty that preferences are highly malleable by their environment the use of choice blindness allows us to establish in what circumstances choice consistency is best elicited. In the next chapter I will attempt to outline which factors in the environment can impact the likelihood of choice blindness being elicited, as well as the factors that affect the likelihood of accepting other forms of false information as demonstrated in the Barnum effect. This will be followed by a discussion of the role of false feedback in determining later behaviours, as well as the cognitive mechanisms put forward for explaining why false feedback acceptance occurs.

Chapter 3.

False Feedback Acceptance: Properties, Influencing Factors, and Cognitive Mechanisms.

3.1 Variables of Interest

To recap, the Barnum effect refers to the tendency to rate invalid feedback about the self as highly accurate (Forer, 1949; Meehl, 1956), whilst choice blindness refers to the inability to detect a mismatch between an alternative we selected during a choice task, and a non-selected alternative presented as feedback (Johansson et al., 2005), and both demonstrate susceptibility to false feedback acceptance. An important question that lies at the core of this thesis is under what circumstance people are more (or less) prone to accepting false feedback about their own characteristics and preferences. In this section I will discuss the progress made in answering this question in past literature, in an attempt to summarise variables that influence false feedback acceptance (see tables 1 and 2, at the end of this chapter for summary of variable discussed for choice blindness and the Barnum effect respectively). Specifically, I will give consideration to the domain for which participants receive feedback, similarity and favourability of the false feedback used compared to the real outcome, ambiguity, and task parameters (specifically for choice blindness) such as time restrictions, framing and the consideration set.

Barnum effect has been demonstrated for a wide range of tools used to measure individual characteristics. Whilst for the majority of assessment devices the elicitation of Barnum effect has been successful at least to some degree (e.g., Forer, 1949; Poškus, 2014; for exception see Layne, 1979), the accuracy ratings provided by participants have been known to vary significantly depending on the tool used in question (Richards & Merrens, 1971; Wyman & Vyse, 2008; for review see Dickson & Kelly, 1985; Furnham & Schofield, 1987; Snyder et al., 1977). For example, people have been shown to discriminate between real and false feedback better when that feedback is in the format used by valid personality measures compared to astrology (Wyman & Vyse, 2008). There is also some variation between the different recognised personality measures, with invalid feedback generated using the Rorschach test more likely to be rated as accurate compared to the Bernreuter Personality Inventory and the Life History Battery (Richards & Merrens, 1971). The false feedback accuracy ratings have also been shown to vary depending on who delivers the test and presents feedback for the task in question (Collins, Dmitruk & Ranney, 1977; Halperin et al., 1976; Snyder & Shenkel, 1976). For example, feedback is rated as more accurate when provided by individuals with higher status (Halperin et al., 1976). There are a number of possible reasons that underlie the differences observed between different domains, including the level of ambiguity associated with the feedback provided and the mystery associated with how that feedback is generated, which will be discussed later. For the moment, I would simply like to highlight that the levels of Barnum effect observed in one domain, cannot be directly generalised to other situations and we must treat the assumptions we make about Barnum effect for new measures with care.

Choice blindness has also been established in a wide range of domains, however, few studies have provided a direct comparison for how the likelihood of experiencing choice blindness compares between these domains. Some studies have compared choice blindness for faces, with choice blindness for other types of stimuli. Johansson et al. (2008) for example reported no significant difference between the proportions of people who successfully detected a switch of faces compared to a switch of abstract patterns. Somerville and McGowan (2016) on the other hand, found that choice blindness experiments with adolescents result in a higher detection rate when choosing between chocolates, compared to choosing between female faces. Sauerland et al., (2014) also report a lower detection rate for school classroom items (e.g., chairs) compared to personal items (e.g., toys).

Somerville and McGowan (2016) propose that such differences can be attributed to prior experience with the stimuli used and not the different type of item presented per se. In other words, it is unclear whether if presented with unfamiliar chocolates or toys such a difference would occur. However, variability in familiarity and experience with different types of domains is in the very nature of such domains being different and therefore it does not seem realistic to set a golden standard for what proportion of detected trials constitutes a 'normal' level of choice blindness.

One possibility is that by repeating a choice we learn which features are important and learn to differentiate these to establish our preference, creating the familiarity effect in choice blindness. Cheung et al., (2014), for example, demonstrated that using specific instruction which hold information on how the decision should be made (e.g., rate the naturalness of the product) can increase the proportion of people detecting a switch compared to general instruction (e.g., rate which one you prefer). This suggests that choice blindness could be reducing when people learn what is relevant to them in a choice, which can be taught or learnt through repetition.

It should also be taken into consideration that a lack of experience in performing the task could also potentially explain why the Barnum effect occurs when rating the accuracy of personality descriptions. It is rare that people are required to perform such task, or even see their own personality descriptions, therefore as the use of personality measures becomes more accepted the Barnum effect may be minimised. Indeed, participants that work within the psychology domain have been shown to demonstrate a higher level of detection of their choice being switched, compared to less experienced student samples (e.g., Bachrach & Pattishall, 1960). Despite a lack of an absolute level of Barnum effect or choice blindness across different domains, some systematic variations can allow us to extrapolate relative effects from one domain to another.

One parameter that appears to play a role in false feedback acceptance is the level of similarity between the outcome thought to be anticipated by the participants based on responses provided and the outcome presented. Whilst this is difficult to establish for the early definitions of the Barnum effect because very general statement used in the feedback lacked specificity that would allow similarity to be compared (e.g., Forer, 1949; Sundberg, 1955), later work using specific feedback in the format of trait personality measures has shown that similarity does indeed impact the accuracy ratings participants provide for invalid feedback (Andersen & Nordvik, 2002).

The role of similarity in false feedback acceptance is not surprising. Consider for example switching a picture of one identical twin, for a picture of the other. In this scenario, it seems reasonable that a person may fail to detect a switch, simply as a result of not being able to tell the two photographs apart. However, the inability to detect a switch of visually different faces cannot be explained in the same manner. For choice blindness, the question of whether similarity has a significant effect on the likelihood of participants detecting an invalid outcome was first put forward by Johansson and colleagues (2005) in the original choice blindness study. The study roughly matched fifteen pairs of faces on attractiveness and asked fifteen independent raters to rate the similarity of the presented pairs to identify high similarity and low similarity stimuli sets. Controversially, the authors did not find a significant difference in detection of switched outcomes experienced for high and low similarity face pairs, moreover they reported that people use features specific to the new, non-preferred faces in their justifications.

Research since has encountered this question on multiple occasions, predominantly with a very different outcome. The effects of similarity have been found to play a significant role in choice blindness for jams (Hall et al., 2010), field studies of eyewitness testimony for voices and faces (Sauerland et al., 2013; Sagana et al., 2013) financial decisions (McLaughlin & Somerville, 2013), and haptic stimuli (Steenfeldt-Kristensen & Thornton, 2013). One possibility which could explain the discrepancy in findings, is that the rough pair matching procedure and use of 15 raters utilised by Johansson et al. (2005) were insufficient to accurately represent the perceived similarity of the stimuli used in the wider population. On the other hand, Sauerland et al., (2014) also found no effect of similarity on choice blindness when using classroom (e.g., chairs) and personal items (e.g., toys), despite having a sample of 60 participants to establish the differences in similarity beforehand. The results appear to be mixed, however, with the majority of findings indicating that similarity is, in fact, an important choice blindness predictor.

Favourability of feedback presented is another parameter that has been known to influence the accuracy with which we perceive invalid feedback. There is a substantial volume of studies using the Barnum effect which demonstrate that favourable feedback is more likely to be accepted than negative feedback (e.g., Johnson et al., 1985; Macdonald & Standing, 2002; Poškus, 2014; for exception see Dmitruk et al., 1973). For example, Poškus (2014) report that personality profiles that are viewed as positive (high trait openness, conscientiousness, extraversion and agreeableness, and low trait neuroticism, Poškus & Žukauskienė, 2014,) are rated as more accurate than the inverse of such profiles. In fact, Macdonald and Standing (2002) propose, that the self-serving bias, or the tendency to attribute positive events to the self and negative event to an external source can cancel out the Barnum all-together. The validity of such proposal is highly debatable, since many studies that report negative feedback being rated as less accurate than positive feedback, still report accuracy ratings of higher than the midpoint that would be associated with being neither accurate nor inaccurate (e.g., Poškus, 2014; Wyman & Vyse, 2008). Nonetheless, it is very clear that favourability of feedback is positively related to the perceived accuracy of the information provided.

The main effect of favourability on choice blindness is more difficult to establish than for the Barnum effect. Facial attractiveness can be construed as a form of favourability or “expected utility” (Kahneman & Tversky, 1979), given that reward processing systems in the brain show higher levels of activation when an attractive face is encountered compared to an unattractive face (Winston, O’Doherty, Kilner, Perrett & Dolan, 2007). Representing faces as high or low utility limits the experimental scope, as this implies that to some degree participants always see the less attractive alternative during the switched feedback as the very action of choosing a different item suggests the ‘imposter’ choice is preferred less by the individual (assuming the existence of such preferences). As a result, when switching a choice for an attractive or unattractive alternative, we are, in fact, switching it for a slightly less or much less attractive alternative and even this approach is not always accurate as there will always be some subjective variation in perception of attractiveness.

Nonetheless, we can hypothesise that a higher discrepancy in attractiveness leads to higher switch detection. In some respects, this can be construed as rational, or in the very least evolutionarily beneficial as switching one item for another of the same utility has little consequence, whereas switching the same item for one with much lower subjective utility can be construed as negative feedback or punishment. Consider for example being offered £5 worth of euros or dollars, assuming you had to exchange them straight away after the experiment. Even if you indicated that you preferred euros, you would probably not complain if you got dollars instead. Now consider if you are offered £10 worth of euros, or £5 worth of dollars, as long as you exchange them straight away you would be expected to prefer the Euro payment, and would care if you receive the dollars instead as this represents essentially losing £5. Surprisingly, the effects of utility have not been directly investigated with respect to choice blindness, and will be addressed in chapter 5 of this thesis.

Another way of interpreting the effects stimuli characteristics have on false feedback acceptance, is by considering how these factors may contribute decision ambiguity and whether such ambiguity may be the mediating factor for how we interpret invalid information about the self. The notion of ambiguity has been key in Barnum effect literature since it was first demonstrated by Forer in 1949, who proposed that false feedback is accepted because it is ambiguous and could apply to almost anyone. Whilst the Barnum effect has since been demonstrated for specific feedback types such as numeric representations of personality traits (e.g., Poškus, 2014), studies show that false feedback based on such measures is rated as less accurate compared to feedback derived from more ambiguous measures, with an unclear relationship between the information provided by the participant and the feedback generated, such as astrology (e.g., Wyman & Vyse, 2008), or the Rorschach test (Richards & Merrens, 1971). There is general consensus that ambiguity is an important element in determining whether the Barnum effect will occur (for discussion see Dickson & Kelly, 1985; Furnham & Schofield, 1987; Snyder et al., 1977), and accordingly, it is not surprising that other factors that introduce ambiguity, such as similarity (Andersen & Nordvik, 2002) would also affect the Barnum effect through changing the level of task ambiguity. However, without an objective measure of the degree of task ambiguity, we cannot say with certainty whether such a property can be fully responsible for the variations observed in the Barnum effect. Furthermore, ambiguity cannot be used to explain the difference in ratings observed for favourable, compared to unfavourable feedback. Favourable false feedback is often rated as more accurate than its real counterpart, even for specific feedback types such as trait personality (e.g., Macdonald & Standing, 2002). Such feedback would not be considered ambiguous, as it is not prone to multiple interpretations because it should be easily rejected by bringing to mind past events countering the stated personality characteristics. And yet despite the low level of ambiguity, positive feedback is rated as highly accurate.

For choice blindness, I have already briefly mentioned that Somerville & McGowan (2016) propose that reduction in ambiguity is likely to be responsible for the high levels of switch detection observed for familiar (brand chocolates) compared to unfamiliar (facial preference) decisions. The reduction in ambiguity is, in turn, thought to be a result of substantial prior experience. Similarly, Sagana et al. (2013) note *'When combining more than one source of ambiguity, the magnitude of the [choice blindness] effect increases dramatically*.' Here Sagana and colleagues (2013) refer to similarity and short decision time as sources of ambiguity, but favourability is likely to also contribute to altering this parameter because when the level of attractiveness is similar the difference in outcome becomes inconsequential. It is also possible that once this ambiguity parameter reaches a certain level, any differences beyond that fail to alert the individual to the invalid nature of the information provided. For example, Merckelbach et al., (2011), interpret the findings that people are more likely to accept falsely elevated psychological symptoms if they already reported having a high level of those symptoms as a result of the increased ambiguity. The ability to minimise choice blindness with specific instructions (Cheung et al., 2014) also provides evidence that ambiguity is likely to play a role in the level of detection exhibited, as it directly minimises uncertainty about how to make the decision at hand.

If we consider the predictive power of stimuli characteristics to be fully mediated by choice ambiguity, this could also explain why individual predictors may have a significant influence on choice blindness on some occasions but not on others. Consider for example if two stimuli are incredibly physically similar, in this scenario ambiguity would remain high even if domain or attractiveness is manipulated thus predicting a low level of detection irrespective of the latter characteristics. Of course, there is a possibility that each of the investigated variables has a unique effect on choice blindness, and ambiguity is a distinct construct with its own contribution to the effect. Researching how ambiguity might influence choice blindness directly, may provide insight into this concept.

It seems apparent that a direct measure of ambiguity is necessary to establish its effects on the susceptibility to Barnum effect and choice blindness. One measure of decision ambiguity is to directly ask participants about their subjective confidence in the responses they provided. Surprisingly, I was not able to find any literature on the relationship between self-reported confidence and the Barnum effect. Interestingly, for choice blindness most studies have focused on analysing the decision confidence after participants are presented with the real or switched feedback (Johansson et al., 2005; Hall et al., 2010; Hall et al., 2012), yet pre-feedback confidence is seldom discussed as a predictor of choice blindness. One study by Sagana and colleagues (2013) reports the post-decision confidence to be a significant predictor of choice blindness. On the other hand, when examining the effects of self-reported certainty of political views on subsequently altered responses to political questions Hall et al. (2013) fail to report a significant relationship. Other studies (e.g., Pärnamets et al., 2015) have included confidence measures in the procedure used to elicit choice blindness but do not report the subsequent relationship to the detection of switched outcomes simply using the rating as a distractor task.

One explanation for the discrepancy in the results reported by Sagana et al. (2013) with subsequent research is the nature of the task. Whereas the majority of choice blindness research tends to focus on preferential choice, Sagana and her team (Sagana et al., 2013; Sauerland et al., 2013) have been working on applying the choice blindness paradigm specifically to procedures resembling eyewitness testimony. As a result, the choices made in this research rely on participants correctly identifying previously encountered individuals, which unlike preferential choice are dependent on memory strength and have objectively correct responses. It is possible that in this research participants have more metacognitive access to their decision quality (e.g., 'I remember the situation in which I encountered these people very well' vs. 'I do not remember the situation being described at all'), compared to subjective judgement of preferences in a novel situation. Nonetheless, the absence of a detected link between decision confidence and switched outcome detection does pose a problem for the ambiguity explanation of choice blindness in preferential choice, as easier decisions where the stimuli are different and vary in their attractiveness would be expected to be reflected in confidence.

The failure to detect an effect of confidence on choice blindness in the majority of research is also surprising in the context of other literature, which has commonly demonstrated that confidence is a good indicator of choice consistency. For example, self-reported attitude confidence has been found to lead to greater consistency between the attitudes and behaviour (Bizer, Tormala, Rucker & Petty, 2006; Fazio & Zanna, 1978; Glasman & Albarracín, 2006; Tormala & Petty, 2004) as well as greater choice consistency over time (Koriat, 2012). Although there are many covariates of choice confidence that may be responsible for these findings, Tormala, Clarkson, and Petty (2006) found that even when only perceived choice confidence was manipulated through providing participants with altered confidence feedback, people's attitudes became more predictive of behavioural intentions as perceived confidence increased. Although, this research demonstrates that increase in perceived confidence, in turn, increases self-reported behavioural consistency, this also highlights the potential problems in using confidence as a measure of consistency. First, just as choice blindness shows that people have a limited ability to monitor their choices and preferences, Tormala et al., (2006) demonstrate that confidence is strongly dependent on external feedback and thus is a malleable concept that we might not have direct access to. Secondly, since this research uses behavioural intentions as opposed to actual behaviour, the authors can only hypothesise that confidence mediates actual behaviour. In the choice blindness paradigm, the equivalent of such a measure would be the likelihood at which participants would judge noticing a mismatch of the outcome with their decision, which Johansson et al. (2006) show to be very high (86%) indicating that in this instance there is a clear dissociation between self-reported and actual behaviour (see section 2.2 for discussion on accuracy of introspective reports).

It appears that the nature of the stimuli with respect to domain, physical similarity and relative attractiveness of the stimuli do influence choice blindness. The results also indicate that the direction of influence of the aforementioned variables is consistent with various levels of decision ambiguity elicited by the options, suggesting the effects observed may affect switch detection indirectly. On the other hand, at least on the subjective level it appears that choice confidence and certainty rarely mediate choice blindness posing that the role of ambiguity should be treated with caution.

Another way of approaching variability of false feedback acceptance is by considering differences associated directly with the procedure itself as opposed to the stimuli used. I will now consider such variables, including decision times, decision strategy, problem framing, number of alternatives and various social effects that may influence false feedback acceptance, discussing ambiguity as a potential mediating factor where appropriate. I was not able to find research investigating how procedural variations of time, framing or number of alternatives (or scale range) impact the Barnum effect, perhaps due to the Barnum effect being less embedded in cognitive psychology fields, and more in social psychology, whilst this is an interesting question to try and address in the future, for the purpose this discussion I can only focus on the procedural variations associated with the choice blindness paradigm. The Barnum effect will be discussed in more detail with respect to demand characteristics and social desirability at the end of this chapter.

In the previous section, I mentioned that Sagana et al. (2013) consider decision time to play an important role in acceptance of false feedback, although this was not empirically tested within their study itself. Such relationship has however, been empirically addressed in other choice blindness research from two distinct approaches; by investigating how limiting decision time influences choice blindness and by measuring how varying decision time affects choice blindness when participants are free to take as long as they want. In the original choice blindness paper (Johansson et al., 2005) participants were assigned to one of three possible decision time conditions; 2 second and 5 second decision time limits, and a condition where participants could take as much time as they like. The results demonstrated that imposing time limits on the choice task, decreases the probability that participants will detect their chosen outcome being switched for a different alternative, regardless of whether the limit is 2 or 5 seconds. Similarly, McLaughlin and Somerville (2013) measured the time participants took to decide on a pension portfolio, before switching some of the funds included in the portfolio. They found that participants that took longer to reach their decision were more likely to notice the switched elements later. It, therefore, appears that taking longer to study the alternatives and reach a choice, results in increased ability to detect a mismatch between intended and actual outcome. On the other hand, Hall et al., (2012) find no significant effects of decision time on the likelihood of experiencing choice blindness for moral judgements.

As far as imposed time limits are concerned, the findings appear to be in line with the hypothesis that ambiguity mediates choice blindness, as imposing time limits on a choice can increase decision difficulty (e.g., Haynes, 2009), thus increasing ambiguity of the outcome. Then again, the mixed findings regarding time taken to make a choice when no time limit is imposed present a challenge for interpretation. Longer time spent making the choice would suggest more information accumulated about the alternatives thus increasing certainty of the choice, yet since ambiguity increases choice difficulty longer time spent on the decision could also be related to higher decision ambiguity (Rolls, Grabenhorst & Deco, 2010). Perhaps the effects of decision time are dependent on the specific circumstances in which they are assessed, and the difference between the financial decisions investigated by McLaughlin and Somerville (2013) and the moral judgements investigated by Hall and colleagues (2012) are sufficient to moderate the effects. Since the majority of choice blindness research (e.g., Hall et al., 2010; Hall et al., 2013; Johansson et al., 2014; Sagana et al., 2013; Sauerland et al., 2013; Pärnamets et al., 2015) does not vary the time constraints imposed on the decision, or measure the time taken to reach the decision, it is difficult to establish when and how time taken to reach a decision can influence choice blindness. For the moment, we can only conclude that artificially restricting decision time leads to a decrease in detection of mismatches between expected and presented outcomes, whereas the role of the time taken to reach a decision requires further research under conditions that allow to control for any other possible variations of decision difficulty.

Since time pressure has been found to influence choice blindness, it seems reasonable that other factors that may limit cognitive processing are likely to also. One such factor is the amount of information participants have to consider, which can be manipulated by increasing the number of choice alternatives presented in the task. There have been two studies that have used multi-alternative variations of the choice blindness paradigm. First, Sagana and colleagues (2013) asked participants to identify a face out of six possible alternatives and reported a slightly higher switch detection rate (59%) compared to other studies on facial choice blindness (e.g., 26% in Johansson et al., 2005; 43% in Somerville & McGowan, 2016; 12% in Johansson et al., 2008). However, as mentioned previously, cross-study comparisons are very difficult given that many procedural differences are likely to impact their outcome. For this study in particular, there were many deviation from the conventional paradigm as the researchers used a field method in order to model the choices as closely as possible to the legal system, thus the choices were made in a realistic setting and the decision was such that it had an objectively correct answer unlike preferential choice. Since Sagana et al. (2013) did not vary the number of alternatives or provide a binary control group under similar conditions the study fails to provide insight into the effect of alternatives, beyond perhaps an indicative trend.

These results may appear surprising, as it seems that increasing the number of alternatives should also increase cognitive load and thus decision difficulty, yet the findings are, in fact, consistent with other related areas of choice stability. For example, DeShazo and Fermo (2002) find that people are more consistent with a single utility function when they have more than two, but less than five, alternatives. Similarly, Collins and Vossler (2009) report that people are more likely to make an optimal decision for ternary compared to binary choices. Overall, the trend observed in the experiment by Sagana et al. (2013) appears to be congruent with other research on choice stability. There are a number of reasons that this might be the case, DeShazo and Fermo (2002) for example suggest that the effect could be a result of a higher probability that a participants will encounter a stimulus to their taste when there are three versus two alternatives; another approach would be to consider how individual parameters (e.g., physical similarity and attractiveness – see beginning of this chapter) may be treated differently when there is a third alternative present, thus potentially increasing the salience of one option which may be more memorable (Huber et al., 1982). Whilst at this stage the effect of increasing the number of alternatives on choice blindness is nothing more than a hypothesis, this will be discussed in more detail in chapter 5 of this thesis.

Another variable of interest within the choice literature is the framing of the task at hand. The way we present the question instructions (Shafir, 1993), possible alternatives (Tversky & Kahneman, 1985) and the response format (Slovic & Lichtenstein, 1968) have all been found to influence task outcomes, even when from the logical perspective, the task is identical. The difference in choice outcomes suggests that the cognitive processes undergone must diverge also, allowing for the possibility that these processes result in varying levels of choice blindness. Cheung et al (2015) demonstrate that task instructions can indeed influence the rate of switch detection, at least where food ingredient based choices are concerned. Cheung et al.'s (2015) study, does not just manipulate the way that information is presented but also provides participants with additional information through specific instructions. It does, however, demonstrate that small changes in the question can indeed influence choice blindness which in the very least does not eliminate the possibility of framing effects.

In the previous section I have touched upon two response modes that have been used to elicit choice blindness; categorical choice versus continuous judgement. The response modes have never been directly compared to my knowledge, and given that there is very little overlap in domain between the choice and judgement studies carried out it is very difficult to contrast the two procedures. Sagana, Sauerland and Merckelbach (2014b) conducted the only choice blindness study using facial stimuli alongside a judgement response, eliciting 59% switch detection. Although the detection rates are at the higher end of those previously reported with facial choice blindness paradigms, they appear to be within the expected range. Furthermore, since the judgement ratings in question were in reference to facial sympathy, and not preference this once again creates a barrier to comparing this study with others. Whether asking participants to provide a judgement or presenting them with a categorical choice can impact choice blindness will remain a mystery for the moment.

Although distinct from the effects of response mode directly, two studies have used judgement ratings alongside the categorical response provided in the choice blindness paradigm. Johansson and colleagues (2008), for example, demonstrated that providing a preferential judgement for each of the female faces that were presented in the choice after they make their decision increases the ability to detect the face being switched. This is likely to be a result of deeper processing when two types of cognitive operations are performed. Conversely, another study that intended to induce a judgement based decision strategy in participants by asking them to judge a number of faces prior to encountering the decision (Cooke, Kirichek, Kusev, in preparation) found that using a judgement strategy, in fact, decreases the level of detection exhibited – a surprising finding. Perhaps this is indicative of the level of choice blindness associated with judgement response mode, or simply indicative of decision fatigue, but the overall conclusion appears to be that inducing judgement based decision strategy can impact choice blindness, reducing detection. On the other hand, combining choice and judgement appears to increase the switch detection.

The general conclusion that can be drawn from the procedural variations of the choice blindness paradigm, is that even small changes in the task such as time restrictions, number of alternatives and response mode could potentially alter the likelihood of choice blindness being induced. However, although I tried to paint a comprehensive picture of the relevant research, it is apparent that more empirical evidence is needed to conclude which procedural differences are of importance and to what degree. In addition, the study of such effects needs to be expanded to the Barnum effect, as well as choice stability, if we are to understand the full extent of what it is about the choice blindness paradigm that makes people accept invalid preferences as their own.

Lastly, I would briefly like to discuss the potential role of social desirability effects or demand characteristics on false feedback acceptance. The majority of research on the Barnum effect and the choice blindness paradigm alike has been conducted face to face, with the experimenter personally presenting the false feedback to participants (e.g., Andersen & Nordvik, 2002; Hall et al., 2010; Johansson et al., 2005; Poškus, 2014). Such proximity to the experimenter could make the participants more vulnerable to demand characteristics, or participants forming an interpretation of the experiment's purpose and subconsciously changing their behaviour to fit that interpretation (Orne, 1962). Consider for example the first choice blindness study (Johansson et al., 2005), where the participants made 15 choices (7th, 10th and 14th manipulated) from alternatives presented physically by the experimenter. The experimenter's behaviour could have influenced the participants' responses either implicitly, or on a conscious level where participants may have felt too embarrassed to point out that the experimenter made a mistake. Furthermore, the manipulated trial is first encountered after participants have already encountered 6 pairs of faces, which poses a problem in that having encountered numerous trials participants may build trust with the experimenter amplifying the demand characteristic which may have resulted from face-to-face experiments in the first place (Kintz, Delprato, Mettee, Persons & Schappe, 1965; Nichols & Maner, 2008). The physical presence of the experimenter, and multiple trials, have not been directly investigated in choice blindness, thus the exact magnitude of any unintentional experimenter effects exerted cannot be established. However, a number of studies (McLaughlin & Somerville, 2013; Sauerland et al., 2013, 2014) have attempted to establish whether demand characteristics play a role by investigating whether individual differences in susceptibility to social desirability (e.g., using Marlowe–Crowne Social Desirability Scale, Crowne & Marlowe, 1960) impact detection of invalid feedback. Since no significant effect was detected, the authors concluded that it is unlikely that social desirability is responsible for choice blindness. Furthermore, cross-study comparison also show that computerised choice blindness tasks (Sauerland et al., 2014; Pärnamets et al., 2015) are successful at eliciting choice blindness, although the detection rates do appear to be higher (54.8-62.8%) than the 20% average detection originally reported in Johansson et al. (2005).

Similarly, it is unlikely that demand characteristics can fully explain the Barnum effect, as the phenomenon has been successfully demonstrated with computerised, as well as face to face tasks (e.g., Guastello & Rieke, 1990; O'Dell, 1972). Direct comparisons of whether people believe the feedback was generated by a psychologist, or a computer (Baillargeon & Danis, 1984; Snyder & Larson, 1972), also showed no significant effect on perceived accuracy of false feedback, although it must be noted that the level of interaction with the experimenter remained high in both conditions, so it is uncertain to what degree this would reduce demand characteristics. On the other hand, the level of Barnum effect has been found to vary with the type of experimenter used to deliver the feedback, with the higher status, as well as better liked experimenters resulting in higher accuracy ratings of false feedback (Collins et al., 1977; Halperin et al., 1976; for studies that failed to detect an effect see Ulrich, Stachnik & Stainton, 1963; Snyder & Larson, 1972). This suggests that social desirability is likely to play a role in false feedback acceptance after all. Furthermore, there also appears to be a significant relationship between the Barnum effect, and participants scores on social desirability scales (Mosher, 1965; Orpen & Jamotte, 1975; Snyder & Larson, 1972), however, this relationship appears to be dependent on the favourability of the feedback presented with higher accuracy ratings provide for positive and neutral false feedback, and lower accuracy ratings for negative feedback (Snyder & Larson, 1972).

The research presented provides some evidence that social desirability could play a role in choice blindness, and even more so in the Barnum effect, however, cannot account for false feedback acceptance completely. This leads us to the conclusion that whilst social desirability effects are likely to contribute to false feedback acceptance, this is likely to interact with other characteristics of the task at hand (e.g., Furnham & Schofield, 1987; McLaughlin & Somerville, 2013). I am by no means suggesting that social influences are irrelevant to the cognitive processes involved in choice, in fact, it is very likely that the desire to appear consistent with our past selves plays a large role in determining our behaviour. However, as I perceive it, the aim of the Barnum effect and choice blindness research is to determine whether we possess the ability to monitor information about the self regardless of social factors, at least in part.

Overall, in this section, I have given consideration to factors that may influence false feedback acceptance. It appears that the tendency to accept false feedback, whether about our characteristics or preferences, is robust across a wide range domains and situations, however, the degree of such acceptance is variable. The task domain, similarity of the false feedback to the expected outcome, and favourability of that feedback all appear to influence the perceived accuracy of the information provided in some circumstances, as well as decision time, response mode, and social desirability effects with respect to the choice blindness paradigm specifically. The ability to use procedural variations to predict the false feedback acceptance are crucial to our understanding of why and when we accept false feedback, and to finding ways to minimise such acceptance, however, as discussed, the evidence is still scarce and often mixed. The original empirical work presented in this thesis largely aimed to contribute to our understanding of the factors that influence false feedback acceptance. Prior to proceeding with presenting the discussion of my findings, the remainder of this chapter will consider the effects false feedback acceptance has on subsequent behaviours, as well as the cognitive mechanisms that have been proposed to explain our varied ability in judging accuracy of information provided with respect to the self.

Table 1. Characteristic of choice blindness studies and the variables of influence.

Reference	Concurrent Detection	Retrospective Detection	Overall Detection	Domain	Response Mode	Response Medium	Manipulated trials	Similarity	Decision Time	Confidence	Importance of Domain	Familiarity	Social Desirability
Aardema et al., 2015	-	-	44%	Impressions of Traffic Accident Story	Judgement (out of 6)	Face-to-face	1 out of 10	-	-	-	-	-	-
	-	-	80%										
Cheung et al., 2015	11-24%	-	-	Food Ingredient Labels	Judgement (out of 10)	Face-to-face	1 out of 1	-	-	-	-	n.s.	-
Hall et al., 2010	14%	6%	33%	Jam Preference	Choice (out of 2)	Face-to-face	1 out of 2	sig.	-	n.s.	-	-	-
	14%	7%	32%	Tea Preference									
Hall et al., 2012	34-49%	1-11%	44-50%	Moral Preference	Judgement (out of 9)	Face-to-face	2 out of 12	sig.	n.s.	n.s.	n.s.	-	-
Hall et al., 2013	-	-	22-53%	Political Attitudes	Judgement (%)	Face-to-face	avg. 7 out of 12	n.s.	-	n.s.	n.s.	-	-
Johansson et al., 2005	13%	-	26%	Facial Preference	Choice (out of 2)	Face-to-face	3 out of 15	n.s.	sig.	-	-	-	-
Johansson et al., 2006	-	-	28%	Facial Preference	Choice (out of 2)	Face-to-face	3 out of 15	-	-	-	-	-	-
Johansson et al., 2008	-	-	12%	Facial Preference	Choice (out of 2)	Computerised	3 out of 15	-	-	-	-	-	-
	-	-	19-39%	Abstract Pattern Preference									
Johansson et al., 2014	10-11%	16-22%	27-32%	Facial Preference	Choice (out of 2)	Face-to-face	3 out of 15	-	-	-	-	-	-
McLaughlin & Somerville, 2013	-	-	29% - 37%	Financial Portfolio Preference	Funds Allocated (%)	Face to Face	3 out of 6	sig.	sig.	-	-	sig.	n.s.
Merckelbach et al., 2011	-	-	63%	Psychological Symptom Severity	Judgement (out of 5)	Face-to-face	2 out of 90	-	-	-	-	-	-
Pärnamets et al., 2015	-	-	63%	Facial Preference	Choice (out of 2)	Computerised	8 out of 36	-	-	-	-	-	-
Sagana et al., 2013	31%	28%	59%	Facial Recognition	Choice (out of 6)	Face-to-face	1 out of 2	sig.	-	sig.	-	-	-
Sagana et al., 2014	-	-	41%	Facial Sympathy	Judgement (out of 10)	Face-to-face	3 out of 20	-	-	n.s.	-	-	-
Sauerland et al, 2013	19%	10%	29%	Voice Sympathy	Judgement (out of 10)	Face-to-face	2 out of 3	sig.	-	-	-	-	n.s.
Sauerland et al, 2014	10-11%	55-58%	-	School Equipment	Choice (out of 2)	Computerised	1 out of 5	n.s.	-	-	n.s	-	n.s.
	51-63%	53-73%	-	Toy Preference		Face-to-face	2 out of 5						
	50-65%	-	-	Erasers Preference		Face-to-face	2 out of 5						
Somerville & McGowan, 2016	21%	-	43%	Facial Preference	Choice (out of 2)	Face-to-face	1 out of 3	-	-	n.s.	n.s.	sig.	-
	64%	-	80%	Preference for Chocolate Images									
	59%	-	88%	Preference for Real Chocolates									
Steenfeldt-Kristensen & Thornton, 2013	21%	25%	46%	Haptic (Touch) Preference	Choice (out of 2)	Face-to-face	3 out of 15	sig.	-	n.s.	-	-	-

Table 2. Characteristic of Barnum effect studies and the variables of influence (due to the volume of literature on Barnum effect, the studies reported here are limited to examples used in section 3.1)

Reference	Accuracy Rating (as % of scale)	Measure Effects	Response Mode	Generality	Similarity	Favourability	Rater Characteristics	Feedback Characteristics	Demand Characteristics	Self-Other
Andersen & Nordvik, 2002	-	Five Factor Personality	Accuracy scale of 1-7	Specific	sig.	-	Gender	-	-	-
Billargeon & Danis, 1984	54-81%	Fictitious 11 Descriptor Personlity Profile	Accuracy scale of 1-9	General	-	sig.	-	Handwritten vs. Printed	-	Favourable Only*
Collins, Dmitruk & Ranney, 1977	52-81%	Manifest Anxiety Scale, North Dakota Null Hypothesis Brair Inventory	Accuracy scale of 1-5	General	-	sig.	-	Test vs. No Test*	-	-
Dmitruk et al., 1973	81-100%	Value survey & Projective Figure Drawing Test	Experimenter Judgement-Binary	General	-	n.s.	-	Psychologist vs. Student	-	-
Forer, 1949	86%	Astrology	Accuracy scale of 1 to 5	General	-	-	Gender, Age, Education	-	-	-
Guastello & Rieke, 1990	69-78%	16 Personality Factor Questionnaire	Accuracy scale of 0 to 10	Specific	-	sig.	-	Computer Rated	-	-
Halperin et al., 1976	76-86%	Rorschach Inkblot	Accuracy scale of 1-5	General	-	sig.	Gender	Disgnostician Status*	-	Sig.
Johnson et al., 1985	40-78%	Astrology & Fictitious Statements	Accuracy scale of 1-9	General	-	sig.	Gender	Personalised vs. General	-	Sig.
Macdonald & Standing, 2002	60-80%	Eysenck Personality Inventory	Accuracy scale of 1-7	Specific	-	sig.	Gender	-	Identifying the Purpose of experiment	-
Mosher, 1965	22-67%	Fictitious Projective Test	Accuracy scale of 1-5 (reversed)	General vs. Directional*	-	sig.	-	-	Social Desirability Scale* (Interaction with Favourability)	-
O'Dell, 1972	69%	16 Personality Factor Questionnaire & Fictitious Statements	Choice of real vs. fake	General	-	-	-	Psychology Test vs. Prosecuting Attorney*	-	-
Orpen & Jamotte, 1975	47-70%	Completed Marlowe-Crowne Social Desirability Scale & Internal-External Locus of Control Scale; Receive Fictitious Profile	Accuracy scale of 1-5	General	-	-	-	Computer vs Psychologist vs Student	Social Desirability Scale	-
Poškus, 2014	71-86%	Five Factor Personality	Accuracy scale of 1-7	General vs. Specific*	-	sig.	Personality*	-	-	-
Richards & Merrens, 1971	74-90%	Rorschach, Life History Questionnaire Bernreuter*	Accuracy scale of 1-5 (reversed)	General vs. Specific*	-	-	-	-	-	-
Snyder & Larson, 1972	76-92%	Completed Marlowe-Crowne Social Desirability Scale & Internal-External Locus of Control Scale; Receive Fictitious Profile	Accuracy scale of 1-5 (reversed)	General	-	-	-	Computer vs. Person, Written vs Oral	Social Desirability Scale	sig.
Snyder & Schenkel, 1976	78-88%	Rorschach Test	Accuracy scale of 1-5	General	-	sig.	Locus of Control*, Gender	Written vs. Oral, Personalised vs. General*		Favourable Only*
Sundberg, 1955	41% Correct Identification	Minnesota Multiphasic Personality Inventory	Choice of real vs. fake	General vs. Specific*	-	sig.	Belief in Measure, Gender	Different Psychologists	-	n.s.
Ulrich et al., 1963	81-88%	Bell Adjustment Inventory and the House-Tree-Person (HTP	Accuracy scale of 1-5 (reversed)	General	-	-	-	Psychologist vs. Student*	-	-
Wyman & Vyse, 2008	67-69%	Astrology & Five Factor Personality	Accuracy scale of 1-9	General vs. Specific*	-	-	Belief in Measure, Gender	-	-	-

* - Denotes significance where details of effect are specified.

3.2 Behavioural Change, False Feedback Acceptance and Relevance for Application

From the discussion so far, it is evident that the tendency to accept invalid feedback about the self as accurate is widespread and robust. As long as the perceived accuracy of such feedback is real, and not a result of faked agreement due to social desirability effects (see section 3.1 for discussion), the very nature of accepting inaccurate information should lead to a subsequent update of our knowledge of the self (at least short term). Given our tendency to remain consistent with past decisions and perceived behavioural propensities, one would expect that if our knowledge of ourselves changes, accordingly so would behaviour. On the other hand, it is possible that some deeper, internal factors such as biological predispositions, or pre-determined attitudes guide behaviour above and beyond momentary changes in our self-perception based on external information. False feedback acceptance paradigms allow us to dissociate such pre-existing propensities to behaviour from perceptions guided by externally presented information and therefore lead to better understanding of how behaviours are brought about. The experiments conducted to date appear to be in consensus, demonstrating that accepting false feedback does indeed lead to behavioural change, for both the Barnum effect and choice blindness (Halperin & Snyder, 1979; Johansson et al., 2014; Kusev et al., in preparation; Merckelbach et al., 2011; Sakamoto et al., 2000).

For Barnum effect, I was able to find two studies which investigated subsequent behavioural effects of being presented with invalid feedback about the self; one looking into effects of enhanced personality feedback on treatment for fear of snakes (Halperin & Snyder, 1979) and the other looking at effects of extraversion feedback on stranger interaction (Sakamoto et al., 2000). Halperin and Snyder (1979) reported that feedback suggesting high potential to change leads to a positive influence on outcome for participants being treated for fear of snakes. Furthermore, this difference is reported for both self-report (Snake Fear Questionnaire) and behavioural (Behavioural Avoidance test) outcomes. The second study by Sakamoto et al. (2000)

investigated the effects of extraversion feedback on self-perception and interaction with a confederate. The study used four questionnaire formats; ‘academic’ personality tests and ‘popular’ personality tests, in multiple choice and open-ended question versions. The results indicated a higher self-image, increase in conversation with a confederate, better impression of the confederate by the participant as well as of the participant by the confederate for the group that receive high extraversion feedback, compared to the group receiving high introversion feedback, although it is important to note that the effect does not appear to be consistent across different survey types.

The research suggests that in accepting fake descriptions of the self, people tend to update their knowledge of the self, in turn changing their behaviours to be more consistent with the received feedback. It is however, important to highlight that both experiments had a number of limitations, in that both failed to measure whether the feedback was actually perceived as accurate, used female only samples, failed to provide a real or neutral feedback control group, and used procedures highly prone to social desirability effects as the feedback was often delivered in prose and face to face. Whether the effects of false feedback remain in more stringently controlled conditions will be further explored in chapter 4 of this thesis, however, for the moment we can only conclude that experiencing the Barnum effect can indeed influence subsequent behaviours.

The effects of accepting false feedback about the self has also been investigated in choice blindness literature (Johansson et al., 2014; Merckelbach et al., 2011; Kusev et al., in preparation). For example, Johansson et al. (2014) examined the effect of choice blindness on future choices of female faces. As per the classic paradigm, the researchers presented participants with pairs of stimuli, in this instance two pictures of female faces, and asked them to select the preferred option. On some trials, following their choice, participants were shown the non-selected option and asked to explain why they chose it. In line with past research on the majority of trials, participants failed to detect that the wrong face was presented and provided a justification for their choice. After the paradigm was complete Johansson and colleagues (2014) asked participants to make their choices again, and interestingly for the sets where participants previously provided a justification for their non-chosen face they were then more likely to select the face they justified and not the one they originally selected. Similarly, Merckelbach et al., (2011) reported that participants that did not detect alterations made to self-reported intensity of psychological symptoms, subsequently changed their perceived symptom intensity in the same direction as the change (tested after 10 minutes, and after a week), whilst participants that did detect the switch exhibited similar responses to those initially reported.

The effect of erroneous feedback altering future choice has also been documented when conducting the choice blindness paradigm with risky choice (Kusev et al., in preparation, reported in Chater, Johansson & Hall, 2011). The study consisted of asking participants to select between hypothetical gambles. Each set of options contained a certain and a risky financial option in the following format: "*What would you prefer: alternative (A) 45% of losing £100, or alternative (B) a certain loss of £50?*". After participant completed their choice they were presented with the stimuli again with the preferred option highlighted and asked to confirm or reject their earlier choice. The study found that not only do the participants fail to notice manipulations of what level of risk they are willing to accept, but they also change their overall risk preferences for repeated choice scenarios, and in some conditions even show a complete preference reversal for the probability levels.

Interestingly, in my own work using graphical representations of risk I failed to replicate this finding. A small-scale study, using the data from 32 participants (19 female), was conducted to assess whether switching stimuli in a pre-set direction (higher risk vs. lower risk) for each participant would result in preference change in the specified direction. Participants completed 35 sets of choices between 'gambling spinners' (pie charts with the size of segment representing probability and colour representing potential gain), providing a confidence rating and justifying why they selected the way they did for each choice. On half of the trials, the participant's choice was switched in the pre-determined direction, if their choice was not already in agreement with that direction. Immediately afterwards they were asked to repeat the choices in a randomised order but without the justification. At the end of the experiment, participants were required to indicate whether they noticed anything unusual, followed by a more specific question of whether they believe the choices they had to justify were switched. Analysis of the individual choice justifications and post-experiment questions revealed that 17 people (53.1%) detected at least one of the switch trials, suggesting a successful induction of choice blindness in half of the participants. However, there was no difference in the preference for risky alternative in the subsequent choice between the participants who received different feedback types. Although the overall consistency between choices in phase one and two was fairly low (average 52.3% consistent choices), and participants in all condition became more risk seeking (from 38.5% to 49.7%) on phase two. Overall, switching responses in a pre-determined condition did not appear to impact later choices irrespective of whether participants experienced choice blindness. The continuation of the study was postponed until the details of the Kusev et al., (in preparation) study are published in order to allow comparison of methodology and stimuli used. However, the results do warrant caution in extrapolating the carry over effects of the choice blindness paradigm to other domains.

The findings that changing our perceived responses can alter subsequent decision is consistent with past findings that repeating a choice strengthens preferential ratings and increases the likelihood that the same selection will be made if the choice is repeated (e.g., Brehm, 1956; Sharot, Fleming, Yu, Koster & Dolan, 2012; Hoeffler & Ariely, 1999). One classic demonstration of how making a choice can alter future preference was demonstrated by the Free Choice Paradigm (FCP; e.g., Brehm, 1956). The FCP involves participants rating a set of alternatives, then choosing between similarly rated items and finally rating the alternatives once again. The findings demonstrate that after participants choose between alternatives they rate the chosen alternative as better and the rejected alternative as worse compared to their pre-choice responses (Brehm, 1956), furthermore, this effect also appears to be long lasting (Sharot et al., 2012). Although the interpretations of this effect have varied (Chen & Risen, 2010; Izuma & Murayama, 2013), it has been consistently replicated in a range of domains (for a review see Harmon-Jones & Mills, 1999). Similarly, making repeated choice has also been found to lead to greater choice stability (Hoeffler & Ariely, 1999). An analogous effect can also be seen for perception of our own attitudes more broadly. For example, stating a belief has often been found to strengthen the attitudes expressed (see Festinger, 1957), and increase the likelihood of behaving in line with that attitude (e.g., Cioffi & Garner, 1996; Freedman & Fraser, 1966). This phenomenon has not only been consistently replicated, but is now also commonly used in behavioural change literature (e.g., Dolan, Hallsworth, Halpern, King & Vlaev, 2010).

Paradigms used to study false feedback acceptance presented an interesting approach to self-consistency as they allowed to dissociate the action made from the behaviour perceived, demonstrating that the latter does contribute to determining behaviour at least in part. Consider, for example, an alternative explanation to describing why stating a choice, or attitude, might impact subsequent behavioural propensities, in that the process of performing an action automatically strengthens an association of the chosen alternative with positive appraisal which in turn would result in perceived consistency. Whilst experiments using the FCP cannot distinguish between this account, and one involving self-perception, the literature on choice blindness allows us to conclude that perception plays an important role in determining behaviour, as the alternative account would predict that the choice blindness paradigm would lead to the originally chosen alternative being preferred – the opposite of the observed results (Johansson et al., 2014).

The observation that changing or reinforcing, information provided about self can alter future behaviour is not surprising, as people tend to seek consistency and patterns in their behaviour, to the extent that even irrelevant environmental cues have been shown to 'anchor' people's choices (see Ariely & Norton, 2008). Evidence from the Barnum effect and choice blindness, provide further support for such arbitrary consistency with past perceived behaviours, although extrapolation of such effects to other domains should be done with caution. Furthermore, it is also important to highlight that the behavioural change is achieved once a participant accepts the invalid information is accurate (e.g., Merckelbach et al., 2011), a phenomenon that is not achieved in one hundred percent of case and varies depending on the exact nature of the task at hand (see section 2.1 for discussion).

When talking about behavioural change it is natural to also consider the potential practical applications of false feedback acceptance. For instance, the demonstration that false feedback about propensity to change can influence susceptibility to phobia treatment (Halperin & Snyder, 1979) has direct implications for clinical applications, as it provides a method that could enhance the effect of therapy, albeit in an ethically questionable manner. Clinical applications have also been considered for choice blindness by Aardema et al (2014), who demonstrated that the rate of detection is related to traits associated with obsessive compulsive disorder. Whilst this is not a direct application per se, it has the potential to aid the diagnosis and even treatment of the disorder. For instance, if we can reduce confabulation associated with choice blindness by training people to detect it, would this in turn reduce the obsessive compulsive disorder symptoms they experience? Whilst Aardema et al. (2014) only demonstrate co-occurrence of choice blindness and symptoms, the authors do hypothesis there is a possible causal effect of susceptibility to choice blindness on the disorder, which with future research may turn out to be a useful tool.

Another applied field which has been impacted by choice blindness is eye witness testimony. Choice blindness has now been successfully demonstrated for recognition of faces and voices encountered (Sagana et al., 2013; Sauerland et al., 2013), and even the recalled details of an observed incident (Cochrane et al., 2015). This poses a grave concern for our judiciary system and puts forth the question of how do we minimise altered recollection and errors in a field where human testimony is often at the core of determining the fate of victims, offenders and the wrongly accused.

The behavioural effects of false feedback acceptance and the potential implication for the real world setting reinforce the importance of understanding when and why such acceptance occurs, and the following chapter will attempt to address this further by considering the theories proposed to explain both the Barnum effect and choice blindness.

3.3 Mechanisms

In the previous sections of this thesis, I outlined the nature of false feedback acceptance and described how variations in the adopted procedures, and the feedback presented itself, can contribute to eliciting different levels of such acceptance. I will now turn to a discussion of why these effects might occur, outlining the theories and mechanisms that previous literature has considered core in explaining the Barnum effect and choice blindness, and the potential compatibility of these theories.

The first question that seems fundamental to explaining why false feedback acceptance occurs, is why people do not simply recall the information they provided about the self and compare it to the information they are provided with as part of feedback. This is easier to account for, for the Barnum effect compared to choice blindness, since the information people see is not necessarily a direct reflection of information they provide, but is judged from some other property in a manner not overtly stated to the participant, such as personality test scoring. Not only does this often involve answering many questions, the responses to which would be difficult to recall (Miller, 1956), but also without knowing how those responses are transformed to individual descriptions an individual cannot have a precise representation of the outcome expected. For example, in judging the accuracy of an astrology generated personality, one would not be certain of what being a Gemini, born in the year of the snake, could possibly determine about their personality. Indeed, people presented with information from more ‘mysterious’ sources, which do not explicitly relate the questions asked to traits provided, and use projective techniques to derive individual profiles, are more likely to experience the Barnum effect (e.g., Snyder, 1974).

Johnson et al. (1985) propose that the availability heuristic, or the reliance on immediate information that comes to mind to make a decision (Kahneman & Tversky, 1973), is responsible for such an effect. They suggest that because we have a large memory store about the self, when presented with general feedback, we are likely to be able to easily identify past evidence that supports such feedback. As we have seen in section 3.1, double headed and general feedback, which could apply to almost anyone, is indeed more prone to the Barnum effect (e.g., Richards & Merrens, 1971; Wyman & Vyse, 2008). Johnson et al., (1985) present further evidence that a memory bias may be in play, by demonstrating that the Barnum effect tends to occur even when the participants are told that the feedback profile is random and not generated for them specifically. Furthermore, they are less likely to judge such feedback as accurate when applying it to someone else compared to themselves, as would be expected because we have fewer examples of other people's behaviour than our own. Additionally, the availability heuristic can explain why people tend to accept false positive feedback more readily than negative (e.g., Johnson et al., 1985; Macdonald & Standing, 2002; Poškus, 2014), as positive information about the self is more likely to be attended to, and in turn stored in memory (Sedikides, Green & Pinter, 2004). Overall, it seems that when comparing information presented, we may be using memories of past behaviours as a comparison after all, however, these memories consist of examples as opposed to some form of personal profile. Perhaps if people were more accustomed to assessing, or at least seeing, personality profiles this effect can be minimised. Indeed, Bachrach and Pattishall (1960) reported that psychiatric residents were less likely to accept false personality feedback than students (see also Greene, 1977; Greene et al., 1979), indicating that level of experience with individual assessment can help eradicate the Barnum effect.

For choice blindness, we are faced with a more perplexing question when it comes to memory; how is it we fail to keep track of a decision we made almost immediately prior to exhibiting such failure (justifying the wrong choice). Limited cognitive resources are often cited as the cause for deviation of observed decision-making behaviours, from what would be expected from choice under an unconstrained environment (e.g., Simon, 1972). It is, therefore, a reasonable suggestion that choice blindness may be a result of insufficient ability to process information or limited memory capacity. In fact, putting environmental constraints on the ability to process information, such as stringent time limits for example (Johansson et al., 2005), has indeed been demonstrated to reduce the proportion of decisions that result in detection of the incorrect outcome being presented. However, whether the decision is limited at the stage of processing alternatives or encoding evidence, memory for the alternatives post choice blindness should demonstrate some impairment if we fail to create an accurate memory representation, which does not appear to be the case (Pärnamets et al., 2015; Sagana et al., 2014a). Research has found that recognition memory for facial stimuli used in a choice blindness task remains intact even after the participants justify their selection, regardless of whether the face presented during feedback was manipulated (Pärnamets et al., 2015). The effect has also been demonstrated with the evaluative judgement variation of the choice blindness paradigm (Sagana et al., 2014a), where participants' memory for previously provided sympathy ratings for female faces did not differ significantly between trials for which participants received altered feedback and those for which they received real feedback. The authors conclude that memory limitations cannot, at least fully, be responsible for the choice blindness phenomenon.

It would appear that recognition memory cannot account for the choice blindness phenomenon, yet the memory of the actual decision outcome must be impaired for choice blindness to occur. Pärnamets et al., (2015) propose that the memory of the decision outcome is altered whilst recognition memory remains intact, because source memory is underpinned by evaluative and reconstructive processes (Johnson, Hashtroudi & Lindsay, 1993; Yonelinas, 1999). Whereas when testing recognition memory, it is sufficient to compare the nature of the perceptual experienced in the recent past with the currently presented alternative, when accessing the source memory, the perceptual information is accompanied by a number of judgements (e.g., 'Does this seem plausible given other things that I know' – Johnson et al., 1993). As a result, environmental factors, such as the incongruent feedback presented in choice blindness can lead to a distorted memory of how the information was used, whilst the recognition of the information remains intact.

A judgement process of assessing the likelihood that the feedback represents the decision made would also allow the ambiguity of the task at hand to influence our acceptance of invalid information, a tendency often noted in false feedback acceptance literature (see section 3.1 for discussion). Recall, that both the Barnum effect and choice blindness are influenced by whether the feedback can be interpreted in different ways, whether due to similarity of the information presented to the real outcome (e.g., Andersen & Nordvik, 2002; Steenfeldt-Kristensen & Thornton, 2013), generality of information (Sundberg, 1955; Merckelbach et al., 2011), or procedural parameters such as time constraints (Johansson et al., 2005). Now let's consider the possibility that on encountering false feedback we recall the situation under which we made the decision fairly accurately, with all the relevant parameters, and have to evaluate the plausibility of the outcome based on those parameters. If the situation is ambiguous, we are more likely to find supporting evidence for the information presented as it would make an alternative outcome plausible if the task was to be repeated, therefore leading to a conclusion that the presented outcome could have been correct. As you can see, a judgement process certainly seems fitting for the description of processes underlying false feedback acceptance.

Pärnamets et al. (2015) further stipulate that since choice blindness appears to be a result of using environmental information to judge our own behaviour, the phenomenon is best described by self-perception theory (Bem, 1967). According to this theory, individuals come to know their own attitudes by observing their own behaviour and the circumstances in which it occurs, and therefore: "*to the extent that internal cues are weak, ambiguous or uninterpretable, the individual is functionally in the same position as an outside observer*." –Bem, 1967, p2. Both the Barnum effect and choice blindness paradigms appear to be the perfect illustration of self-perception theory. Not only do they demonstrate that people will use external cues above their previously indicated responses to infer their own characteristics and preferences (e.g., Forer, 1949; Johansson et al., 2005), but research has also demonstrated that accepting false feedback can also affect future behaviour (Halperin & Snyder, 1979; Johansson et al., 2014; Merckelbach et al., 2011; Sakamoto et al., 2000), indicating that instead of accessing a stable representation of the self, we think back to the responses we made and use these to infer our attitudes, just as we would for someone else.

The main problem for the application of self-perception theory arises when one considers what 'weakness of internal cues' entails, and how this can distinguish between individuals and situation more predisposed to false feedback acceptance. Bem (1967) proposed that there are, in fact, a number of differences in self-perception and interpersonal perception. First, the theory does recognise that we can discriminate between some internal stimuli (e.g., the amount of effort exerted), however, poses that such discrimination is very limited. Second, when making attributions to the self as opposed to another, one has vast amounts of knowledge about past behaviours which could inform the attitude in question, alongside the current overt behaviour. And lastly, Bem (1967) recognises that personal motivations, such as seeking to protect self-esteem, are likely to influence how we interpret our own behaviour.

Somerville and McGowan (2016) help clarify when external or internal cues may be used to make a judgement, though the Discovered Preference hypothesis (Plott, 1966) which poses that when people are faced with a new decision, in a new environment, they encounter uncertainty of which action would be in their best interest, which results in a strong element of randomness, and reliance on external cues. However, as people gain experience with the decision in question by repeating the choice process and receive consequential feedback their behaviour will evolve and knowledge of preference stabilise. In other words, as randomness plays a large part when making unfamiliar choices, detection of incongruent outcomes under such circumstances would be low (high choice blindness), whereas for choices that we have previously encountered on multiple occasions we will be familiar with our preference and detection will be high (low choice blindness) – exactly the pattern of behaviour described by Somerville and McGowan (2016).

Whilst the application of discovered preference hypothesis appears congruent with the self-perception theory account of false feedback acceptance, as familiarity could explain internal cues becoming strong enough to be accurately recognised, the pattern of experience reducing false feedback acceptance can also result from building up a sufficient wealth of information about our past behaviours to detect that the outcome is out of character, equivalent to the constructed preference approach (Lichtenstein & Slovic, 2006). The inability to distinguish which theory is better suited to describing the mechanisms underlying choice blindness, is a weakness of the poorly defined 'internal cues' described in self-perception theory.

Similarly, the pattern of positive feedback being more readily accepted in the literature on Barnum effect can be explained using two separate sources of internal information differentiated by Bem (1967). The first would be the heightened ability to recall positive information (as discussed with respect to availability heuristic) resulting in increased amount of knowledge about socially positive attitudes which can be used to judge the situation, thus making the negative feedback more likely to appear incongruent leading to its rejected. The second would be the direct effect of cognitions guided by personal motivation, such as the self-serving bias (Heider, 1958), or the tendency to attribute positive events to the self and negative events to external factors (i.e., the profile was positive because of my traits, versus the profile was negative because of the poor instrument design). Macdonald and Standing (2002), argue that self-serving bias is the main, if not the only explanation required for the Barnum effect. Yet other evidence suggesting that the Barnum effect can be induced for negative feedback (e.g., Dmitruk et al., 1973; Poškus, 2014; Wyman & Vyse, 2008) indicates this should not be considered in isolation. Regardless, whether the tendency to rate positive feedback as accurate is a result of a memory bias, or motivated cognition that takes place in the moment, self-perception theory does not help us discriminate between the two.

Generally, self-perception theory is capable of accommodating most effects observed in false feedback acceptance literature. However, at this stage it is too broad and general to provide us with a precise prediction of when internal cues are weak or strong, and fails to provide a description of what happens when we judge our behaviour using external cues, or 'self-perceive'. This is not to say the information provided is not useful, simply that we need more if we are to gain full understanding of when and why false feedback is accepted.

The predictions made by self-perception theory can also be explained by cognitive dissonance (Festinger, 1957). Cognitive dissonance theory states when there is an inconsistency between cognitions, such as beliefs, attitudes or behaviours, people experience mental discomfort, or dissonance, which results in one or more of the cognitions to be adjusted in order to establish consistency. In the context of accepting invalid feedback, cognitive dissonance theory would predict that when we receive feedback about our own characteristics or choices that is inconsistent with the earlier expressed responses, we would experience mental discomfort and either change our belief about the earlier choice or reject the feedback itself, depending on which of the two elements receives more support from either internal or external cues.

It seems that cognitive dissonance provides a plausible alternative to self-perception in the explanation of what leads to acceptance of invalid feedback, through the introduction of a new cognitive element – dissonance, or mental discomfort. To my knowledge, there has been no research which allows us to identify whether discomfort is experienced for participants who are presented with invalid information about the self in the Barnum effect. For choice blindness, neither the analysis of people's justification for erroneous feedback (Johansson et al., 2006), nor the decision confidence post feedback manipulation (Johansson et al., 2005) reveal any sign of introduced discomfort or uncertainty. Although this is not necessarily unusual, since neither of the measures were designed to capture mental discomfort experienced by the participants as a result of the mismatch between choice and outcome. Nonetheless, cognitive dissonance introduces a new mental variable that to this date is yet to be established, making the self-perception approach the more parsimonious one.

Overall it appears that some of the approaches used by researchers to explain choice blindness date back as far as half a century (e.g., Bem, 1967), yet these approaches are still largely under debate with no unified theory established to how the different explanations fit together (e.g., Gawronski & Strack, 2004). However, as long as the theories are in agreement as to the description of how choice blindness should behave they should all be treated as possible contenders for the mechanism underlying choice blindness. Self-perception and cognitive dissonance theories both predict that when faced with inaccurate choice feedback, either we will misremember our past choice or reject the current feedback, with an increase in rejection for choices that have been repeated many times previously. Similarly, whether choices are discovered or constructed, we expect them to be less stable (thus susceptible to choice blindness) before we become familiar with these choices and rely more on the environmental conditions in which they are presented, which in turn can affect choice blindness by creating more or less ambiguous conditions. Sadly, the lack of a single coherent theory of cognition still prevails across all of psychology, but hopefully by describing how various behaviours manifest, we are taking small steps to deciphering what exactly is going on in our brains when we fail to detect the errors in information concerning our own traits and preferences. My own research, presented in the following chapters, attempts to contribute to such a description by establishing the circumstances under which we are likely to accept invalid feedback (chapters 5 through 7), and whether this results in updating our beliefs about our own behavioural propensities (chapter 4).

Research Papers

Chapter 4.

The Barnum effect and its consequences: can bogus feedback change behaviour.

(Paper I)

Mariya Kirichek – *Warwick Business School*

Nick Chater– *Warwick Business School*

4.1 Abstract

Past research has demonstrated that accepting bogus feedback about the self, known as the 'Barnum' effect, can affect human behaviour (Halperin & Snyder, 1979; Sakamoto et al., 2000). However, this research has failed to compare the effects of the bogus feedback to that of real feedback, making it impossible to conclude whether different types of feedback equally contribute to altering behaviour. Furthermore, the methodology has been limited to using written prose feedback, delivered to female only samples, posing potential problems for the generality of the research. In the current paper, we use validated personality and risk attitude measures to revisit this research area over two experiments; looking at how altering personality feedback can affect willingness to volunteer for psychology experiments and how risk attitude feedback can affect risky choices respectively. Both experiments partially elicit the Barnum effect, with feedback altered to suggest personality associated with high likelihood to volunteer for experiments, and lowered financial risk preference rated as of similar accuracy to real feedback. The type of feedback presented did not impact the responses provided on the subsequent measures, with data consistent with the null hypothesis, suggesting false personal feedback does not impact subsequent behaviour. The possible reason for the findings differing from previously reported results are discussed.

4.2 Introduction

The ‘Barnum effect’ refers to the tendency to accept bogus feedback about the self, whether it is applicable to everyone or simply untrue, because it is supposedly derived from personality assessment procedures (Furnham, 1989; Meehl, 1956). Forer (1949) was the first to demonstrate this phenomenon, by administering a personality test to 39 students, and asking them to indicate how accurately they believed the profile generated from the personality measure described them. Forer (1949) pretended to score the tests himself, however, instead he copied a personality description from an astrology column of a newspaper and gave it to his participants as their personalised descriptions. The average accuracy score provided by the students was of 4.3 out of 5, suggesting high perceived accuracy of the statements provided.

Although initially known as the fallacy of personal validation and the Forer effect, the term ‘Barnum effect’ was later popularised by Meehl (1956), so named after P.T. Barnum – a circus entertainer with the catch phrase ‘we have something for everybody’. This phenomenon has since been well documented, with a lot of attention being given to why the effect occurs and how it varies across different situations (see Dickson & Kelly, 1985; Furnham & Schofield, 1987; Snyder et al., 1977). Although many researchers warn of possible negative effects associated with the Barnum effect (e.g., Dickson & Kelly, 1985; Furnham, 1989; Furnham & Schofield, 1987; Snyder et al., 1977) the literature has been predominantly concerned with the issues the phenomenon poses for psychometric measure validation, whilst the consequences of accepting bogus personality feedback on the individual have seldom been researched. In particular, might people start to behave in a way that assimilates with the false feedback? That is, if people fall for the Barnum effect, could this shape their later behaviour?

To date we were able to identify two studies looking into the impact of Barnum effect on subsequent behaviour; one examining the effects of bogus feedback suggesting high susceptibility to phobia treatment on response to the treatment of fear of snakes (Halperin & Snyder, 1979), and the other looking at effects of extraversion feedback on stranger interaction (Sakamoto et al., 2000). Both studies successfully demonstrate that accepting bogus feedback can alter subsequent behaviours in the direction associated with the traits suggested by the feedback. This finding carries important implications for feedback use in clinical treatment, as well as non-clinical behavioural change (Halperin & Snyder, 1979), and can help us understand how external information shapes human behaviour (e.g., Self-Perception theory, Bem, 1967). It is, however, surprising that despite the broad scope of domains to which the Barnum effect can be applied, only two such studies are reported. Furthermore, both studies suffer from limitations that complicate outcome interpretations and question the validity of results. Accordingly, we will present these limitations for discussion and introduce two new experiments designed to test Barnum effect consequence in two oft-cited domains in psychometric and behavioural literature – personality measures and risk behaviours respectively.

The first study by Halperin and Snyder (1979) reported that presenting female participants with bogus, supposedly personalised, feedback suggesting high potential to change leads to enhanced effects of clinical treatment for fear of snakes. The feedback used in this study was a hand-written personality assessment, which suggested that a previously taken personality test revealed that the participants' personality was well suited to change. Participant's responsiveness to treatment was then compared between the group that received the aforementioned feedback, and a group that received no feedback at all. A positive effect of feedback was reported for both self-report (Snake Fear Questionnaire; Klorman, Weerts, Hastings, Melamed & Lang, 1974) and behavioural (Behavioural Avoidance test; Nawas, 1971) outcomes.

The second study by Sakamoto et al. (2000) investigated the effects of bogus extraversion and introversion feedback on female participants' self-image and nature of their interaction with a confederate. The study used four questionnaire formats to assess personality; 'academic' personality tests and 'popular' personality tests, in multiple choice and open ended question versions. Bogus feedback was designed for each of the questionnaires, either suggesting the participants were highly extraverted or highly introverted. The authors conclude that giving participants bogus extraversion feedback results in a higher self-image, increase in conversation with a confederate, better impression of the confederate by the participant, as well as of the participant by the confederate, compared to participants who received the bogus introversion feedback.

Let us first discuss the validity of the conclusions provided by each experiment individually. Halperin and Snyder (1979) demonstrate ecologically valid results that are consistent throughout the behavioural and self-report measures used. However, as the authors themselves note the lack of an appropriate control group (a group that receives neutral, or opposite feedback) means the observed effects may be a result of receiving any diagnostic feedback, or even simply the extra time commitment to the experiment. Although it is argued that in a clinical setting providing negative feedback is not ethical, the use of a neutral group that is provided with real or neutral feedback may provide an appropriate solution for future research. Alternatively, we can extrapolate from the bi-directional feedback effects observed by Sakamoto et al. (2000), where introversion and extraversion feedback result in behaviour associated with each trait respectively. Yet, this approach also fails to provide a definitive conclusion, as without a neutral control group we cannot determine whether extraversion and introversion feedback both lead to behavioural change or whether one simply maintains the behaviour at a level similar to neutral feedback, or no feedback at all.

The research by Sakamoto et al. (2000) also provides evidence that bogus feedback can impact behaviour related to the feedback across a range of dependent measures, however, these measures comprised only a third of the outcome variables reported, with the remaining eight measures showing no significant effects of feedback (sitting distance, time for conversation, period of eye contacts, acceptability of result, pleasure in test, relaxedness, comfortableness, positive feelings). This complicates the interpretation of results since no explanation is provided for why bogus feedback may impact some, but not other, behaviours hypothesised to be related to the feedback trait. Furthermore, the effects observed also appear to be inconsistent across the different questionnaire types used, for example when participants believed they were receiving feedback for an open-ended popular personality test the direction of the effects observed was the opposite to that concluded for the impression of the person by confederate as well as time spent talking to a stranger, despite the authors' observation that popular tests are better at eliciting feedback effects. Overall it appears that despite the proposed conclusion that providing people with feedback that suggests they are extraverted makes them behave in a manner more akin to that of an extraverted person, we must be cautious with regards to accepting the validity of the results at face value.

Aside from the limitations specific to each of the methodologies used, the aforementioned research shares a number of traits that may preclude us from reaching a valid conclusion. For instance, both studies used specifically designed written feedback, delivered to the participants by the experimenter. This may run a risk of providing participants with additional information about the experimenters' expectations and thus increase demand characteristics, especially in the study by Sakamoto et al. (2000) where the experimenter read the feedback to the participants face to face. Although this may still be considered an effect of the bogus feedback, its presence can only be concluded for participants' behaviour when they know they are being observed by the people who provided the feedback in the first place. The use of written prose is also inconsistent with personality assessments commonly encountered in academic or organisational settings (e.g., Myer's-Briggs Type indicator – Myers, 1962; Myers, McCaulley, Quenk & Hammer, 1998; International Personality Item Pool (IPIP) – Goldberg et al., 2006), which organise individuals along trait scales creating directly interpretable numeric feedback. Since many individuals may be aware of the common practice of using numeric personality descriptors, they may have been suspicious of the vague format and not experienced the Barnum effect, which cannot be directly established since feedback accuracy was not measured in either of the experiments.

It is important to note that the lack of establishing Barnum effect, or measuring the accuracy of the feedback provided, can in itself complicate the interpretation of results. For example, without the ability to demonstrate that participants accepted the feedback observed, there remains a possibility that the behavioural changes observed by Halperin and Snyder (1979) and Sakamoto et al. (2000) does not require the Barnum effect at all, and are a result of a more direct associative process, such as priming (e.g., Herr, 1986). On the other hand, if the post feedback effects observed are dependent on believing the feedback received, there is a possibility that there is a limit to the types of behavioural change that can be achieved using this methodology. For instance, recent research has shown that although people are likely to accept positive feedback about the self, they tend to reject negative feedback (Macdonald & Standing, 2002), thus suggesting that the Barnum effect can only exist for positive feedback and in turn can only be used to achieve behaviours associated with socially positive traits.

Another concern with the research that demonstrated the behavioural consequences of the Barnum effect is that both experiments only used female samples, thus providing uncertainty as to whether the results can be generalised to males also. This presents a concern as females tend to accept feedback more easily than males (Layne, 1998). In addition, the systematic personality differences (higher on neuroticism, extraversion, agreeableness, and conscientiousness – Schmitt, Realo, Voracek & Allik, 2008), and higher sensitivity to the experimenter effect (Deaux, 1985) in females compared to males could confound the experiment, making it difficult to pinpoint the role gender may have played in the experiments.

Overall, it appears that research into the consequences of the Barnum effect indicates that accepting false feedback can subsequently affect human behaviour. However, this research is limited to only two domains, does not establish whether the Barnum effect actually occurs, lacks appropriate control groups and uses a very limited sample making the interpretations of the results limited at best. Accordingly, here we present two experiments that explore whether providing false feedback impacts subsequent responses perceived to be related to that feedback. In order to make sure our results are widely generalizable we picked domains that dominate psychometric measurement and behavioural change fields respectively – personality assessment and risk taking behaviours, in order to test the impact false feedback can have on individuals.

4.3 Experiment 1

The main aim of experiment one was to establish whether altering participants' personality scores in the direction perceived to be associated with high or low likelihood to volunteer for psychology experiments would impact self-reported likelihood to volunteer for experiments in the future.

Personality assessment was selected as the independent variable because of its prominence across a range of disciplines, including academic, clinical and personnel recruitment. Specifically, the International Personality Item Pool (Goldberg et al., 2006; IPIP, 2006) measure of the Big Five personality factors was selected as the measure of choice because it is publicly available (IPIP, 2006), widely used, and is organised according to one of the most widely known personality taxonomies (Gow, Whiteman, Pattie & Deary, 2005).

Self-reported likelihood to volunteer for psychological experiments was selected as the dependent variable because the measure is perceived to be easily verifiable post experiment, providing participants with an incentive to answer as honestly as possible. To ensure the bogus feedback provided to participants was associated with either low or high likelihood to volunteer, a pilot study was conducted that established how people perceived volunteering for psychology experiments to relate to personality traits.

Although the Big Five approach to personality is one of the most recognised psychometric tools, there are a number of concerns in applying it to the Barnum effect and behavioural change. First, the use of multiple traits makes it difficult to control for the individuals' personality due to the covariance of the traits, willingness to accept false feedback and susceptibility to the aforementioned feedback. Additionally, the introduction of all five traits as behavioural predictors poses a problem for the power of the experiment. Although we ensure that the base personalities of the participants do not differ between conditions, and exclude the traits from further analysis, this is not ideal but cannot be overcome due to lack of a single trait measure that can comprehensively capture a personality profile.

Second, when using the Big Five trait measures the elicitation of the Barnum effect has been shown to be limited, with participants rating personality profiles as less accurate the more they differ from their real profile (Andersen & Nordvik, 2002). It has also been shown that there is a high likelihood that a profile will be rejected if the personality profile used to describe the participant is perceived a negative (Macdonald & Standing, 2002). Although the possibility that no participants will accept the Bogus feedback is certainly a concern, as long as a proportion of participants rate the bogus feedback as more accurate than chance, the experiment allows us to establish whether the Barnum effect can impact subsequent self-report willingness to volunteer for psychology experiments (by examining the interaction effects of feedback type with feedback accuracy).

Overall, the study took on an independent measures design, with *type of bogus feedback* (consistent with high likelihood to volunteer, consistent with low likelihood to volunteer, and real) and participants' *accuracy rating* of the seen feedback (rating on a Likert scale of 0 to 10) as independent variables, and *likelihood to volunteer* across four different types of experiment (online, via phone, face to face with experimenter, group setting) as the dependent variable. The focal question of the experiment was whether believing altered feedback can impact subsequent self-report likelihood to volunteer.

4.3.1 Method

Participants

One hundred and seventy-nine participants (86 female) were recruited online using the Prolific academic online recruitment platform with a mean age of 29.40 (SD=10.12). All participants were required to satisfy age (between 18 and 80 years old) and language requirements (English as a first language) in order to take part in the experiment.

Materials

Personality. The self-report personality measure used was the short version (50-items) of the International Personality Item Pool (Goldberg et al., 2006; IPIP, 2006), a publicly available measure of the Big Five personality factors – Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The measure uses 10 questions to measure each trait on a ten point Likert Scale, with half of the items negatively scored. The measure is widely used and has good internal consistency and correlations with the oft-cited Costa and McCrae's (1992) NEO-FFI measure (Gow et al., 2005).

Personality feedback was created based on the personality traits people think of when asked to describe a person that is either very likely to volunteer for psychology experiments, or very unlikely to volunteer for psychology experiments depending on condition. A volunteer sample of 22 people was asked to imagine a person most likely to volunteer for psychology experiments, and rate this person's Big Five traits on a scale of one to five. One indicating very low on that trait, and five indicating very high on the trait. The responses indicated that people perceived high trait Openness, Conscientiousness, Extraversion, and Agreeableness and low trait Neuroticism to be associated with people likely to volunteer for psychology experiments.

Accordingly, Openness, Conscientiousness, Extraversion, and Agreeableness scores were transformed onto a scale of 51 to 100, and Neuroticism scores onto a scale of 1 to 50, for feedback associated with high likelihood to volunteer. For simplicity purposes, this will be referred to as enhanced feedback. For feedback associated with low likelihood to volunteer for psychology experiments participants' Openness, Conscientiousness, Extraversion, and Agreeableness scores were transformed onto a scale of 1 to 50, and Neuroticism scores onto a scale of 51 to 100. This will be referred to as reduced feedback. For the control feedback condition participants were shown their actual scores, referred to as real.

The feedback was provided alongside a question asking 'How accurately would you say this describes you' for each trait. The answers were given on a Likert scale of 0 to 10, ranging from 'Completely Accurate' to 'Completely Inaccurate'.

Volunteering Questions. The Volunteering questions were selected to cover a wide range of psychology research study types. Intrusive research, such as medical interventions and physiological measures, was not included to avoid privacy, ethics and health concerns, as well as to make it seem realistic that participants might be approached to volunteer at a later date. Four research types were selected; online research, over the phone research, face to face research and group research. For the latter two categories, participants were told the research would take place varying locations and they would be able to pick the location most convenient for them. Participants were required to report how likely they would be to volunteer for each type of experiment on a scale of 0 to 10, from 'Definitely Not' to Definitely Yes'.

Procedure

The study was presented using the Qualtrics Software platform. The survey consisted of demographic questions, the IPIP NEO 50 personality scale, followed by feedback accuracy ratings and volunteering questions. Within each questionnaire, all questions were randomised to avoid order effects. The participants were required to provide informed consent before starting the study, and were provided with debrief information as well as contact details of the researcher in case of any concerns at the end.

4.3.2 Results

Fifteen (of 179) participants were excluded from the analysis because participant scored below 5 or above 95 on one or more personality traits, which would result in no difference in feedback between reduced and real (floor effect) or enhanced and real (ceiling effect) scores respectively.

IPIP personality scores. The average trait scores were 66.85 (SD=13.66) for Openness, 60.93 (SD=14.77) for Conscientiousness, 45.65 (SD=18.06) for Extraversion, 65.34 (SD=13.24) for Agreeableness and 46.19 (SD=18.59) for Neuroticism. Multivariate GLM analysis revealed no significant difference in personality profiles between different conditions (F (10, 316) =.955, p=.483)

Barnum effect. The average accuracy rating for the personality feedback was 6.21 (SD=1.73). The accuracy scores were 6.99 (SD=1.47) for enhanced, 5.18 (SD=1.48) for reduced and 6.73 (SD=1.88) for real feedback conditions. The difference between conditions was significant (F (2,161) =24.630, p<.001). Pairwise comparisons revealed that the accuracy ratings in the reduced feedback condition were significantly lower than the accuracy ratings for the real feedback condition (p<.001) and enhanced feedback condition (p=.001), there was no significant difference between the real and enhance feedback accuracy scores. The scores were significantly higher than the mid value (neither accurate, nor inaccurate) of 5 (p<.05) for enhanced and real feedback, but not for reduced feedback (t (64) =1.019, p=.312).

No effects of gender were detected for feedback accuracy, overall (male M=6.16, female M=6.27*;* F (1, 162) =.194, p>.1), or by feedback type (p>.1).

Feedback Effects. The average likelihood to volunteer (across all experiment types) was 5.16 (SD=2.45) overall, 4.96 (SD=2.64) for enhanced, 5.00 (SD=2.42) for reduced, and 5.70 (SD=2.18) for real feedback conditions.

A Multivariate GLM analysis was carried out using the likelihood to volunteer for the four separate experiment types (online, phone, face to face and group) as dependent variables, and feedback condition (enhanced, reduced or real) as a predictor variable. Feedback acceptance was not included in the analysis due to its strong relationship with condition, which would have resulted in collinearity. There was no significant effect of condition (F (2, 158) =1.679, p=.190) overall, or for any of the likelihood to volunteer measures in isolation (p>.05)

The analysis was further repeated using a Bayesian linear regression approach, to establish whether a model using condition as a predictor variable was more likely than the null hypothesis. A BF_{01} of 5.933 was observed, suggesting that these data are 5.933 times more likely to be observed under the null hypothesis[1].

No effects of gender were detected for feedback accuracy, overall (male M=5.10, female M=5.24; F (1, 162) =.125, p>.5), or by feedback type (p>.5).

4.3.3 Experiment 1 Summary

The results of experiment one suggest that the Barnum effect can be successfully induced, with false feedback rated as highly accurate by the participants, however, this only occurs for enhanced feedback, whereas accuracy of reduced feedback is rated as neither accurate nor inaccurate (midpoint value). No subsequent effects of feedback type presented on self-reported willingness to volunteer for psychology experiments was detected, even when controlling for the perceived accuracy of the feedback. In fact, the differences in willingness to volunteer between the people who saw different feedback types were so small, that it is more likely that feedback type had no effect, than that it did.

Whilst it is apparent that no effect of condition can be detected on subsequent willingness to volunteer in the current sample, there are a number of limitations which may have contributed to this. First, the feedback augmentation method was based on real personality scores which varied between individuals, in turn resulting in varied feedback scores. This variation made it difficult to control for similarity of feedback without introducing participants' real traits as a confounding variable, and could have precluded the effect that may have been apparent if fixed feedback was used. However, this method allowed to control for personality and ensure that the direction in which the feedback changes personality was indeed inaccurate.

[1] Both the GLM and Bayesian Regression analysis were validated with analysis controlling for participants' personality scores, and the same pattern of results was observed. There was no significant effects of condition. Neuroticism (F (4,150) =2.938, p=.023) and Extraversion (F (4,150) =3.062, p=.018) were significant predictors of self-reported likelihood to volunteer, whereas effects of Openness, Conscientiousness and Agreeableness were not significant (p>.05). Bayesian analysis revealed that the data were more likely to be observed under a model only using personality traits as predictors, compared to a model including feedback condition and accuracy ratings (BF_{01}=2.557).

In piloting this study fixed numbers were initially used, revealing that extreme profiles (e.g., 100 or 0 on all traits) were not believable, and moderate profiles (75 or 25 across traits) often augmented feedback in the direction opposite to the one desired. Another problem arises when we consider that personality has often been found to impact the propensity to accept invalid feedback as well as volunteering behaviour, making it difficult to tear apart the effects it may directly exert on self-reported likelihood to volunteer, from the effect mediated by the feedback acceptance or any interactions of the two. With five variables being manipulated it is exceedingly difficult to control for the effect of original and presented personality, as there are multiple correlation and interactions that may be taking place and precluding a real effect. Accordingly, in experiment two, we only manipulate one trait which is directly relevant to the subsequent behaviour measured, which would allow us to better envision an interaction between real feedback and the direction in which it is augmented.

4.4 Experiment 2

The focal aim of experiment two was to establish whether altering participants' financial risk attitude scores in either the risk seeking or risk averse direction would impact subsequent preferences for risky versus certain monetary lotteries.

Risky choice was selected as the domain of interest given its prominence in the behavioural science literature. Although risky choice is often perceived as an economics domain with unique cognitive mechanisms, it has been proposed that risky choice should be treated the same as other types of choice and thus should be susceptible to general mechanisms such as feedback effects (Chater, Johansson & Hall, 2011).

The Domain Specific Risk Taking scale (DOSPERT, Blais & Weber, 2006) was used to measure risk attitudes, as it is a validated measure with good predictive power of risk taking behaviour (Harrison, Young, Butow, Salkeld & Solomon, 2005). The subsequent risk taking behaviour was assessed through choices between certain and risky lotteries designed by Lauriola, Levin and Hart (2007).

The DOSPERT scare measures ethical, financial, health and safety, recreational and social risk attitudes. In the experimental manipulation, bogus feedback was created for the financial risk preference trait only. This is because people who exhibiting high levels of risk taking in one content area can be quite risk averse in other risky domains (Hanoch, Johnson & Wilke, 2006), and financial risk attitude is the most relevant to the lottery task used as the dependent variable in the current experiment. In addition, this eliminates some of the covariance and power concerns discussed in experiment one. Two types of altered financial risk attitude feedback were created – enhanced and reduced. Real feedback was used as a control group.

Overall, the study took on an independent measures design, with type of bogus *financial risk preference feedback* (enhanced, reduced, and real) and participants' accuracy rating of the *feedback* (rating on a Likert scale of 0 to 10) as independent variables, and the financial risk preference as exhibited in *risky choice* (proportion of trials on which the risky choice was preferred over the certain choice) as the dependent variable. Based on past literature, it is hypothesised that participants will be more likely to select risky choices if they accept feedback that they are financially risk seeking, and less likely to select risky choices if they accept feedback that they are financially risk averse.

4.4.1 Method

Participants

One hundred and eighty-four participants (71 female), with the average age of 29.56 (SD=8.92) took part in the study.

Materials

Risk Attitudes. Risk attitudes were measured using the DOSPERT scale (Blais & Weber, 2006). The scale was designed to measure the likelihood with which the participants would partake in risky activities across five domains; ethical, financial, health and safety, recreational and social. The measure is comprised of 30 items, with 6 positively scored items per domain. The answers are rated on a 7-point Likert scale ranging from 1 (*Extremely Unlikely*) to 7 (*Extremely Likely*).

Feedback. Feedback was presented as a number between 1 and 100, for each domain. For ethical, health and safety, recreational and social domains, participants saw the real scores generated by the DOSPERT scale. For financial feedback participants saw their scores transformed onto a scale of 51 to 100 for the enhanced condition, 1 to 50 for the reduced condition and real feedback in the control condition. Participants were asked to indicate how accurately they think the feedback on each trait describes them on a Likert scale of 0 to 10 ranging from Completely Accurate to Completely Inaccurate. Participants were allocated randomly to each condition.

Risky Choice. A subset of lotteries designed by Lauriola et al. (2007) was used to examine behaviour in a risky decision making task. In the task, participants are told to imagine that they are presented with two contracts, one of which they have to agree to sign and the other reject. One of the contracts offers a sure thing amount, whereas the other offers an uncertain amount proportionate to the riskiness involved.

The original measure constituted a 2 x 5 x 6 independent measures factorial design. Half of the trials involve choosing between gains, and the other half between losses. Five levels of probability were used for the uncertain choice (0.02, 0.25, 0.50, 0.75 and 0.98), and the sure thing amount was set between 50 cents and $50000 in logarithmic steps of value.

In this study, the contracts were presented in pounds instead of dollars to suit local currency. The 50000 value was excluded from the measure, due to the time-consuming nature of the task identified by volunteers in the piloting phase, leaving a 2 x 5 x 5 design (i.e., with six levels).

The final measure was comprised of 50 items, varying across the three described levels of factors that may have a possible effect on risky decision making.

Procedure

The experiment was carried out online. The materials were combined with questions about age, gender and education, and a survey was created using the Qualtrics software platform. Questions within each questionnaire, as well as the risky and non-risky lottery options were randomised to avoid order effects.

Participants were recruited and paid through the Prolific Academic crowdsourcing platform. They were required to enter a valid Prolific Academic ID at the beginning and to click on a link at the end of the study to confirm study completion. Only participants with English as their first language, and between the ages of 18 and 80 were allowed to take part, to ensure full understanding.

4.4.2 Results

10 (out of 184) participants were excluded from the analysis because participant scored below or above 5 on the financial risk attitude score on the DOSPERT scale, which would result in no difference in feedback between reduced and real (floor effect) or enhanced and real (ceiling effect) scores respectively.

Financial Risk Attitude. The mean score for financial risk attitude on the DOSPERT scale was 33.94 (SD=17.30), which did not differ significantly between feedback conditions (F (2, 171) =.076, p=.927).

Barnum effect. The average accuracy rating for the Financial Risk preference feedback was 6.75 (SD=2.20). The accuracy scores were 5.98 (SD=2.35) for enhanced, 7.03 (SD=2.25) for reduced and 7.05 (SD=1.88) for real feedback conditions. A GLM analysis using financial risk preference scores, feedback condition and an interaction of the two as independent variables, and perceived accuracy as a dependent variable (F (5, 168) =6.583, p<.001), revealed that financial risk preference was not a significant predictor of accuracy ratings (F (1, 172) =.125, p=.724), whereas condition (F (2, 171) =15.290 p<.001) and the interaction of condition and financial risk preferences (F (2, 171) =11.512, p<.001) were significant. Pairwise comparisons of accuracy ratings between feedback condition revealed that the accuracy ratings in the enhanced feedback condition were significantly lower than the accuracy ratings for the real feedback condition (p=.036) and reduced feedback condition (p=.034), all other comparisons were not significant. The scores in each condition were significantly higher than the mid value (neither accurate nor inaccurate) of 5 (p<.05). A Pearson's correlation analysis further demonstrated that when participants saw enhanced feedback, financial risk preference was positively correlated with accuracy ratings (r=.405, n=47, p=.005) and when they say reduced feedback financial risk preference was negatively correlated with accuracy ratings (r=-.406, n=68, p=.001). There was no relationship between the two measures when participants received real feedback (r=-.073, n=59, p>.05). This effect is most likely a result of change in similarity between the altered and real feedback, with feedback similar to real preferences more likely to be accepted.

No effects of gender were detected for feedback accuracy (male M=6.72; female M=6.76; not disclosed M=10), overall (F (2,172) =1.103, p>.1) or by feedback type (p>.1).

Feedback Effects. The average risk preference (the proportion of lotteries on which participants preferred the riskier choice) was .255 (SD=.223) overall, with .236 (SD=.237) enhanced, 0.247 (SD=.203) for reduced, and .280 (SD=.235) for real feedback conditions

A GLM analysis was carried out using participants' risk preference as the dependent variable, and feedback condition and the real financial risk attitude scores. Accuracy ratings for financial risk attitude score were not included in the model due to the observed collinearity with condition. The model was significant overall (F (5, 168) =4.097, p=.002; R^2=.109), with a significant main effect of financial risk attitudes (F (1, 168) =18.798, p<.001). Neither the effect of condition (F (2, 168) =.833, p=.437), nor interaction of condition and financial risk attitudes (F (1, 168) =.569, p=.567) were significant.

The analysis was further repeated using a Bayesian linear regression approach to establish whether the type of feedback received can predict risk preference scores with a better likelihood than the financial risk attitude scores alone. The data was more likely to be observed under model including financial attitude scores only (BF_{01}=4.95) compared to a model including financial risk attitude scores and condition as predictor variables.

No significant effect of gender on the proportion of risky choices made was detected (male M=.27; female M=.24; Not Disclosed M=.30), overall ($F(2,171) = .363$, p=.696) or by feedback type (p>.05).

4.4.3 Experiment 2 Summary

The results of experiment two suggest that the Barnum effect can be successfully induced for risk attitudes, with false feedback rated as highly accurate by the participants. This is only the case however, for reduced risk preference, whereas enhanced risk preference feedback is rated as neither accurate nor inaccurate (midpoint value). No subsequent effects of feedback type presented on preferences for hypothetical certain versus risky lotteries were detected. The differences in choices made between the people who saw different feedback types were so small, that it is more likely that feedback type had no effect than that it did.

Experiment two addresses some of the concerns observed in experiment one, such as the inability to control for the role of personality traits in feedback acceptance due to multiple traits being altered. For instance, by only varying one of the traits reported, we are able to assess whether the variation in real financial risk preferences could impact the propensity to accept feedback, which in this experiment does not appear to be the case. On the other hand, we do find that similarity of the altered feedback to real trait scores generated does appear to impact the likelihood of feedback acceptance. Whilst this makes it difficult to keep a consistent level of deviation from real personality across participants, as mentioned previously this was the optimal method identified in augmenting personality relative to the real traits, without making the feedback too extreme to be believable.

4.5 Discussion

The research presented was designed to establish whether accepting bogus feedback about the self, or the Barnum effect, has subsequent effects on participants' behaviours. Two experiments were carried out investigating how believing altered personality and financial risk preference feedback impacts self-reported willingness to volunteer for psychology experiments and risky choice respectively. Both experiments demonstrated a partial Barnum effect. For experiment one participants rated feedback consistent with low likelihood to volunteer for psychology experiments as less accurate than feedback consistent with high likelihood to volunteer or their real personality scores. For experiment two, participants rated enhanced financial risk preference feedback as less accurate, compared to reduced financial risk attitude feedback, or their real scores. Both experiments failed to demonstrate any effect of feedback type on participants' predictions of how they would behave in hypothetical scenarios. For both experiments, the data was more likely to be observed under a null hypothesis compared to the alternative hypothesis.

The findings are somewhat surprising, given that both Halperin and Snyder (1979) and Sakamoto et al., (2000) have previously found that bogus feedback can significantly alter participant behaviour. Recall that Halperin and Snyder (1979) found that giving participants feedback that suggests higher propensity to respond to therapy, resulted in higher reductions of fear of snakes following snake phobia treatment. Sakamoto et al., (2000), further reported that giving participants bogus extraversion feedback resulted in increased interactions with a stranger (confederate), and better impressions of the confederate by the participant, and of the participant by the confederate, in comparison to participants who received bogus introversion feedback. Nonetheless it is the shortcomings of these studies, such as not using a real or neutral control conditions to understand the contribution of each feedback type to Barnum effect consequences, and using a female only sample, that have lead us to re-visit the subject in the first place, and thus we propose that the likely conclusion is that there is no robust effect of believing false information about the self on subsequent behaviours.

There are a number of possible differences between the research presented here and the past methodologies, which can account for the different conclusions. Perhaps the most prominent is the failure to demonstrate a full Barnum effect in the experiments presented here, as only certain types of bogus feedback appear to be accepted in both experiment one and experiment two. However, we cannot directly compare this to the past experiments, as neither Halperin and Snyder (1979), nor Sakamoto et al. (2000) actually measure the feedback acceptance in their experiments, and therefore may suffer from the same limitation. In fact, it is likely that this is a manifestation of a robust effect described by Macdonald & Standing (2002), where participants are only willing to accept the positive feedback about the self, and thus we would expect other experiments to present a similar Barnum effect pattern thus failing to explain the difference in findings observed. Furthermore, although we do not observe the Barnum effect for all conditions, for the conditions where it is manifested we can test the hypothesis that accepting false feedback would lead to subsequent changes in related domains – which does not appear to be the case.

Another possible reason for the discrepancies in the current findings and those of past research is the format of the feedback provided. Halperin and Snyder (1979) and Sakamoto et al. (2000) alike used pre-designed verbal feedback to elicit the Barnum effect, whereas in the current research we enhanced or reduced the scores presented to the participants by transforming the numeric outcome from a scale of 1 to 100, to either 1 to 50 or 51-100. With the personal nature of verbal feedback, and more opportunity for the experimenter to motivate the participants, it is possible that the former elicited behavioural effects due an experimenter effect which numbers would not capture thus leading to a more accurate conclusion. On the other hand, the multiple ways in which numeric feedback can be interpreted may have failed to capture a real effect – although participants were told that the number is designed to reflect their performance relative to other people, the average performance of the majority is not well defined which could have led to confusion. Although this may limit the conclusions based on the observed data to numeric feedback type, the findings remain consequential as this comprises a very common format of feedback delivery.

Another key difference between the experiments presented is that for the first time we use a mixed gender sample, whereas in the past female only samples were used. There is strong evidence to suggest that females are more susceptible to feedback effects as they are more sensitive to the Barnum effect (Layne, 1998) and the experimenter effect (Deaux, 1985). However, we find no gender difference in feedback susceptibility in the current study, suggesting this is unlikely to be the determining factor.

It is also important to note that both the Barnum effect and the consequences it may have on subsequent behaviour may simply be domain dependent, with phobia treatment and stranger interaction simply more susceptible to change compared to self-reported willingness to volunteer and risky choice – although the reasons for this unclear. Perhaps some domains simply have a closer relationship with personal feedback. For instance, the manipulation of numerous personality traits could introduce complexity which could lead to uncertainty on how the traits relate to volunteering behaviours. Or the perceived relationship between the personality profiles and the dependent variable may have simply not been as strong as that of propensity to change and susceptibility to treatment (Halperin & Snyder, 1979), or extraversion and stranger interaction (Sakamoto et al., 2000). However, this should not have been a concern in experiment two given the direct relationship between financial risk attitudes and financial risky choice. Future research investigating the Barnum effect and its consequences across a wide range of feedback types is necessary to establish whether the observed pattern is universal or domain dependant.

Lastly, we must note the possibility that lack of appropriate control condition and inconsistent effects of past research has led to concluding an effect that may simply not be there. For instance, in the experiment by Sakamoto et al. (2000) dependent variables are analysed one by one and only statistically significant effects are discussed in conclusion – neglecting the possibility that a multivariate approach to the problem may, in fact, result in opposing effects balancing to produce a null effect overall.

The possibility that believing bogus information about the self does not alter subsequent behaviour is very consequential. First, it suggests that altering feedback cannot achieve behavioural change, clinical or otherwise, as originally proposed by Halperin and Snyder (1979). Second, it demonstrates that although we may use external information to guide our behaviour (Bem, 1967), this does not appear to extend to relying on feedback we think is derived from psychometric measures.

In conclusion, although past research has indicated that bogus feedback can impact subsequent behaviour we fail to detect the same effect. The data presented here suggests that bogus personality feedback, and financial risk attitude feedback does not have an impact on willingness to volunteer for psychology experiments and risky choices respectively, even when it is viewed as credible.

Chapter 5.
Ternary choice blindness: increasing the number of choice alternatives enhances the detection of mismatch between intention and outcome.
(Paper II)

Mariya Kirichek –*Warwick Business School*
Alex Cooke – *Kingston University London*
Paul Van Schaik – *Teesside University*
Petko Kusev – *Kingston University London*

5.1 Abstract

Choice blindness is the failure to detect a mismatch between intention and outcome when respondents make a choice (Johansson et al., 2005). The phenomenon is well established across various decision types and domains, from moral opinions (Hall et al., 2012) to haptic stimuli (Steenfeldt-Kristensen & Thornton, 2013). Research, however, has been largely limited to binary choice. Accordingly, we studied how increasing the number of decision alternatives to three options affects respondents' ability to detect the switch from their chosen stimulus to a non-selected alternative. We found that increasing the number of decision alternatives leads to an increase in switch detection, however, only when the alternative wrongly presented as the preferred choice was less attractive and dissimilar to the chosen stimulus. The roles of salience and preference strength are discussed as possible explanations for the observed effect.

5.2 Introduction

Choice blindness refers to the behavioural failure to notice a mismatch between choice and outcome. The basic procedure involves presenting participants with two alternatives and asking them to select their preferred option. Participants then undergo a distractor task, typically rating confidence in their choice (Pärnamets et al., 2015), and are then presented with their selected option once again. However, in the following feedback stage, their choice is switched for the non-selected option and they are asked to explain why they preferred the presented option out of the two alternatives originally presented. This is followed up by assessing detection retrospectively through asking whether the participant noticed anything unusual on completion of the experiment (Johansson et al., 2005). The proportion of trials reported as detected in past research has varied from 12% to 88%, depending on the type of stimulus used in the choice (Johansson et al., 2005; Somerville & McGowan, 2016).

The choice blindness phenomenon was first demonstrated using female faces as stimuli and involved using a sleight of hand to switch a participant's preferred choice for the non-chosen alternative before asking them to explain why they preferred the presented image (Johansson et al., 2005). Since it was first reported, the choice blindness paradigm has become a well-established psychological method and has been demonstrated with a wide range of stimuli, from moral opinions (Hall et al., 2012) to haptic stimuli (Steenfeldt-Kristensen & Thornton, 2013). The empirical robustness of the choice blindness phenomenon across domains and task types has been demonstrated beyond what would be expected from random errors in evaluation (Loomes & Sugden, 1998; Woodford, 2014). This has raised concerns about models of consumer choices as well as theories of human cognition based on assumptions of stable and coherent preferences (e.g., van Schaik, Kusev & Juliusson, 2011). Therefore, exploring this psychological phenomenon further is fundamental for understanding the nature of preference formation and decision consistency.

To our knowledge, the choice blindness procedures used in previous research have been limited to either judgment tasks or tasks using binary choice, with the exception of one study which explored eye-witnesses' recognition of one face among six concurrently presented faces (Sagana et al., 2013). Sagana et al. (2013) reported a higher level of switch detection than previously reported for choice blindness demonstrated using facial stimuli (e.g., Johansson et al., 2005). However, as a result of procedural differences, as the study used a field method (closely resembling the legal procedure), it is impossible to draw any valid conclusions from cross-study comparisons. Furthermore, without a binary choice control condition, it remains uncertain how increasing the number of alternatives may impact choice blindness. Accordingly, in the current experiment, we use binary and ternary variations of the choice blindness paradigm to examine how detection of switched choices may be affected by the number of choice alternatives.

The change from two to three alternatives could influence preferences in a manner greater than simply reducing the likelihood of alternatives being chosen. For example, Collins and Vossler (2009) report less deviation from induced preferences in a ternary choice task than in a binary choice task. Rolfe and Bennett (2009) also found that increasing the number of choices can reduce decision uncertainty, as demonstrated by the proportion of people indicating that they are unsure about their choice.

Furthermore, choosing from triplets as opposed to pairs could alter how respondents treat different characteristics of the presented alternatives, such as their attractiveness and similarity. Since similarity of stimuli has been found to impact choice blindness in the past (e.g., Sagana et al., 2013; Steenfeldt-Kristensen & Thornton, 2013), and differences in attractiveness are known to impact preference strength and choice certainty (Olsen, Lundhede, Jacobsen & Thorsen, 2011) we anticipate that if we enhance the perceived difference of physical features, or the level of attractiveness, between alternatives the chance of detecting a switch is likely to be increased. Since past research suggests that increasing the number of alternatives is capable of producing such change, this further presents an interesting avenue to explore.

Evidence that perceived similarity is dependent on the number of alternatives is quite abundant. For instance, presenting two different alternatives alongside a third alternative that is similar to one option but not the other, can increase the visual salience of the dissimilar alternative (Taylor & Fiske, 1978) as well as the salience of the dimensions on which it differs (Bordalo, Gennaioli & Shleifer, 2012). The increase in salience can, in turn, enhance the memory for the dissimilar item (Pedale & Santangelo, 2015), as well as its perceived magnitude of dissimilarity to the other alternatives (Tversky, 1977) – characteristics likely to increase a person's ability to differentiate between the alternatives after encountering them and thus increasing detection of mismatch between intended choice and outcome. Accordingly, we anticipate the similarity of stimuli may impact how the number of alternatives affects choice blindness.

Whilst the effects of adding a third alternative are well researched for alternatives of similar attractiveness and asymmetrically dominated alternatives (Ariely, 2008; Huber et al., 1982), the perceived distance between alternatives where one is clearly dominant is less established. Nonetheless, the evidence that certainty increases with the number of alternatives (Rolfe & Bennett, 2009) may suggest that people make a stronger distinction between the utility of chosen and non-chosen alternatives. As a result, we anticipate that the perceived difference in attractiveness may be enhanced when three alternatives are available compared two.

The research question of the current study is what effect, if any, increasing the number of alternatives in a choice task has on participants' ability to detect a discrepancy between their indicated preference and the stimulus presented during feedback. In addition, we explore whether varying the similarity and attractiveness of one of the alternatives presented in the choice has an impact on any relationship observed. Based on the past findings that people are more consistent in their preferences for ternary compared to binary choice (e.g., Collins & Vossler, 2009), we hypothesise an increase in the detection rates exhibited, especially when the alternatives presented in the choice are dissimilar and varied in attractiveness. In order to maintain consistency with past literature, we use facial stimuli used in the first choice blindness study (Johansson et al., 2005) to investigate this effect.

5.3 Method

An independent measures 2×2×2 design was employed, with independent variables consisting of *number of alternatives* (binary or ternary choice), *similarity* (similar or dissimilar), and *attractiveness* (uniform or varied). In accordance with past research (e.g., Johansson et al., 2005; Sagana et al., 2013; Somerville & McGowan, 2016), two dependent variables were used to assess whether participants experienced choice blindness: *concurrent detection* and *retrospective detection. Concurrent detection* (detected or undetected), was determined from the justifications participants provided for selecting the non-chosen alternative immediately after the manipulation. *Retrospective detection* (detected or undetected) was measured through a post-experiment questionnaire which required participants to indicate if they noticed anything unusual, followed by the instructions to describe what it was they thought was unusual. For both dependent variables, the written responses were assessed to determine whether participants detected the switch. Any reference to the face being different from that selected, choosing the wrong face in error or being presented with the face the participant did not prefer resulted in the trial being coded as detected. All remaining trials were coded as not detected. Since retrospective detection is considered to be a conservative measure of detection, the intention was to focus the analysis on this variable, however the pattern of results unusually demonstrated a higher proportion of trials detected concurrently and a composite variable *overall detection* was created to ensure the analysis includes all forms of detection.

Stimulus Standardisation

Overall sixteen combinations of faces were created: eight triplets and eight pairs (see Appendix). Specifically, two highly similar face pairs were selected from the stimuli used by Johansson and colleagues (2005) and extended to triplets using head on shots of female faces from The Psychological Image Collection at Stirling (PICS) online face database (*http://pics.psych.stir.ac.uk/*). One of the pairs was high on attractiveness, whilst the other low on attractiveness. Four triplets were created with each of the pairs, by adding a third face which was either similar to the pair in question both physically and on the level of attractiveness, similar physically but of different attractiveness, dissimilar physically but of similar attractiveness, or dissimilar both physically and on attractiveness. When selecting the faces different on attractiveness, if the pair to be extended was unattractive high attractiveness faces were chosen, when the pair was attractive low attractiveness faces were chosen. Eight triplets were created overall. Similarity and attractiveness were defined based on the ratings of 350 (173 female) people with the mean age of 48.35(SD = 12.65) in a pilot study.

An additional eight pairs of faces were selected to create binary choice stimuli. Following the same principle as for ternary stimuli, each of the two faces from one of the high similarity pairs used by Johansson et al. (2005) was combined with one of four possible images: either similar physically and on attractiveness, similar physically but of different attractiveness, dissimilar physically but of similar attractiveness, or similar both physically and on attractiveness.

Participants

The experiment was carried out on-line using the Qualtrics Survey Platform. Overall 521 participants took part in the experiment. 54 people were excluded from the analysis due to selecting a less attractive stimulus in the set, thus deeming the stimulus manipulation unsuccessful. For the 467 (39.2% male; mean age 46.18) participants included in the analysis, 225 (36.9% male; mean age 46.99) took part in the binary choice blindness paradigm and 242 (41.3% male; mean age 45.42) in the ternary choice blindness paradigm.

Procedure

The procedure was taken from the first study to demonstrate choice blindness with female faces (Johansson et al., 2005) and adapted to fit with a single (either one binary or one ternary choice), computerised, online decision task in order to minimise any interference of previous choices on the manipulated decision. Participants read the instructions and, after giving consent, proceeded to the main experiment. One of the 16 possible choice sets (8 binary and 8 ternary) were presented to each participant at random, and they were instructed to select their preferred alternative. The participants then provided a rating of confidence in their choice on a scale from 0 to 10, partly to assess the effect of confidence on the switch detection and partly as a distractor task.

Participants were further presented with a face that they did not select and asked to justify why they chose it. Since for ternary choice two of the alternatives were always visually similar and matched on attractiveness, if one of these faces was selected this was switched for the face outside of the pair to maximise the variability of faces encountered in the switch. For participants who chose the face that was not a member of the matched pair, one of the other alternatives was presented at random. For binary choice, the only other non-selected alternative was presented.

Lastly, the retrospective detection of the switch was assessed by asking the participants to indicate whether they noticed anything unusual in the experiment, by selecting ‘Yes’ or ‘No’. If they answered ‘Yes’ they were then required to provide a description of what they thought was unusual. After the experiment participants received full debrief information, and given the researchers’ contact details in case of any concerns.

5.4 Results

The proportion of participants that detected a mismatch of choice and outcome was 56.5% (53.3% binary; 59.5% ternary) concurrently and 42.8% (38.7% binary; 46.7% ternary) retrospectively. As retrospective detection, usually a conservative estimate of detection, was lower than concurrent detection, an ‘overall detection’ variable was created, coded as 1 (‘detected’) when participants detected the switch either concurrently or retrospectively, to capture all possible types of detection present, all other trials were coded as 0 (‘not detected’). The proportion of participants to demonstrate overall detection was 60.0% (56.9% binary; 62.8% ternary).

Visual analysis of the data (see Figure 1.) revealed a different pattern of effect of similarity and attractiveness on binary and ternary choice. For binary choice the range of detection rates was lower (38.3% to 69.2%) compared to ternary choice (38.8% to 96.2%).

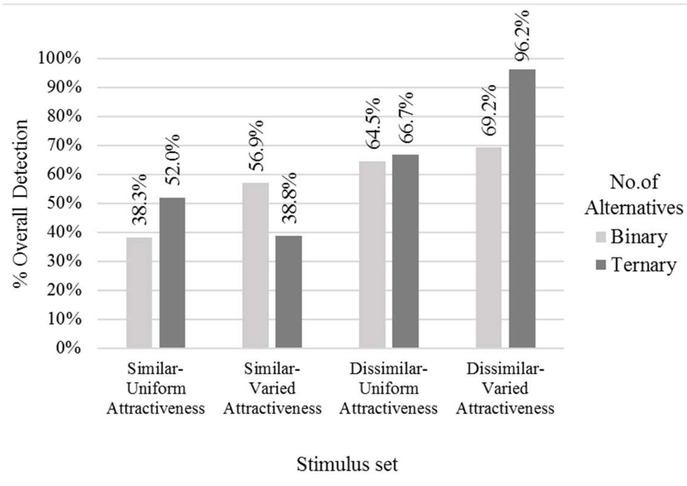


Figure 1. Proportion of detected trials by number of alternatives, similarity and attractiveness.

Effects of stimulus similarity on overall detection were significant for binary (χ^2=9.008, p=.003, df=1) and ternary (χ^2=27.995, p<.001, df=1) choice. The effects of attractiveness were not significant for binary or ternary choice (χ^2=2.995, p=.084, df=1; χ^2=2.250, p=.134, df=1).

A binary logistic regression was conducted on overall detection as a dependent variable. The independent variables were number of alternatives (binary=0 or ternary=1), attractiveness of faces (varied=0 or uniform=1[1]) and physical similarity of faces (dissimilar=0 or similar=1).

The three-way interaction was significant (odds ratio = 36.63, *CI* (95%) = [4.97; 270.07]). Therefore, interpretation of the two-way interactions and the main effects was precluded and subsequently, the effect of alternatives was analysed by similarity and attractiveness (see Table 1). The effect of alternatives was significant when the faces were dissimilar and varied in attractiveness (odds ratio = 11.11, *CI* (95%) = [2.40; 51.37]), but not for any other similarity-attractiveness conditions.

[1] Two variants of attractiveness coding were considered: one coded for the proportion of unattractive faces in the set, and the binary category coding reported here. The latter was deemed more appropriate as it resulted in a better model fit (-2 Log likelihood 564.72 vs, 557.78).

Table 1
Overall detection by number of alternatives, similarity and attractiveness.

Similarity	Attractiveness	Alternatives	*N*	*p* (detected)	Odds	OR	CI(95%)	
							Lower Limit	Upper Limit
Dissimilar	Varied	Binary	52.00	0.69	2.25	11.11	2.40	51.37
		Ternary	52.00	0.96	25.00			
	Uniform	Binary	62.00	0.65	1.82	1.10	0.53	2.28
		Ternary	66.00	0.67	2.00			
Similar	Varied	Binary	51.00	0.57	1.32	0.48	0.22	1.07
		Ternary	49.00	0.39	0.63			
	Uniform	Binary	60.00	0.38	0.62	1.74	0.87	3.47
		Ternary	75.00	0.52	1.08			

Confidence. The mean confidence reported by participants was 7.20 (*SD*=2.04).Analysis of variance revealed no significant difference in self-reported decision confidence between the participants that demonstrated overall detection and those that did not (F (1,465) =2.042, p=.154). Visual analysis (see Figure 2. overleaf) of confidence by number of alternatives, similarity and attractiveness revealed a disparity in confidence for detected compared to non-detected trials when the participants were presented with binary stimuli varied in physical features and attractiveness, with a mean confidence of 6.06 for non-detected and 8.17 for detected trials. The difference was significant F (1, 51) =18.586, p<.001). No significant difference was detected for any of the other stimulus types.

Decision Time. The average decision time was 4.48 seconds (*SD*=12.99). A general linear model analysis with decision time as a dependent variable and overall detection, number of alternatives, similarity and attractiveness and their interactions as predictor variables was not significant overall (F (15, 466)=.488, p=.947). None of the individual predictor variable reached significance (p>.05).

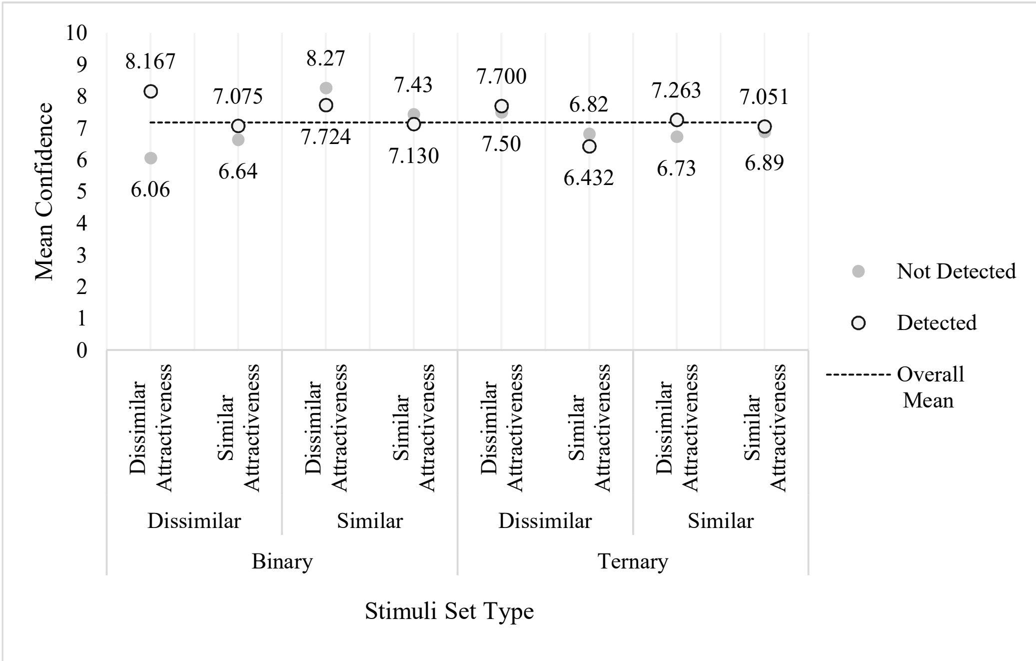


Figure 2. Mean confidence for detected and non-detected trials, by stimulus set presented.

5.5 Discussion

We investigated the effect of number of alternatives, similarity and attractiveness on choice blindness: people's inability to detect a switch from their stated preferred face to a non-chosen alternative. Our principal finding was that participants are more likely to detect the switch when selecting from three, compared to two alternatives, however, only when the choice used in the switch is dissimilar to the other alternative and less attractive. It must be noted that due to the unusual pattern of data, the analysis was performed on overall detection, a combination of concurrent and retrospective detection. Whilst this is not the convention in the choice blindness literature, this was done to ensure the highest possible sensitivity to detection. As a result, there is a possibility that the level of detection is over exaggerated compared to other work, however the pattern was representative of the results apparent for concurrent and retrospective detection in isolation.

Overall our findings appear to be partially consistent with past literature that suggests increasing the number of choice alternatives can lead to an increase in choice consistency (Collins & Vossler, 2009); however, for choice blindness this has been limited to very specific sets of stimuli. The interaction between the stimulus parameters and number of alternatives is, however, not surprising, as introducing an array of more than two objects can alter the relative salience of alternatives, and thus the perceived differences between them (e.g., Taylor & Fiske, 1978; Bordalo et al., 2012).

Consider the scenario in which choice detection is enhanced for ternary compared to binary choice. For binary choice, the alternatives consist of two visually different alternatives *a* and *b*. For ternary choice, we add a third alternative, a' which is similar to *a*. This results in psychological grouping of *a* and a' (Tversky, 1977), making the differences of alternative *b* more salient for ternary compared to binary choice (Taylor & Fiske, 1978; Bordalo et al., 2012). After the participants' make their choice, it is switched to a dissimilar alternative (either *a* or a' to *b*, or *b* to *a* or a'), with the differences more salient, and therefore likely to be better remembered (Pedale & Santangelo, 2015) for the decision-maker who encountered the alternatives in a ternary as opposed to binary choice. It is likely that by making the differences between alternatives more salient, we are simply amplifying whatever mechanism leads to similarity-induced differences in binary choice. This, in turn, could be because people are more likely to able to recall the differences between the faces presented.

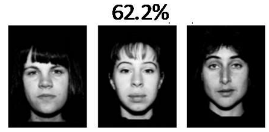
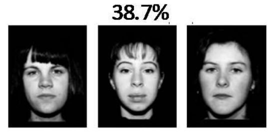
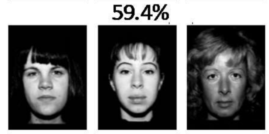
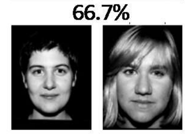
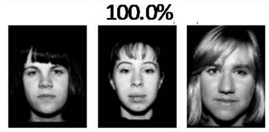
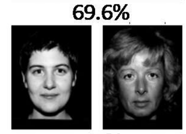
Furthermore, the effect of number of alternatives was only present when the chosen alternative and the alternative presented during feedback differed in their level of attractiveness. We hypothesised that this may be a result of increased perceived difference in attractiveness, as may be suggested by the increase in certainty for ternary compared to binary choice (Olsen et al., 2011). Whilst, we find that confidence decreases with ternary choice, and only appears to distinguish between detectors and non-detectors for binary choice with stimuli varied in attractiveness and similarity, this is in fact not surprising since participants still had to choose between two similar variables even if they eliminated one of the alternatives with certainty. To establish the certainty of eliminating an alternative early, one would need to ask participants to indicate their confidence in preference for every possible pair of faces presented, which is likely to influence subsequent detection rates due to repeated decisions strengthening the preference (Hoeffler & Ariely, 1999). Accordingly, it is impossible to establish whether participants' certainty is altered when one of the alternatives differs in physical features and attractiveness, compared to physical features alone.

One possibility is that participants do not only need to have the enhanced ability to differentiate between alternatives, but also need to justify why that difference resulted in the choice made. There is evidence to suggest that participants are able to remember the alternatives they encountered during their choice, but not the decision made and thus when provided with feedback reconstruct their preference based on the newly available information (Pärnamets et al., 2015). When reconstructing their preference, a participant must consider 'why would I have preferred this option', and if they cannot distinguish between choices or are indifferent to the outcome they can justify any alternative. On the other hand, if the physical difference of alternatives is enhanced the ability to differentiate between the two increases, and thus justification becomes dependent on whether the two are equally liked.

Overall, it appears that increasing the number of alternatives from two to three and the level of confidence in the original choice can significantly reduce choice blindness as predicted by past research (e.g., Collins & Vossler, 2009) in terms of the proportion of trials in which participants fail to detect a mismatch between their intended choice and presented outcome. However, this only occurs when one of the alternatives in the choice set is dissimilar to the others and less attractive. Although it is uncertain whether the number of alternatives impacts choice blindness through manipulating the perceived magnitude of similarity, or attractiveness, or both, one or more of these mechanisms are likely to be responsible for the observed phenomenon.

5.6 Appendix

Face combinations used in the experiment (eight triplets and eight pairs) with respective overall detection rates, by similarity and attractiveness.

Physical Similarity	Attractiveness	Ternary	Binary
Similar	Uniform	1.  62.2%	1.  26.7%
		2.  42.1%	2.  50.0%
Similar	Varied	3.  38.7%	3.  66.7%
		4.  38.9%	4.  62.5%
Dissimilar	Uniform	5.  59.4%	5.  66.7%
		6.  73.5%	6.  62.5%
Dissimilar	Varied	7.  100.0%	7.  69.6%
		8.  80.0%	8.  69%

Chapter 6.

Choice Blindness for preferred versus non-preferred stimuli. (Paper III)

Mariya Kirichek – *Warwick Business School*
Nick Chater– *Warwick Business School*

6.1 Abstract

Procedural invariance is one of the fundamental assumptions of rational choice theory, yet past research has demonstrated that choosing versus rejecting alternatives can result in preference reversal (Shafir, 1993) and affect the congruence of the choice outcome with previously stated priorities (Kogut, 2011). This study asks whether the framing of a choice task and subsequent justification task, in terms of choosing or rejecting, also affect participants' ability to detect a mismatch between the choice made and feedback presented as part of the justification task, or the extent to which participants exhibit choice blindness (e.g., Johansson et al., 2008). Our findings suggest that participants are more likely to detect a mismatch when they are required to select their preferred alternative and asked to explain why they preferred the (non-chosen) choice, compared to when they are required to choose or justify their least preferred choice, or both. The direction of the effect appears to contradict the previous finding that participants are more likely to exhibit consistent preferences when the task is framed negatively (Kogut, 2011). The possible procedural and cognitive differences between the consistency task used in the past and the choice blindness tasks presented here are discussed as a possible explanation of the diverging results.

6.2 Introduction

Preferences are highly subjective in their nature, making the quality of preferential choice difficult to evaluate. However, it is widely considered that stable preferences, which result in consistent choices being made over different points in time, are better than unstable preferences, at least from the rationality perspective (e.g., Rieskamp et al., 2006; Samuelson, 1938; von Neumann & Morgenstern, 1944). Accordingly, volumes of research have been dedicated to identifying what leads to violations of choice consistency, by establishing which seemingly irrelevant variables can lead to preference reversal. Many such factors have now been discovered, including positive and negative framing of the task (Shafir, 1993; Tversky & Kahneman, 1985), presence of other alternatives (Huber et al., 1982), recently encountered information (Herriges & Shogren, 1996), and even the font in which the information is presented (Novemsky, Dhar, Schwarz & Simonson, 2007). Interestingly, even when faced with identical choices over different points in time people often fail to be consistent (e.g., Camerer, 1989; Hey, 2001; Loomes & Sugden, 1998). For instance, Hey (2001) found that where risky choice is concerned people tend to reverse their choice on 1 to 21% of trials.

Over the past decade, a new phenomenon, choice blindness (Johansson et al., 2005) has been used to demonstrate that preferences are often unstable. The choice blindness paradigm involves measuring a persons' ability to detect a mismatch between their indicated preference, and the later presented non-preferred outcome, by switching the alternative and asking the participants to explain why they selected the alternative presented. The results indicate that participants often fail to detect the switch and coherently justify the wrong choice (Johansson et al., 2005; Johansson et al., 2006). The estimated proportion of trials on which choice blindness, or the inability to detect a switch, occurs has been shown to vary between 12-88% depending on domain and procedure adopted (e.g., Johansson et al., 2005; Somerville & McGowan, 2016), with this variability suggesting that there is potential to use the paradigm as a tool to investigate which task parameters lead to greater, or lower, choice stability. Surprisingly, the paradigm has been seldom applied to exploring the factors thought to lead to preference reversal. There is however, reason to suggest choice blindness is closely related to choice consistency. Just as different response across two different trials indicates preference reversal, the inability to recognise that a presented alternative is not the one we indicated as a preference, accompanied by the ability to explain why we preferred the non-chosen alternative, also appears to suggest that preference change occurred. Similarly, the patterns observed with choice consistency have also been detected for choice blindness. For example, just as choice consistency improves as participants repeat trials, and become familiar with the products at hand (Brown et al., 2008), a very high level of detection is reported for products with which participants are already familiar (Somerville & McGowan, 2016). Here, we attempt to use the choice blindness paradigm to investigate one factor that has been implicated in preference reversal, namely whether the task is framed in terms of accepting or rejecting an alternative.

The very existence of choice blindness may come as a surprise, however, since it was first reported a little over a decade ago, using a magic trick to switch out participant preferences for female faces (Johansson et al., 2005), it has become a well-established phenomenon and has been demonstrated across a wide range of domains, from haptic stimuli (Steenfeldt-Kristensen & Thornton, 2013), to consumer choices (Hall et al., 2010), and moral decisions (Hall et al., 2012). The most prominent stimuli used, however, have been female faces. Choice blindness for female faces has now been established not only using a physical sleight of hand, but also using computerised paradigms (e.g., Pärnamets et al., 2015), and has been used to establish choice blindness in experiments designed to maximise ecological validity (Sagana et al., 2013), as well as to investigate the role of memory processes in choice blindness (Pärnamets et al., 2015). Accordingly, the use of faces to demonstrate choice blindness is the most validated and established approach and will also be applied to investigating the effects of framing in the current experiment.

The effects of task framing have been investigated with respect to a wide range of decisions, from hypothetical ice cream preferences (Shafir, 1993), to decisions on child immunisation (e.g., Abhyankar, O'Connor & Lawton, 2008; see also Tversky & Kahneman, 1985). For the purpose of understanding how selecting the most versus least preferred alternative can impact choice, let us consider the work carried out by Shafir (1993). In his experiments, Shafir (1993) manipulated the framing of the task by asking participants to either accept or reject one of two alternatives, and gave participants a choice between an enriched (higher positive and higher negative features), and impoverished (fewer positive and fewer negative features) option. To illustrate, in one of the scenarios participants were asked to imagine that they are planning to take a short break. They were then presented with two alternatives, but whilst half of them were instructed that they could choose which option to book, the other was told that reservations were being held for them in both places and they had to select which option they would like to cancel. The two alternatives were:

Spot A	***Spot B***
Average weather	Lots of sunshine
Average beaches	Gorgeous beaches and coral reefs
Medium-quality hotel	Ultra-modern hotel
Medium-temperature water	Very cold water and very strong winds
Average nightlife	No nightlife

When participants were asked to indicate which spot they would like to book, a 67% majority preferred Spot B, however, when they were asked which option they would cancel 52% rejected Spot A. Logically, the two numbers should have been the same, however, it became apparent that a higher proportion of people preferred Spot B (enriched alternative) when choosing which holiday to book, compared to when choosing which holiday to cancel. Shafir (1993) went on to demonstrate a similar effect in child custody decisions, ice cream preference, financial lotteries, elections of town council president, and educational course selection. The overall observation was that when asked to select which option is preferred, or should be accepted, participants preferred the enriched option, whereas when asked which option is the least preferred, or should be rejected, participants tended to prefer the impoverished choice.

Shafir, Simonson and Tversky (1993) explained this effect through adopting the view that decisions can be explained by taking into account reason based choice. The approach involves identifying the reasons and arguments that enter the decision process, and explains choice in terms of the reasons people would have considered for and against the possible alternatives. Shafir et al. (1993) argue that when asked to select a choice, people look for arguments in support of the available options, resulting in the selection of the choice with the highest positive features, whereas when asked to reject a choice people look for the possible reasons to eliminate an alternative, thus choosing the alternative with the lowest negative features.

Whilst Shafir and colleagues (Shafir, 1993; Shafir et al., 1993) provide ample evidence to suggest that people choose differently when asked to accept, or reject, one of the alternatives presented in the task, little is known about how this may influence choice consistency if the task is repeated more than once. Based on the pattern observed, we can hypothesise that choices would be more consistent across trials that use the same framing (i.e., accept task followed by an accept task, and reject task followed by reject task). For the purpose of this experiment, this suggests that framing both the choice and justification portions of the choice blindness task in the same way should lead to higher chance of switch detection. Consider for example a person making a choice of which alternative they prefer, and then being presented with the choice they did not select and asked to justify why they preferred it. From the reason based choice approach they would have considered the advantages of the alternatives during the task, and would also try to identify the positive traits of the alternative in the justification task. Having already evaluated the positive features before, the person would then be likely to recall that the positive features of the alternative have changed and in turn detect that the alternative presented was not the one they preferred. However, if they are asked to explain why they did not prefer the choice, having seldom considered the negative features that would explain lack of preference, the person would be less familiar with the negative features of the task involved and thus more likely to accept the erroneous choice as their own. Whilst we hypothesise that framing consistency across a choice blindness task would lead to increased detection of invalid feedback, we are still left with the question of whether just the accept or reject framing of the task could alter consistency, or in turn the likelihood of experiencing choice blindness.

Building on Shafir et al.'s (1993) theory (see also Yaniv, Schul, Raphaelli-Hirsch & Maoz, 2002), Kogut (2011) proposed that since the accept framing of a task is likely to result in participants focusing on the positive attributes of alternative, the attractiveness of the forgone options would become increased after it has been deliberated, thus bringing the attractiveness of the two options closer together and resulting in less choice consistency. Although Kogut (2011) did not demonstrate this using a conventional choice re-choice paradigm, he attempted to capture the effect of using an inclusion versus exclusion strategy in multi-alternative choice to establish which of the approaches results in higher consistency with earlier stated preference order.

To carry out the experiment, Kogut (2011) asked participants to rate programs designed to improve a range of aspects of schools or universities according to their importance, from very important to not important at all. After a distractor task participants were presented with the programs again and asked to imagine that they were to meet with a donor and discuss which of the programs are to be supported. In order to prepare, participants had to mark the programs that were most or least important to them (accept vs. reject frame). After they made their decisions, the participants underwent the final choice, where they were instructed to choose only two best programs to present to the donor. Kogut (2011) was interested in two outcome variables in this experiment; the size of the consideration set used to make the choice, which appeared to be higher for the reject frame, and the consistency between the initially stated preference and the final two choices, which were found to be more consistent under the reject frame. Choices were considered consistent if no other programs were initially rated higher than the final selections. Kogut (2011) further provided evidence that supported his proposed mechanism underlying the aforementioned difference, by demonstrating that when instructed to consider both advantages and disadvantages, the choice consistency for positive framing is improved.

Overall, the research by Kogut (2011) suggests that reject framing results in higher choice consistency than accept framing. Accordingly, we hypothesised that exclusion framing would be more likely to result in responses consistent with initial preference stated in the choice blindness paradigm, in other words, increased ability to detect a mismatch between intended and presented outcome. However, since the framing in Kogut's (2011) experiment is thought to exert influence prior to the final selection, we can only equate this to the initial selection process in the choice blindness task, combining this with the earlier proposal that consistent framing should result in increased detection of incongruence between choice and outcome, we can hypothesise that participants will be more likely to exhibit detection when the choice and justification instructions are framed as rejections.

It must be noted that given the binary nature of the choice blindness paradigm, due to the difficulty in presenting more than one non-chosen alternative as the single selected option, it is impossible to create an experiment directly parallel to that of Kogut (2011) as instead of using multiple options, only two can be initially presented. In fact, Kogut (2011) hypothesised that part of the mechanism responsible for the observed effect is the differences in consideration size – something that cannot differ within the choice blindness task. Consequently, this may lead to different patterns brought about by the procedure used by Kogut (2011), and that used in choice blindness. On the other hand, if we consider the reason based approach explanation to apply to both Kogut's (2011) work and choice blindness, the pattern observed should remain the same.

The study conducted by Kogut (2011) is certainly not a typical example of demonstrating choice consistency, or a direct parallel to choice blindness, however, it provides us with the best available hypothesis at this stage. It is also evident that the mechanisms best used to explain the framing effects on consistency, are at least in part transferrable to the choice blindness paradigm.

We, therefore, test whether participants would be more likely to detect that their choice has been switched when the choice aspect of the choice blindness paradigm was framed as an exclusion task versus inclusion task. In order to eliminate the possibility that any effect observed is a result of a change in task framing between the initial choice and justification as often demonstrated in preference reversal literature (e.g., Shafir, 1993), we also manipulated the framing of the justification task. Overall the experiment took on a 2 x 2 independent measures design, with two levels of choice framing (select preferred vs. select least preferred) and two levels of the justification task framing (why preferred vs. why least preferred). The stimuli used in the experiment were female faces taken from the Psychological Image Collection at Stirling (PICS) in line with the first choice blindness experiment of Johansson et al. (2005), in order to maintain consistency with past research. It was hypothesised that participants would be more likely to notice a mismatch between their chosen face and the one presented when they were required to choose and justify the face they do not prefer, compared to when they were required to choose and justify the face they did prefer.

6.3 Method

Stimuli

Photographs of female faces were taken from The Psychological Image Collection at Stirling (PICS) online face database (pics.psych.stir.ac.uk) in accordance with the first choice blindness study (Johansson et al., 2005). Eight pairs of faces were selected for the experiment. The stimuli sets used were counterbalanced for similarity in physical appearance and perceived attractiveness based on previously reported ratings (see Kirichek, Cooke, Van Schaik & Kusev, submitted). The final combinations were comprised of two face pairs that were similar physically and on attractiveness; two pairs that were physically dissimilar but of similar attractiveness; two pairs that were physically similar but of dissimilar attractiveness; and two pairs dissimilar both physically and on attractiveness (see Appendix).

Participants

Participants were recruited via the Prolific Academic recruitment platform. One hundred and thirty-nine participants (87 male) took part in the experiment with a mean age of 30.58 (*SD*=10.16).

Procedure

Participants were randomly allocated to one of four conditions that determined the framing of the choice and justification tasks within the choice blindness paradigms. After consenting to take part in the study and providing demographic information, participants completed a single choice blindness paradigm task, which consisted of the choice phase, justification phase and detection assessment.

During the choice phase participants selected between two faces, indicating either their preferred, or least preferred face depending on the task framing condition they were allocated to. Participants then provided a confidence rating in their choice as a distractor task, in accordance with past literature using computerised variations of the choice blindness paradigm (Pärnamets et al., 2015).

During the justification phase, participants were presented with a face on screen and required to justify why they prefer (not prefer) the face. The picture shown was always the one incongruent with the instructions. To achieve this, in some conditions participants were shown their selected image, whereas others the non-selected alternative. For instance, if the choice phase is framed as preferred and the participant chooses face *a* over *b*, for preferred framing of the justification task participants would be presented with picture *b* and instructed to justify why they preferred it, whereas for non-preferred framing of the justification task participants would be presented with face *a* and asked to justify why they did not prefer it. It must be noted that participants allocated to the preferred framing of both the choice and justification tasks underwent the conventional choice blindness paradigm featured in other research (e.g., Johansson et al., 2005).

In accordance with past research (e.g., Johansson et al., 2005; Sagana et al., 2013; Somerville & McGowan, 2016) two variables were used to assess whether participants detected the switch: *concurrent detection* and *retrospective detection.* Participants' justifications provided for preferring (not preferring) the presented face were used to establish c*oncurrent detection* (detected or undetected), a post-experiment questionnaire which required participants to indicate if they noticed anything unusual, followed by the instructions to describe what it was they thought was unusual was used to establish *retrospective detection* (detected or undetected). For both dependent variables, the written responses were assessed to determine whether participants detected the switch. Any reference to the face being different from that selected or choosing the wrong face in error resulted in the trial being coded as detected. All remaining trials were coded as not detected. In order to ensure all detected trials were included in the analysis a composite variable termed *overall detection* was created, where any trial detected concurrently or retrospectively was coded as detected, all other trials were coded as undetected.

Lastly, participants were provided with debrief information and researcher contact details in case of any concerns and paid for their time.

6.4 Results

The proportion of participants to detect a mismatch of choice and outcome (coded 0 for not detected 1 for detected) was 40.1% concurrently and 38.1% retrospectively. 47.4% of trials were detected overall.

Visual examination (see Table 1) of overall detection by framing of the choice task (coded 0 for preferred 1 for least preferred) and the framing of the justification task (coded 0 for preferred and 1 for least preferred) revealed that when choice and justification tasks were both framed as preferred, the proportion of participants detecting the switch was visibly higher compared to other conditions. A chi-squared analysis revealed that the detected proportions were significantly different across the four task variations (χ^2=23.945, p<.001, df=3). Pairwise comparisons further confirmed that the proportion of detected trials when both task and justification were framed in terms of the preferred alternative was significantly higher compared to preferred choice and non-preferred justification framing (χ^2=17.241, p<.001, df=1), non-preferred choice and preferred justification framing (χ^2=11.470, p=.001, df=1), or when both choice and justification were framed as non-preferred (χ^2=19.625, p=.001, df=1). No other comparisons were significant (p>.05).

Table 1. Overall detection of mismatch in choice and outcome, by choice and justification task framing.

	Prefer Choice Framing		Not Prefer Choice Framing		Overall Mean
	Prefer Justification	Not Prefer Justification	Prefer Justification	Not Prefer Justification	
Overall Detection	***84.4%***	34.3%	43.8%	31.6%	47.4%

Confidence. Mean choice confidence rating was 7.54 (SD=2.07) on a scale of 1 to 10. A general linear model using confidence as a dependent variable, and choice and task framing and overall detection as predictor variables was not significant overall (F (7,136) =1.087, p=.375). No predictors were significant in isolation (p>0.05).

Decision Time. Average decision time was 9.70 seconds (SD=23.17). A general linear model for decision time as a dependent variable and choice and task framing and overall detection as predictor variables was not significant overall (F (7,136) =1.181, p=.318). No predictors were significant in isolation (p>0.05).

Stimulus Characteristics. Overall detection ranged from 37.9% to 58.3% depending on similarity and attractiveness of stimuli. The variation across conditions was not significant (χ^2=3.515, p<.319, df=3). A chi-squared analysis was carried out to ensure the distribution of stimuli did not vary by framing of choice and justifications framing conditions, this was not significant (χ^2=22.347, p<.380, df=21).

6.5 Discussion

In the current study, we investigated the effects of task on participants' ability to detect a mismatch between their indicated preference and presented outcome. The framing was manipulated at two stages of the choice blindness task, which consisted of participants being asked to select the female face they prefer (or not prefer) out of a pair of stimuli, followed by being shown a face and instructed to justify why they preferred (or did not prefer) it, the face however, never matched the instructions. We found that when at least one of the tasks (choice or justification) was framed negatively the detection of the mismatch was much lower, compared to when both tasks were framed positively (as in the original choice blindness paradigm).

The findings were the opposite of what we hypothesised based on past literature into how task framing can affect choice consistency. More specifically, Kogut (2011) found that asking participants to identify improvement programs they found least important resulted in their final choices being more consistent with their original ratings of importance, compared to when they were asked to identify the most important programs. In other words, the research suggested that negative framing of the tasks is likely to result in higher choice consistency – the opposite pattern to that observed here if we consider choice blindness and consistency to be closely related.

One possible explanation for the difference in the data observed to that hypothesised is that the relationship between inconsistent choice and choice blindness is not as direct as theorised. Although both presuppose a stable preference order to which preferences could be compared, there are other factors that differ between the two. For instance, whereas in Kogut's (2011) choice consistency task participants encounter all alternatives in the second elicitation task, only one option is presented to them in choice blindness, and it is well established that the alternatives amongst which a choice is presented can alter participants' behaviour (e.g., Huber et al., 1982). Furthermore, in re-choice consistency paradigms providing inconsistent responses is not incorrect per se, whereas in choice blindness not detecting the switch is equivalent to not detecting an error which may be prone to different psychological pressures, yet is more prone to demand characteristics as through reporting a switch a participant contradicts the researcher.

It is more likely, however, that the procedural differences in the tasks utilised by Kogut (2011) and that presented here are responsible for the divergence of results. One of the key differences here is that the former used multiple alternatives, whereas the current study only uses binary choice. Kogut (2011) propose that one of the mechanisms through which participants achieve higher consistency is through the higher consideration set they identify when selecting the least compared to most important alternatives (Yaniv & Schul, 1997), a factor not present in the binary choice used in the choice blindness paradigm. On the other hand, Kogut (2011) also suggests that it is the consideration of positive attributes in the positive framing of the problem that results in the possible preference change, as the previously non-selected alternatives become more attractive (Yaniv et al., 2002), a factor that one would expect to influence the alternatives in the choice blindness paradigm as well as the aforementioned consistency experiment.

There are two other variables that may confound the observed pattern of results: whether the face presented was the one selected (with instruction to explain why it was not selected) or not selected (with instruction to explain why it was selected), and the congruence of framing across the two task. These features vary for participants who are presented with positive or negative feedback for both parts of choice blindness thus making the framing between the two conditions congruent and the face presented the non-chosen alternative, compared to when they are presented with negative framing for one of the tasks and positive for the other, thus making the framing incongruent and the face presented the chosen alternative. For example, when participants undergo a positively framed choice task, however, are then asked to explain why they did non-prefer a presented alternative they must be shown the face they chose in the first instance in order for the alternative to be incongruent with the instructions. This could impact whether participants notice the switch, as selected alternatives are known to be better remembered than the chosen alternative (McClelland, Stewart, Judd & Bourne, 1987). In addition, we also know that positive and negative frames of questions have been known to impact participants in a systematic direction (Shafir, 1993), thus the switch in phrasing could also theoretically impact the observed consistency. However, if this was indeed the reason for the different level of detection, we would expect to see the difference between the two conditions with congruent framing compared to the two with incongruent framing for the two tasks – yet we only observed the higher detection for positive congruent framing, but not negative.

The level of detection observed when the task is framed in a manner akin to the conventional choice blindness paradigm is much higher at an average of 76% compared to the highest detected proportion of 26% reported by Johansson et al. (2005). This might indicate that the observed result is an anomaly. However, even if the figure is higher than the average detection level that would be observed using this procedure in the population more generally, given the large effect size observed the difference between the positively framed task and the other framing combinations is unlikely to be eliminated. The procedural differences between the currently presented experiment and past choice blindness work are much larger than between the conditions in the current experiment, suggesting these are more likely to be responsible for the observed discrepancy than assuming that an almost 50% difference in conditions is completely due to chance. One such difference is that classically choice blindness experiments were carried out with a physical trick and experimenter present (Johansson et al., 2005; Steenfeldt-Kristensen & Thornton, 2013), whereas the current experiment used a computerised task. The detection rates in other computerised choice blindness tasks have indeed been similar to what we observe here (e.g., 63% – Pärnamets et al., 2015). In addition, the paradigm we use requires participants to only make a single decision, whereas in the conventional paradigm they complete 15 trials, seventh, tenth, and fourteenth of which are manipulated. The completion of the initial six choice tasks may build confidence, or trust, in the paradigm confounding the observed choice blindness with social effects not present in the one decision variation of the task.

Overall, it appears that we find a strong effect of framing on choice blindness, with positive framing resulting in higher detection of mismatch between intended choice and its outcome. The observed phenomenon is still, however, to be explained. One potential discrepancy is that processing information in its negative form (e.g., not prefer vs. prefer) is neurologically harder for the participant to process. For instance, Fischler, Bloom, Childers, Roucos, and Perry, (1983) show that when asked to judge statements such is 'A sparrow is a tree' participants show significant negative scalp potentials in the region of about 250 to 450 msec. following their presentation. However, when the same sentence is phrased in the negative (e.g., 'A sparrow is not a tree'), the same activity is observed even though in this instant the content of the statement holds true. This is also not surprising given that it is a lot more common encounter decision of what we prefer, compared to not prefer. For example, we choose which groceries to buy on regular basis, yet it is very rare for us to select our least preferred items from those available. Accordingly, it seems participants are more likely to recognise when a positively framed statement is untrue, thus explaining why the positive framing of the choice blindness tasks is more likely to result in higher detection of an erroneous outcome. On the other hand, the participants' ability to then provide justifications for the question 'Why did you not prefer the face' suggests that they could not have simply misunderstood the instructions and must have processed the negative information.

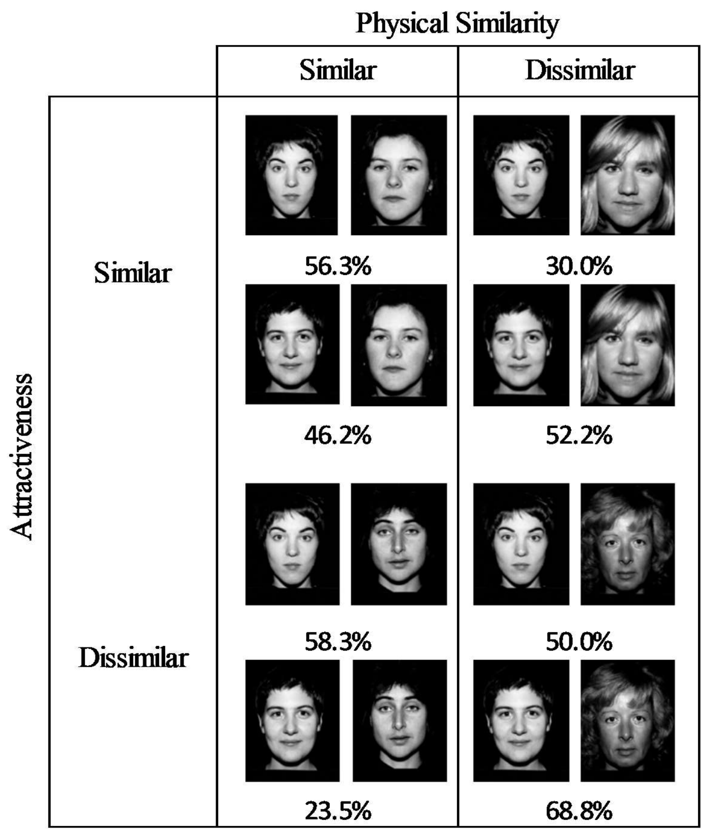
The finding that neither confidence nor decision time varied with detection, nor the task at hand, also provides an interesting insight. Assuming participants are less familiar with negative statements, and find it harder to process these we would expect this to reflect in longer decision times and lower subjective confidence, which is not the case. Whilst it must be noted that the reliability of decision time for online experiments is somewhat weak, the confidence data does raise the question of whether task difficulty can really be responsible for the pattern observed. On the other hand, the dissociation of confidence from choice blindness has also been reported in the past (Hall et al., 2013) and it is likely that people have limited insight into the stability of their choices (the likelihood that the same alternative is chosen if the task is to be repeated) and accuracy of their own performance.

It seems at this stage the data is not sufficient to identify the mechanisms that underlie the differences in choice blindness for positively and negatively framed tasks. Furthermore, we cannot definitively say whether the same mechanisms may guide choice consistency and choice blindness, as at least for the multiple alternative consistency task the effects of framing appear to be the opposite to those observed for choice blindness. If we are to answer these questions, future research into both choice blindness and choice consistency is required. First a consistency study using only binary choices, over two time points, and varying framing of the choice at each time point is necessary to establish whether framing affects consistency differently for multiple compared to binary choice, and in turn providing a more direct comparison for the choice blindness paradigm. Second variations of the choice blindness paradigm would help unravel why framing may play a role in choice blindness. For instance, including explicit instructions to consider both negatives and positives like Kogut (2011) would help identify whether justification strategies may impact the framing effects; or alternatively the use of negative particles can be replaced with single words (e.g., ‘not prefer’ with ‘hate’), to eliminated processing difficulties as the possible explanation.

The results presented here clearly demonstrate an interesting effect of framing on choice blindness: positive framing across all of the choice blindness paradigm results in a much higher proportion of participants detecting that the outcome presented is incongruent with their response, compared to when at least one of the tasks (choice or justification) is framed negatively. Such findings comprise an effect opposite to that hypothesised and further research is needed to understand this surprising discovery.

6.6 Appendix

Face combinations used in the experiment and the respective proportions of trials detected by similarity and attractiveness levels.



Chapter 7.

Choice blindness for stimuli external to the choice. (Paper IV)

Mariya Kirichek – *Warwick Business School*

Nick Chater– *Warwick Business School*

7.1 Abstract

Choice blindness refers to the inability to notice the switch of a previously selected preferred stimulus for a non-selected 'imposter' alternative during feedback (Johansson et al., 2005). The level of choice blindness is measured by asking participants to justify why they selected the imposter choice in place of the actual choice, followed by a request to indicate whether they noticed anything unusual. Although the levels of imposter detection vary by domain, it has become an established phenomenon. The processes underlying choice blindness, however, are not yet well understood. In the current paper, we ask whether people accept imposter choice regardless of having encountered it before; because they had seen it before; because they have evaluated it before; or because it was encountered as a part of comparative choice, even if that choice was separate from the one they are receiving feedback for. We found that stimuli not encountered previously, just seen previously, and previously evaluated in isolation induced a high level of detection. The detection significantly decreased when the participants encountered the stimulus in a preferential choice task preceding the actual choice they are receiving feedback for. We propose that although there appear to be a number of factors that contribute to choice blindness, the imposter choice needs to undergo a similar cognitive task to the one at hand in order to be falsely identified as the real preference made which in the case of the classic choice blindness paradigm entails being deliberated as a part of a comparative choice process.

7.2 Introduction

Choice blindness refers to the failure to detect a mismatch between intention and outcome when making a preferential choice (Johansson et al., 2005). This is usually demonstrated by switching the alternative selected by participants in a preferential choice task with the non-chosen alternative, or an 'imposter choice', and measuring the proportion of people that report noticing something unusual. Lower proportion noticing the switch is then taken as an indication of higher level of choice blindness and vice versa. For example, in the original choice blindness experiment (Johansson et al., 2005) participants were presented with two female faces printed on card and asked to select their preferred option. Their choice was then replaced with the non-chosen face using a sleight of the hand and the participants were asked to justify the presented option. The majority of people complied with providing a justification and failed to report noticing anything unusual having taken place. Furthermore, participants used features specific to the 'imposter' face in their justification suggesting the effect extends beyond not being able to tell the two alternatives apart Johansson et al., 2008).

The existence of choice blindness is now broadly accepted across many different domains, including faces (Johansson et al., 2008; Johansson et al., 2006; Sagana et al., 2013, 2014a), consumer products (Cheung et al., 2015; Hall et al., 2010; Sauerland et al., 2014; Somerville & McGowan, 2016), self-reported psychological health symptoms (Merckelbach et al., 2011), moral preferences (Hall et al., 2012), political opinions (Hall et al., 2013), personal finance (McLaughlin & Somerville, 2013) haptic stimuli (Steenfeldt-Kristensen & Thornton, 2013), and witness testimony for incident details (Aardema et al., 2015; Cochran et al., 2015) and faces and voices of individuals encountered (Sagana et al., 2013; Sauerland et al., 2013).The effects shown have been demonstrated beyond what would be expected from error in valuation alone (Loomes & Sugden, 1998; Woodford, 2014), and present a problem for models of consumer choice (Somerville & McGowan, 2016) and the existence of stable ordered preferences more broadly (Chater et al., 2011). The notion that our preferences do not appear to have a stable order of alternatives is not a new one of course, with a wide range of literature demonstrating that people fail to exhibit consistency in their preferences across alternatives, and different points in time (for review see Rieskamp et al., 2006). Nonetheless, choice blindness introduces a distinct approach to choice stability, as not only does it question just how stable our preferences actually are, it also highlights our own inability to monitor intentions as well as cognitive processes that gave rise to these intentions within a single choice (Johansson et al., 2005).

To date, a number of factors have been investigated as mediators of choice blindness (e.g., similarity – e.g., Steenfeldt-Kristensen & Thornton, 2013; time constraints – e.g., Johansson et al., 2005; choice certainty – e.g., Hall et al., 2013; and choice familiarity – e.g., Somerville & McGowan, 2016), however, all of the research to date assumes that choice blindness is specific to switching the alternatives that the participants were given to select their response from. The current paper takes a new look at choice blindness and when it occurs. More specifically, we attempt to identify at which stage of the choice process a stimulus becomes susceptible to being mistakenly recognised as a previously stated preference. To identify which aspect of the paradigm may be responsible we considered how the stages leading up to the decision unfold over time and noted four distinct stages; pre-task stage with no knowledge of the stimuli, encountering the stimuli, evaluating the stimuli, and comparing between the alternatives to make a selection.

If the stages identified here comprehensively cover the process leading up to the choice, the characteristics that make a stimulus susceptible to choice blindness must occur at one of these stages. It also follows that since the latter stages (e.g., comparing alternatives) also require participants to undergo the earlier processes (e.g., encountering and evaluating the alternatives), once we observe choice blindness in earlier stages it will also be present in the latter ones – and thus choice blindness becomes more likely with each proceeding step. Additionally, aside from the increase in the number of processes undergone, each interaction stage is also likely to exert an additive effect on liking or preference (e.g., Freedman & Fraser, 1966; Zajonc, 1968), as well as memory (e.g., Craik & Lockhart, 1972) of the stimuli encountered, the two processes at the core of monitoring preferential choice over time.

Literature suggests that exposing participants to items prior to evaluating them can result in more favourable ratings when assessing these stimuli later (Zajonc, 1968). Deeper interaction with such stimuli (such as evaluation or decision making) can also increase the cognitive ease of processing such stimuli further thus enhancing our evaluations of these items even more (Oppenheimer, 2008). This suggests that with each of the identified stages leading up to a choice participants would experience an increase in liking as each step requires additional, and thus deeper, processing. This could, in turn, be responsible for bringing about choice blindness as having already experienced positive evaluations of a stimulus we are more likely to agree to it later (e.g., foot in the door technique – Freedman & Fraser, 1966), and we are also more likely to judge such stimuli as true (Reber & Unkelbach, 2010).

Similarly, the memory trace for the stimuli encountered at each of the proceeding stages is also anticipated to increase as the depth of processing of the material increases (Craik & Lockhart, 1972). It has been proposed that memory failure may be able to account for choice blindness (Pärnamets et al., 2015; Sagana et al., 2014a), in which instance inability to recognise stimuli should not affect the susceptibility to choice blindness and stimuli that have never been encountered before would have the same chance of being mistakenly accepted as previously stated preferences. However, recognition memory of the stimuli presented in the choice blindness task has, in fact, been found to remain intact after participants undergo the choice blindness paradigm, regardless of whether they detect the mismatch of the presented outcome or not (Pärnamets et al., 2015; Sagana et al., 2014a). This would suggest that participants would be capable of detecting that they had not encountered a stimulus when a brand-new alternative is presented to them during feedback, and thus exhibit a very high level of detection when a brand-new face is presented. Additionally, the increase in memory strength with each stage of the decision would also increase the likelihood that we will mistake the face for one recently seen, decreasing the chance of the stimulus being recognised as external to the current task and thus decreasing the chance of detecting that the stimulus cannot be the one that was selected.

Although recognition memory seems to remain intact, in order for choice blindness to occur some form of memory failure seems to be essential. Pärnamets et al. (2015) propose that source memory is where such a failure occurs. Source memory refers to the knowledge of when and where something was learned (Pandey, 2011), in the case of the choice blindness example referring to knowledge that the image presented was seen during the choice presented in the experiment, when it was either rejected or chosen. Source memory is considered to be more reconstructive in nature than recognition memory (Johnson et al., 1993; Yonelinas, 1999), using in the moment judgements to determine a plausible set of events and thus more susceptible to suggestion. This suggests that as long as participants recognise the alternative presented to them during feedback, there is a likelihood that they will misjudge the alternative as the selected one when, in fact, it was not, resulting in choice blindness. Source memory misattribution, however, is very reliant on similarity of the sources in which information is presented (Johnson, Foley & Leach, 1988; Johnson et al., 1993), and therefore since each of the latter decision stages that we identify shares more with the decision we provide feedback for in the choice blindness paradigm the chances of falsely accepting the imposter choice also increases with each consecutive stage.

Overall, based on effects of preference and memory discussed above, we hypothesise that the susceptibility to accepting an imposter stimulus as one that we indicated as our preference would increase with each stage of the decision. Given the ability to identify a novel stimulus as one never encountered we hypothesise very high chances of detection (little to no choice blindness) when participants are requested to justify a never seen face. Conversely, given the similarity of the memory source when encountering a stimulus in a decision similar to the target task we anticipate such a scenario to elicit a similar level of detection to the choice blindness paradigm. For the two stages in between, a gradual increase is our best hypothesis.

In order to empirically test the stipulated hypothesis, we created four experimental conditions for each of the identified decision stages; novel, pre-exposed, pre-judged and pre-choice conditions. In the novel condition the imposter stimulus is one the participants have never encountered before and are not familiar with. In the pre-exposed condition, participants saw the imposter choice prior to the experiment in an array of alternatives and were simply instructed to look at the alternatives presented. In the pre-judged condition participants saw the imposter choice amongst the same array and were instructed to provide three positive points about the face to ensure they have evaluated the possible reasons for its selection[1]. And in the last, pre-choice condition, participants encountered the imposter choice as a part of a binary selection task preceding the choice task that will consequently receive the erroneous feedback. We used an alternative from the preceding trial as opposed to the classical choice blindness in-trial switch in order to differentiate the act of choosing from any direct comparison or association with what would be the alternative presented as correct feedback. In addition, the choice presented was always one that was not selected, in order to ensure that the feedback remained at odds with the actual preferences previously stated.

The experiment used female faces as choice blindness has been most commonly tested within this domain (Johansson et al., 2005, 2006, 2008; Pärnamets et al., 2015), thus ensuring that choice blindness can occur with the stimuli in question and providing a reliable baseline for the current experiment. As similarity and attractiveness are thought to affect the level of facial choice blindness (e.g., Sagana et al., 2013) we control for the two parameters within the experiment and data analysis. The procedure used in the experiment follows the computerised version of the choice blindness paradigm (Pärnamets et al., 2015) in order to maintain consistency with other literature. We did make one potentially consequential change in procedure in

[1] The prejudgement task asked participants to provide positive evaluations of the stimuli presented as opposed to unspecified or negative evaluations as comparative choice processes are likely to be contingent on the positive information in a preferential choice task (e.g., Mitsuda & Glahot, 2014), whereas in a sequential evaluation task there is a possibility that participants would attend to the negative aspects of the stimuli.

exposing participants to one as opposed to 36 trials, manipulating the feedback provided on every trial. This method was used as once participants experience a number of trials it is difficult to establish what carry over effects trust and visual memory interference may have on the manipulated trial.

Through this research, we aim to identify the boundaries of the situations under which people are susceptible to accepting false feedback as their own, by establishing at which stage of a choice process the stimuli become susceptible to being accepted as a self-reported preference. The failure to detect a switched choice at each level can help us identify the components that facilitate choice blindness, in turn narrowing down the potential range of mechanisms that underlie the phenomenon. Furthermore, this allows us to identify the conceptual and ethical concerns with how we view preference and decision making in our society if choice blindness does, in fact, extend to stimuli beyond those present in the current problem.

7.3 Method

The study took on a between subject 4 x 2 x 2 factorial design, with four levels of *interaction with the imposter stimuli* (novel, pre-exposure, pre-judgement & pre-choice), two levels of *'imposer' similarity* (similar or dissimilar to the choice pair) and two levels of *'imposter' attractiveness* (high or low). Choice blindness was measured by the proportion of trials in which participants reported noticing the switch. Detection was assessed at three stage of the paradigm – (i) *concurrent detection* (detected or not detected) which was measured through the justifications provided by the participants for selecting the 'imposter' feedback, with any mention of a change in face, mistaken selection or anything unusual coded as detected; (ii) *retrospective detection* (detected or not detected) which was measured by asking participants to indicate whether they noticed anything unusual and describe what they found unusual if the response was confirmatory, any mention of a change in face or mistaken selection was coded as detected and (iii) *informed detection* (detected, not detected, or uncertain) where participants were informed of the possibility of being shown the wrong face and asked to indicate whether they believed they experienced the switch by selecting the most suited response from three possible alternatives – yes, no, or unsure. For the purpose of analysis, an additional variable *overall detection* (detected or not detected) was created, coded as detected if the participant detected the switch either concurrently or retrospectively. Overall detection was used as the primary dependent variable, due to the inconsistent demonstration of sensitivity for concurrent and retrospective detection in past literature (i.e., sometimes people demonstrate higher detection concurrently and at other times retrospectively).

Materials

Black and white photos of female faces from The Psychological Image Collection at Stirling (PICS) online face database (pics.psych.stir.ac.uk) were used in the current experiment, to create stimuli in line with the original study. In order to control for similarity and attractiveness of the faces, ratings of the parameters were obtained through an online survey distributed to 350 people (173 female; mean age of 48.35; see Kirichek et al., submitted, for details).

A total of eight faces were selected for use across the four conditions. Four of the faces formed two 'core' pairs, each pair high on similarity and one of the sets high in attractiveness and the other low in attractiveness. The remaining four faces were designed to fulfil the role of the imposter stimuli, two of which were similar to the core images, and two dissimilar. Each of the similarity levels further contained one image high on attractiveness and one low on attractiveness.

Procedure

The experiment was carried out using the Qualtrics Online Survey Creator (Qualtrics, 2016), and participants were recruited using Prolific Academic recruitment platform (Prolific Academic, 2016). Four separate surveys were created for each of the conditions – novel, pre-exposed, pre-judgement and pre-choice. The surveys were then posted for participants to sign up to on the recruitment platform, with participation possible only once, creating a quasi-random allocation of people to the different conditions. At the start of each survey, participants were allocated to one of four stimulus type conditions, and one of four similarity-attractiveness conditions using a random number generator which determined which stimuli they will be exposed to.

Prior to the experiment participants were informed of the nature of the task, revealing as much as possible of the procedure without interfering with the experiment. They then had to confirm they understand their right to confidentiality and to withdraw at any time, and verify that they are satisfied with the information received and are happy to proceed. The different conditions varied in the procedure the participants underwent prior to the core choice blindness task, and in the face shown during feedback. For the novel choice condition participants were not presented with any additional information prior to the choice blindness task, and saw a face they had not been exposed to during the feedback condition.

For the pre-exposure and pre-justification conditions, participants were presented with six out of the possible eight faces, in accordance with the stimuli condition they belonged to. The set contained the four images varying in similarity and attractiveness, and one of the two core pairs. For the pre-exposure condition the participants were instructed simply to look at the faces presented on the screen, whereas in the pre-justification condition they also had to provide three positive characteristics of each of the faces presented. The seen core pair was then used for the choice phase of the study, and one of the remaining four images randomly allocated for presentation during feedback depending on the similarity-attractiveness condition the participant was allocated to.

For the pre-choice condition participants were required to complete the preferential binary choice task three times. The first choice was made between one of the core face pairs, the response to this choice was not analysed and only used so that the participants encountered six faces like they did in the justification condition, keeping the number of items in memory consistent. The second choice was made between a pair of faces as determined by the randomly allocated similarity-attractiveness condition. The last choice was made between the unseen core face pair. The face used during feedback was always the non-chosen alternative in the second choice task, in order to keep the procedure as close as possible to the original choice blindness study, and to avoid introducing a difference of retrospective preference plausibility. The attractiveness and similarity of the imposter face was coded accordingly with the face presented.

The core choice blindness procedure remained the same for all participants: as the last phase of the study all participants were given a binary choice task, where they selected their preferred face. After their selection, they indicated their confidence in their choice, partially as a measure of interest and partially as a distractor task in line with past research (see Pärnamets et al., 2015). After this, participants were presented with a face that was not one of the alternatives they saw in their choice, and required to provide a justification for selecting the presented image. Participants were then asked to indicate if they noticed anything unusual about the task, and if they answered yes were instructed to describe what it was that they found unusual. We informed the participants that the face they saw during feedback may have been switched, and asked them to indicate whether they believed that they experienced the switch by selecting between three options – yes, no and unsure. The participants in the pre-choice condition were also required to answer whether they thought the face shown was, in fact, from the most recent trial, to ensure all variations of switch detection were covered. After the experiment, all of the participants were fully debriefed on the nature of the experiment and purpose of the research, and given the details of the researchers to contact in case of any concerns. They were then provided with a code to authorise payment for their participation.

Participants

One hundred and ninety-nine people (72 female) took part in the experiment, with the average age of 30.12 (SD=9.29). Participant were recruited using the Prolific Academic online recruitment platform (Prolific Academic, 2016), with eligibility limited by age (18-80 years old) and

English ability (Fluent or Native). Three participants were excluded from the analysis due to incomplete responses that did not allow for detection to be assessed.

Fifty (19 female; mean age 31.84) participants took part in the novel condition, 52 (17 female; mean age 30.92) in the pre-exposure condition, 50 (21 female; mean age 31.14) in the pre-judgement condition and 44 (14 female; mean age 26.22) in the pre-choice condition. The difference in mean age between groups was significant (F (3,195) = 3.582, p = .015).

7.4 Results

An average of 51.3% of people reported noticing something unusual concurrently, 60.7% retrospectively (retrospective detection). In order to ensure all forms of detection are included in the analysis an overall detection variable was created, coded as detected (1) if the participants reported detecting the switch either concurrently or retrospectively, and non-detected (0) otherwise. The proportion successful overall detection was 65.1%.

Visual analysis of overall detection rates by experimental condition (see Table 1.) suggested that a lower proportion of people experience detection in the pre-choice condition, compared to other experiment types.

Table 1. Proportion of overall detected trials by experimental condition (no exposure, pre-exposure, pre-choice and pre-choice).

	Experimental Condition				Overall Mean
	No Exposure	Pre-Exposure	Pre-Judgement	Pre-Choice	
Overall Detection	72.0%	72.5%	72.0%	40.9%	65.1%

A logistic regression was carried out to predict the level of overall detection exhibited by participants using experimental condition, similarity, attractiveness and age as a continuous predictor variable to control for the difference in age observed between the conditions. Interaction terms were not included in the model, due to the limited sample size restricting the number of predictor variables to be used. A test of the full model against a constant only model was statistically significant, indicating that the full set of predictors is better at predicting when participants notice something unusual (χ^2= 14.089, p=.020, df=6). The Wald statistic revealed that the experiment condition is a significant predictor of the number of people who notice something unusual (Wald=10.456, p =.015, df=3). A by-category comparison using the novel choice condition as the reference category revealed a significant decrease in detection for the pre-choice condition (β =-1.327, p =.008), and no significant difference for pre-exposure (β =.042, p =.926) or pre-judgement conditions (β =.017, p =.970). Attractiveness (β =-.209, p =.539), similarity (β =-.009, p =.980) and age (β =.013, p =.434) were not found to be significant predictors of detection.

Informed Detection. After participants were informed of a potential switch in their choice an additional 19.0% reported that they experienced the switch and 6.7% were unsure (in addition to 65.1% overall detection). The proportion of people exhibiting informed detection or uncertainty did not differ significantly between experiment types (χ^2= 2.334, p = .504, df=3). When informed detection and uncertain individuals was combined with overall detection 90.8% of participants appeared to at least suspect the switch was possible, a proportion significantly lower than 100% that would be anticipated if no choice blindness had occurred (t (194) =-4.442, p<.001).

For the pre-choice condition participants were also informed that the face they saw as feedback was from a different trial, and asked to indicate whether they experienced this. Overall 26.2 % out of 44 participants noticed the wrong order and 26.2% reported being unsure. Whilst all participants that noticed the order switch also experienced other types of detection, 9 % of the participants were uncertain about the order even though they didn't report other types of detection.

Decision time. The average detection time was 6.21 seconds (SD=5.95). A general linear model using decision time as dependent variable and experiment type, overall detection and the interaction of the two as independent variables was significant overall (F (7,194) =4.051, p<.001). Experiment type was the only significant predictor (F (1,194) =7.701,

$p<.001$). The decision time was highest for no-exposure condition (8.05 secs), followed by pre-exposure (7.35 secs), pre-judgement (6.48 secs) and pre-choice conditions (2.47 secs).

Confidence. The average confidence was 7.76 (SD=1.90) on a scale of 1 to 10. A general linear model using confidence as dependent variable and experiment type, overall detection and the interaction of the two as independent variables was not significant overall (F (7,194) =.808, p=.582).

7.5 Discussion

In the current study, we investigated the type of interaction with a stimulus that is required for a participant to be susceptible to mistakenly recognising the stimulus as their own stated preference. In line with the original choice blindness study (Johansson et al., 2005), we presented participants with two female faces to choose from, and then asked them to justify their preference for a face that they did not select. However, instead of presenting participants with the non-selected alternative in their choice we presented them with a face that was either brand new to the participant; previously seen; previously evaluated; or a part of a previous choice. Our principal finding was that choice blindness can be successfully induced using stimuli external to the choice task. Furthermore, participants were significantly less likely to detect that the face they are shown is not the one they selected when the face was encountered as a part of a previous choice task, compared to when the face was previously evaluated, previously seen, or not encountered at all. The finding that people are very good at recognising whether they had previously encountered the alternative presented, indicates that recognition memory remains intact during the choice blindness paradigm, in-line with past literature (e.g., Pärnamets et al., 2015). On the other hand, the results also suggest that whilst some form of source memory failure (inability to recall which alternative was chosen) must occur for choice blindness to take place, to some extent source memory must remain intact in order for participants to recognise that the task for which they are receiving feedback, is different to the task in which the image presented was initially encountered.

The overall switch detection rate in the study was 65.1%, slightly above the previously reported range for choice blindness paradigms using facial stimuli (13%-63%; Johansson et al., 2005; Pärnamets et al., 2015). However, once participants are informed of the possibility that they might have been shown the wrong face as many as 91% of participants report detecting something unusual (83% detected & 8% uncertain). Nonetheless, the procedure used successfully established the presence of choice blindness using stimuli external to the task in question. It does, however, appear that the lower level of detection experienced by participants in the pre-choice condition is predominantly responsible for the significant levels of choice blindness observed overall. In fact, when participants that report being uncertain are coded as detected for informed detection, the levels of choice blindness observed for the novel and pre-exposed condition are not significantly different from 0% and for the pre-exposure condition the difference is only marginally significant.

It appears that a very small proportion of people fail to detect a mismatch between their intended stated preference and one presented during feedback, when that feedback consists of a stimulus that was not subject to a comparative choice process. It must be noted however, that it remains a concern that a significant proportion of people do not report identifying the switch when first requested to do so even for conditions seldom prone to choice blindness. Whether this is due to genuine error or demand characteristics, this lack of reporting the mismatch could have detrimental consequences for real life decisions. For example, if a third of people accept anything they are given as their preferred choice this poses problems for models of consumer choice (see Somerville & McGowan, 2016) regardless of whether they may have noticed that something is not quite right. This in turn highlights the importance of a robust method of preference elicitation in the study of preferential cognition.

For the pre-choice condition, however, the proportion of people exhibiting choice blindness (40.9% overall detection) does appear to reflect a limitation of human cognition. As with the classic choice blindness paradigm, this presents ethical concerns for malleability of choice and for our ability to monitor our own behaviour, and the environment conditions of the choice that lead to that behaviour (e.g., Johansson et al., 2005). There are two possible task characteristics that may be responsible for the low level of detection observed in the pre-choice condition in comparison with the novel, pre-exposed and pre-judged stimuli; the added comparative component of the choice task and the level of similarity between the task in which the stimulus is encountered and the task we are evaluating the feedback for.

As proposed in the introduction each stage of the decision process, which for the pre-choice condition is reflected in the addition of comparative component to the judgement stage, is accompanied by an increased depth in processing. This change in processing could in turn account for the increased levels in the pre-choice condition compared to the others increasing memory strength (Craik & Lockhart, 1972) and affective evaluation (Oppenheimer, 2008) of the imposter stimuli. For better remembered stimuli participants would be more likely to identify them as recently seen and therefore assess them to be a part of the recently undertaken choice task. For the stimuli with better affective evaluations (achieved through increased fluency), participants would be more likely to judge the stimuli as a likely candidate for what they preferred. This explanation is contingent on a plausibility judgement taking place when we are presented with information about our own choices, which is likely to occur due to reconstructive nature of source memory as proposed by Pärnamets et al. (2015). There is also a concern regarding the lack of gradual increase in choice blindness across the different conditions, which should be apparent if depth of processing is at least in part responsible as this should differ between each of the conditions, however, this could be explained by a detection ceiling effects, with the maximum level of detection being reached for the pre-judgement condition and thus all the decision stages preceding it.

On the other hand, if the comparative nature of the task is responsible for choice blindness we would not anticipate choice blindness to occur when the feedback is manipulated for a judgement rating and not a choice, yet a choice blindness equivalent has been consistently demonstrated in judgement task (e.g., Hall et al., 2012, 2013; Merckelbach et al., 2011; Sagana et al., 2014a). Although a valid criticism at face value, it fails to take into account that every study using judgement to demonstrate choice blindness, has done so by asking participants to select a numeric value from a fixed numeric scale as opposed to providing false verbal narrative for example. As a result, participant still have to undertake a comparative process between the available numerals when choosing the most appropriate thus approximating the choice element in the current experiment. The equivalent of the current experiment for such judgement procedure would be to present participants with feedback that lies outside the scale from which they selected their response. It would be reasonable to hypothesise that in this instance a similar pattern will be observed where a numeral outside the scale that is brand new or was simply shown to the participants earlier would result in high switch detection, whereas a number selected on a different scale prior to the task would result in relatively low detection.

The alternative explanation is that in this experiment the pre-choice task simply took on the form most similar to the task being manipulated. Although source memory has been found to be reconstructive in nature (Johnson et al., 1993; Yonelinas, 1999), source misattributions are more likely to occur when the source is similar (Johnson et al., 1988, 1993) which could account for why misattributions occurred for the stimuli from pre-choice conditions but not others. Further research using judgement tasks with imposter feedback originating from previously undergone choice task may shed light on which aspect of the pre-choice condition is responsible for choice blindness, however, for now it remains unclear whether a comparative process is what causes us to mistakenly accept erroneous feedback or whether the feedback being encountered in a similar environment may be sufficient.

The secondary measures within this study such as similarity, attractiveness, self-reported confidence and decision time failed to demonstrate any statistically significant, meaningful relationships with choice blindness. Although this is in part surprising given that all of the aforementioned have been identified as variables that affect choice blindness (e.g., similarity & attractiveness – Steenfeldt-Kristensen & Thornton, 2013; confidence – Sagana et al., 2013; decision time – Johansson et al., 2005), a number of other studies have also failed to find these effects (e.g., similarity & attractiveness – Johansson et al., 2005; confidence – Hall et al., 2013; time – Kirichek et al., submitted). The mixed findings regarding the effects of stimuli and task characteristic could be explained through other mediating factors-for example, time, confidence, similarity and attractiveness have all been found to contribute to choice difficulty, however, do not account for a hundred percent of its variability (Liberman & Förster, 2006). Perhaps when task difficulty reaches a certain threshold, the aforementioned variables stop having an effect, however, without explicit and precise knowledge of other factors this is merely a speculation, and the current paper is not sufficient to determine why similarity, attractiveness, confidence and time constraint appear to play no role in choice blindness when using this particular procedure.

On the other hand, decision time did appear to vary between the conditions, with the longest time taken to reach a choice for no exposure, and the shortest time for pre-choice condition. This in turn supports the notion that each proceeding task type shares more with the choice paradigm, as practicing the task beforehand should reduce the reaction time in the target task. It must be noted that the reduction in decision time would result in less exposure to the stimuli at hand which could cause a weaker memory for the presented alternatives and accordingly a lower level of detection. If this were the case we would anticipate a gradual reduction in detection for no exposure, pre-exposure and pre-judgement conditions, however we see an almost perfectly uniform pattern. Accordingly, it seems unlikely that decision time is driving choice blindness, especially since no significant effect on detection was identified.

Overall the study demonstrates that choice blindness can occur when stimuli external to the choice itself are presented during feedback as the selected option. This however, is only the case when the imposter stimulus is encountered in a past comparative choice setting, in turn ruling out novel, previously seen and previously evaluated stimuli as potential falsely accepted choices. We pose two possible explanation for the observed pattern; either the comparative process is crucial in giving rise to the choice blindness phenomenon, or that choice blindness occurs when the target task and feedback source are procedurally similar. Further research using a variety of different target tasks is necessary to establish the role of comparative processes in choice blindness.

Conclusion

Chapter 8.

Summary and Conclusions

Past research has shown that people often accept false feedback about their own characteristics and decisions they made in the past, even when that feedback contradicts the information they provided shortly prior to receiving that feedback. In this thesis, I explored two phenomena used to demonstrate such false feedback acceptance, the Barnum effect (Meehl, 1956) and choice blindness (Johansson et al., 2005), with the aim of identifying the conditions required for false feedback to be accepted, and in turn to create a false perception of one's choices.

Chapter 4 of this thesis was dedicated to investigating whether the Barnum effect, or the tendency to accept false feedback about one's personal characteristics, occurs for feedback about one's personality and risk attitudes, and whether this has a subsequent effect on people's perception of related choices they would make in the future. The Barnum effect was successfully induced for both personality and risk domains, with the pattern of false feedback acceptance demonstrating consistency with past research that suggests people are more likely to accept socially positive feedback, but not negative feedback (Johnson et al., 1985; Macdonald & Standing, 2002). However, there was no subsequent effects of receiving, or accepting false feedback, on subsequent choice, which is surprising in light of past research (Halperin & Snyder, 1979; Sakamoto et al., 2000).

In chapters 5 through 7, I focused on the choice blindness phenomenon, which demonstrates that people often fail to notice a mismatch between their indicated preference in a choice task and the outcome presented, when that outcome is in-fact switched for a different alternative. The aim of the experiments discussed was to identify which variables influence the likelihood of choice blindness being experienced, specifically number of alternatives presented in the choice, framing of the choice and justification of outcome tasks, and the level of prior interaction with the switched alternative. All three variable were found to have a significant effect on the proportion of people who experience choice blindness, however the effect of number of alternatives was dependent on the nature of stimuli used.

In this chapter, I discuss the implications of the research presented in this thesis to the wider literature on the subject. Building on the discussion in chapters 2 and 3, I consider how the findings contribute to our knowledge of when false feedback is likely to be accepted, how this may be relevant to the related research fields of introspective access, ability to detect errors, and preferential choice (specifically relevant for chapters 5 to 7), as well as how this contributes to our understanding of cognitive processes that determine whether the Barnum effect or choice blindness occur.

8.1 Variables that Influence False Feedback Acceptance

In chapter 3 of this thesis, I discussed a number of variables that appear to affect false feedback acceptance: the domain of the task for which the feedback is delivered, similarity of the real and altered outcome, favourability of the feedback, subjective confidence and demand characteristics. In addition, I outlined a few parameters that are specifically relevant to the choice blindness paradigm, including the number of alternatives in a choice, framing of the problem and familiarity with the stimuli presented. Here I revisit that discussion in light of the findings reported in this thesis, with consideration given to each of the aforementioned variables that are thought to impact false feedback acceptance.

Past literature suggests that the likelihood of people accepting false feedback about one’s characteristics, or choices, is dependent on the domain of the task for which the feedback is provided. For Barnum effect, people have been found to rate false feedback as less accurate when the measure from which it is thought to be generated is specific, and the methodology used to generate the profile is transparent (see Dickson & Kelly, 1985; Furnham & Schofield, 1987). For example, profiles from trait personality measures are less susceptible to the Barnum effect, compared to profiles generated using astrology (e.g., Wyman & Vyse, 2008). The likelihood of experiencing the

Barnum effect, therefore, appears to vary depending on the domain of the measure used. In the research presented in chapter 4, we investigate two distinct feedback domains; personality and risk attitudes. The personality measure used in the first experiment is based on the five-factor personality model (reporting trait openness, conscientiousness, extraversion, agreeableness. and neuroticism), and the effects of such feedback format on the Barnum effect have been tested in the past with somewhat mixed results (Andersen & Nordvik, 2002; Wyman & Vyse, 2008; Poškus, 2014). For example, Andersen and Nordvik (2002) conclude that participants successfully identify and reject invalid five-factor personality feedback, whilst Poškus (2014) reports successfully inducing the Barnum effect when providing false five-factor personality feedback. Here we provide support for the susceptibility of such personality feedback to the Barnum effect. That said this appears to only be the case when the feedback is socially positive, suggesting domain alone cannot determine whether the Barnum effect can be induced. In experiment two, I introduce a new domain into the Barnum effect literature, the risk-taking attitude measure. Overall the accuracy ratings given to the risk attitude feedback did not differ significantly from those for personality feedback, for real or altered scores (in aggregate or if compared by socially desirability of conditions), suggesting participants perceive the measures to be of similar accuracy for real and altered feedback types across the personality and risk attitude measures. We propose that when the feedback provided is based on real, validated measures, and augmented using the same procedure, the changes in accuracy ratings show little variation between domains. Whilst for positive feedback the reported accuracy ratings are similar to those for real feedback, negative feedback is perceived as neither accurate nor inaccurate.

Similarly, for choice blindness, the domain in which the choice is made can impact the likelihood of people accepting false feedback. For example, Somerville and McGowan (2016) report a much higher level of detection for brand chocolates compared to faces. In this thesis, however, we only explore choice blindness in one domain, namely preferential choice for female faces, a domain most prominent when it comes to studying the choice blindness phenomena. The average detection of false feedback reported in each experiment presented in this thesis (47-65%) is slightly above the range of detection reported in past literature on choice blindness using faces (12-63%), however this is likely due to the difference in dependent variables used between the work presented and past literature. I combined concurrent and retrospective detection for the purpose of my analysis and in the past concurrent detection has been the conventionally reported measure. Whilst this complicates between study comparisons, it appeared to be a necessary precaution to account for the unusual pattern of results observed in my thesis: higher levels of detection observed for concurrent detection compared to retrospective in two out of three studies focusing on choice blindness. I suspect the pattern is a result of the study being carried out online, preventing the experimenter from clarifying what would be deemed as unusual. In turn, if participants believed that the switch resulted from their own mistake or a system error they might perceive it as irrelevant.

Examination of concurrent detection in isolation does appear to lie within the previously reported range at 40.1-56.5% providing additional evidence that the likelihood of experiencing choice blindness in facial preference tasks, lies within a certain range. Whilst other domains are not explored here, one could expect the rate of choice blindness to be different for other domains. However, although the baseline may change, the systematic changes observed in the experiments presented here would also impact choice blindness in a similar manner for other areas of paradigm applications. Specifically, it is proposed that the impact of similarity, favourability, framing, interaction with the switched stimuli, and potentially time and confidence would alter choice blindness in the same direction as observed here, across different domains.

The most intuitively relevant factor to the acceptance of false feedback is the difference in favourability between the outcome one would expect, and that presented. The research on Barnum effect has often found that social positive feedback is more readily accepted than negative feedback (e.g., Johnson et al., 1985; Macdonald & Standing, 2002; for exception see Dmitruk et al., 1973). The results discussed in chapter 4 appear to support this finding. For experiment one, participants were more likely to accept personality feedback when openness, conscientiousness, extraversion and agreeableness trait scores were increased and neuroticism trait scores were decreased compared to the inverse. Whilst the feedback was designed to more closely resemble the traits of a person thought to be more (vs less) likely to volunteer for psychology experiments, this exact pattern has been found to be seen as more positive (vs negative) by Poškus and Žukauskienė (2014; in Poškus, 2014). Similarly, for experiment two participants were more likely to accept feedback associated with low risk taking behaviour, compared to high risk taking behaviour, which is often evaluated more negatively (Alhakami & Slovic, 1994), and is potentially associated with socially negative traits such as gambling, or promiscuity.

For choice blindness, the nature of the paradigm makes it difficult to assess how changes in favourability between the selected alternative and the presented choice impact acceptance of incorrect feedback, because by the very nature of choosing one alternative participants indicate that they prefer it the most. This makes it impossible to switch the chosen alternative for a better one, since the best alternative is always selected. In chapter 5 of this thesis, we try a different approach, by varying the difference in attractiveness of the face alternatives presented. We find that when the difference is higher, making the false feedback less favourable compared to the actual choice made, a higher proportion of participants detect that mismatch between choice and outcome compared to when the levels of attractiveness are similar however this fails to reach significance. The only exception to this observation is when participants are faced with a choice of three alternatives and the less attractive face is also dissimilar visually. Accordingly, it is difficult to ascertain the role of favourability on choice blindness. Whilst in some specific scenarios it appears that favourability of the feedback presented does affect the likelihood of that feedback being accepted as accurate, for choice blindness as well as the Barnum effect, the effect is not present throughout.

Similarity of the false feedback to the real feedback is another factor anticipated to impact false feedback acceptance. If the two alternatives are so similar people cannot tell them apart, it would not be surprising that people would rate them as equally accurate. Imagine, for example, completing a personality test where the real scores suggest you scored high on trait extraversion, say 80 on a scale of 1 to 100. If you are then presented with feedback saying you scored 81, it is highly likely you would not notice the discrepancy, however, if the feedback is the inverse of your score at 20, one would suspect that detection would be very likely. Indeed, Andersen and Nordvik (2002) demonstrated that participants rate false feedback as less accurate when that feedback is dissimilar to their real personality profile. In the experiments conducted in chapter 4, the feedback was transformed from a scale of 1 to 100, to either a scale of 1 to 50, or 51 to 100, therefore we cannot look at the effects of difference between real and false feedback without introducing real personality as a confounding variable. The results pertaining to risk attitude acceptance do, however, suggest that similarity plays a crucial role in the acceptance of altered information, with people demonstrating stronger risk preference more likely to accept feedback when it is altered to suggest high risk preference and less likely to accept feedback altered to suggest lower risk preference. On the other hand, the slight increase in accuracy ratings when participants received socially positive feedback compared to real feedback, suggest that the deviation from real profiles is not responsible for differences in the accuracy ratings. It is likely that there is an interaction between favourability and similarity, where a person is more prone to accepting the feedback if it is either favourable or similar.

For choice blindness, varying the similarity of alternatives used has received somewhat mixed results. Whilst the majority of research has reported significant effects of similarity (Hall et al., 2010; McLaughlin & Somerville, 2013; Sagana et al., 2013; Sauerland et al., 2013; Steenfeldt-Kristensen & Thornton, 2013), a few studies failed to find such effects (Johansson et al., 2005; Sauerland et al., 2014). In chapter 3, I discussed the possibility that a lack of a robust measure of similarity may have been responsible for the lack of similarity effect reported in the first choice blindness experiment (Johansson et al., 2005). To ensure that the manipulation of similarity in experiments presented in this thesis is representative of the population, we used a large sample of 350 raters to establish the characteristics of the facial stimuli used (see chapter 5). However, despite the same method of establishing similarity being used across the experiments presented in chapter 5 through 7, similarity only appeared to be a significant predictor of choice blindness in the experiments presented in chapter 5. Furthermore, the stimuli assessed on similarity included those used by Johansson and colleagues (2005) in their experiment, and we found that the ratings we collected corresponded with the similarity categories originally assigned to the face pairs, indicating that the procedure of allocating faces to different similarity conditions could not be responsible for the lack of difference reported. It is therefore proposed that similarity may contribute to a higher level cognitive representation, such as ambiguity, which in turn determines the likelihood of choice blindness occurring (Sagana et al., 2013; Somerville & McGowan, 2016). Once this variable reaches a certain threshold, the contribution of similarity ceases to have an effect because even when the stimuli are dissimilar the situation remains ambiguous. Consider for example the experiment presented in chapter 6, where the negative framing of the choice or justification task are found to decrease the proportion of manipulated trials that are detected. It is likely that negative framing impacts the same 'ambiguity' factor as similarity, resulting in no effect of similarity on choice blindness being identified. Indeed, when the positive framing condition is examined in isolation, a trend begins to emerge of higher detection when stimuli are dissimilar (81%) compared to when they are similar (66%), although this does not reach significance, perhaps due to the small sample of people in this subset of the data.

Whilst I did not investigate the level of ambiguity with respect to the Barnum effect in the original research presented here, the very origin of the phenomenon suggests that some ambiguity factor does indeed play a role in determining whether false feedback is accepted as accurate. As I outlined in previous chapters, the Barnum effect was initially considered to be a result of information being general and applicable to a wide range of other people (Forer, 1949; Sundberg, 1955), such lack of specificity in itself creates ambiguity. Direct comparison of generic feedback such as that used in astrology, compared to trait personality measures, further confirmed that acceptance of feedback about the self is higher when the feedback is general (e.g., Wyman & Vyse, 2008), suggesting that increase in ambiguity increases the likelihood of false feedback acceptance.

If decision ambiguity is indeed the determining factor in whether false feedback is accepted or not, we would expect this to be reflected in participants' subjective confidence in their decision. As outlined in chapter 3, however, the research on the relationship between confidence and false feedback acceptance has been scarce. I am not aware of any such studies being carried out for the Barnum effect, but a few have attempted to understand the relationship of self-reported confidence and choice blindness. One study by Sagana and colleagues (2013) reports the post-decision confidence to be a significant predictor of choice blindness. On the other hand, when examining the effects of self-reported certainty of political views on subsequently altered responses to political questions, Hall et al. (2013) fail to report a significant relationship. Similarly, I fail to find a relationship between confidence and choice blindness in the experiments presented in chapters 5 through 7 of the current thesis. This poses an interesting dilemma, since we know that factors contributing to ambiguity also contribute to choice blindness, yet this is not reflected in self-reported confidence. One possibility that could explain such results is a floor effect of confidence in preferential choice, where no participants exhibiting high confidence levels would limit differentiation between detected and non-detected trial. However, since the average choice confidence observed in my research (chapters 5-7 ranged) from 7.2 to 7.8 on a 0 to 10 scale, we can reject such hypothesis. Another explanation for this discrepancy could be that subjective confidence and actual level of ambiguity are not as closely linked as we intuitively suspect. After all, confidence ratings have often been found to be poor predictors of decision quality (e.g., Tversky & Kahneman, 1975). If this is the case, however, we are faced with a different challenge of finding a direct measure of what it means for a choice to be 'ambiguous'.

So far, I have discussed variables that have been investigated with respect to both the Barnum effect and choice blindness paradigm. There are a number of variables considered in my research, which I will now discuss, that are more relevant for the choice blindness procedure, specifically number of alternatives presented in the choice task, positive and negative framing of the task, and the prior interaction with the stimulus used in the switch. Whilst the Barnum effect will be mentioned where relevant, it will be seldom discussed in the remainder of this section.

In chapter 3, I proposed that the number of alternatives presented in a choice might impact choice blindness. For example, when asking participants to recognise a face from a line-up formed of six alternatives, and switching the selected face for a different alternative in a subsequent justification task, Sagana et al. (2013) reported a higher detection rate of the switch compared to other choice blindness studies using facial stimuli. I proposed that this may be a result of methodological deviation from the original paradigm, since the study gave participants a task with an objectively correct outcome, to recognise an earlier encountered individual. However, in chapter 5 we report that increasing the number of alternatives can indeed increase the level of switch detection, which could explain the results reported by Sagana and colleagues. Whilst overall, the detection for three compared to two alternative tasks was higher, we must be cautious when interpreting these results since such effect was only observed when there was a variation in physical similarity and attractiveness of the faces used. In chapter 5, I propose that the mechanism underlying the observed phenomenon is the enhanced salience of differences between the alternatives, this however, does not negate the pattern observed.

It remains unclear why the difference in choice blindness for binary and ternary choice only occurs when both similarity and attractiveness are varied, but not one or the other. As noted earlier in this chapter, for binary choice variability in similarity alone is sufficient to produce a change in proportions of people who detect wrong feedback being presented. Whilst this is also true for ternary choice, a combination of variation in the two parameters seems to produce a much larger change in detection. This is potentially a result of the difficulty in creating ternary stimuli sets equally spaced across the similarity and attractiveness dimensions, for example for stimuli that are designed to be equally similar and attractive one pair within the three will always be slightly more similar than the other unless a perfect level of control is achieved through artificially creating stimuli. Whilst the reason for this pattern requires further research, it does allow us to predict under which circumstances detection of erroneous outcomes can be increased, and the results highlight that extrapolating the findings of binary choice to multi-alternative choice should be done with caution.

In chapter 2, I briefly discussed research evidence on how framing, or more specifically the response modality of the question, may impact choice blindness, concluding that whilst studies using judgement rating demonstrate slightly lower levels of detection compared to the original paradigm, further research directly comparing judgement and decision tasks whilst controlling for other factors is required. In chapter 6, I explored another form of framing variation, positive versus negative question frame that is presented to participants. Past research has demonstrated that choice outcome (Shafir, 1993), as well as choice consistency (Kogut, 2011), vary with the manner in which the task instructions are formulated. More specifically a study by Kogut (2011), reported that higher choice consistency can be observed when participants are provided with negatively framed tasks (i.e., please select the options you would like to reject) compared to a positively framed task (i.e., please select the option you would like to keep). Surprisingly, I found that participants exhibited the highest levels of detection when both the choice and justification elements of the choice blindness task were framed positively, which coincidentally was in line with the original procedure of the choice blindness paradigm (Johansson et al., 2005). Detection rates for participants who underwent the task with at least one negative frame (i.e., were required to select their preferred face and required to justify why they did not prefer it, or required to select the face they did not prefer regardless of the justification frame) were significantly lower compared to when the original procedure was used, but showed no significant difference between each other.

There are a number of possible differences in the research approaches that could explain this difference. First, it must be noted that the definition of choosing and rejecting framing is very different in the experiment presented in chapter 6 and the research conducted by Kogut (2011); whereas we asked participants to either select their preferred or non-preferred alternative, in keeping with the choice blindness paradigm, the choice consistency researchers asked participants to include their preferred alternatives or exclude their least preferred alternatives out of a host of possibilities. Although both sets of research were designed to measure the effects of choosing versus rejecting alternatives, the difference in the wording of the tasks themselves could produce a difference in the choice outcomes observed. Additionally, the research presented in chapter 6 used binary choice with one outcome, whereas Kogut (2011) used multiple alternative tasks that required participants to narrow their selection in stages. The decision domains themselves could have determined how framing can affect choice consistency in different ways. Alternatively, a possibility remains that the cognitive processes involved in choice consistency and outcome mismatch detection are, in fact, different, and the choice consistency work simply measures something different to the choice blindness paradigm. The reasons for the observed discrepancy can only be theorised, until further empirical work can establish the effect of all of the aforementioned variables, so for the moment, we can only conclude that probability of switched outcome detection is at its highest when both the choice and feedback tasks are positively framed.

In chapter 3 of this thesis, I discussed how familiarity with a domain can reduce choice blindness, as demonstrated by Somerville and McGowan (2016) who applied the choice blindness paradigm to preferential choice decisions with familiar chocolate brands, demonstrating a much lower level of accepting switched feedback. The last original paper in this thesis (Chapter 7) explored the effects of another type of familiarity, which was manipulated within the experiment itself. This was achieved by presenting false feedback that consisted of pictures of faces not encountered during the actual choice task for which the feedback is being provided, and varying the levels of prior interactions with that face. The results were somewhat reassuring for our ability to monitor the validity of feedback provided for our choices, in that we found that a very high proportion of people (more than 85%) detecting that their choice had been switched if they are presented with an alternative that they had never seen before, or only encountered in a different task, such as only studying the alternative or providing evaluative comments for it when presented amongst a range of other alternatives. However, we did find that about a third of participants who encountered the face as part of a preceding choice task failed to detect that their choice had been switched, suggesting choice blindness can be induced using alternatives outside the task as feedback. The implications of this finding are two-fold: first it demonstrates that the range of false feedback people may accept is wider than previously anticipated, and second it highlights the importance of source memory similarity for the choice selected and choice presented as feedback in choice blindness, which will be discussed in further detail later in this chapter (section 8.3). Another important contribution of this experiment is that it provides us with a new procedural approach to choice blindness that can now be used to explore a wider range of circumstances, such as whether detection of a manipulated outcome can be reduced when the feedback provided is, in fact, better than the choice made.

The last variable I would like to consider, as a determining factor of false feedback acceptance is demand characteristic, such as investigator effects (see chapter 3). Such effects are of some concern for all psychology experiments, as it is hard to determine whether people are responding with their genuine thoughts, or with what they think the experimenter expects from them (Orne, 1962). For example, some people may be embarrassed to suggest the experimenter has made a mistake when presented with incorrect feedback, or alternatively if they have received accurate feedback from the experimenter in the past they may develop trust and assume that the experimenter's judgement is better than their own, even when they suspect something is not right. Interestingly, neither the Barnum effect (Orpen & Jamotte, 1975; Snyder & Larson, 1972), nor choice blindness (Sauerland et al., 2013, 2014) appear to be affected by individual differences in susceptibility to social desirability as measured by the Marlowe-Crowne Social Desirability scale. However, it remains possible that the propensity to demand characteristics in the tasks used is so high that even the participants less susceptible to social desirability exhibited the effects. Such effects can be minimised by eliminating direct interaction with the experimenter, and ensuring there is no past experience which can impact their judgement.

Whilst I do not directly manipulate the factors associated with demand characteristics in any of the studies presented, I have taken some steps to minimise the chances of such circumstances having an effect. For the experiments outlined in chapter 4, investigating the Barnum effect, the data was collected online without the experimenter present, minimising potential experimenter effects and suggesting the effect is a real cognitive phenomenon as opposed to the participants' response to perceived expectations. Other research using computers to supposedly generate the feedback has also consistently reported successfully inducing the Barnum effect (Baillargeon & Danis, 1984; Guastello & Rieke, 1990; O'Dell, 1972). Furthermore, in experiment one I manipulate all of the trait scores presented, preventing participants from learning to treat feedback as accurate. Whilst only one trait is manipulated in experiment two, any effects resulting from the lack of manipulation of the other three traits presented would have been visible in comparison to experiment one.

Similarly, in the experiments using the choice blindness paradigm (chapters 5 through 7) I employ an online procedure with a one-shot decision to investigate choice blindness to minimise experimenter effects. To my knowledge past experiments have always presented false feedback amongst other trials which provided real feedback, and this is the first time such procedure has been employed. Whilst I do appear to demonstrate higher detection rates compared to the range previously reported, this change is inconclusive without conducting research that can compare these variables in otherwise controlled empirical research. Generally, my findings further support the conclusion that social desirability cannot account for choice blindness, demonstrating that it is indeed a robust, stable cognitive phenomenon.

Up to this point, I have discussed the variables that determine the likelihood of false feedback acceptance, and the contribution of the original papers presented in this thesis to the understanding of these variables. In summary domain, similarity, in some instances favourability and specificity appear to be significant predictors of when false feedback acceptance occurs for both the Barnum effect and choice blindness. For choice blindness specifically, I report that the number of alternatives, framing of the choice and justification tasks, and the type of interaction with the feedback presented prior to justifying it have a significant effect on false feedback detection. Interestingly, subjective confidence and demand characteristics appear to have little to no relationship with acceptance of false feedback. In the following chapters, I will briefly discuss the contributions of the work presented here to our understanding of the consequences of false feedback effects (specifically the Barnum effect), as well as to our understanding of the mechanisms associated with accepting false feedback. Lastly, I will discuss the broader implications of the work for introspection, error detection and preferential choice, as well as the possible practical applications of the research presented.

8.2 Consequences of False Feedback Acceptance

Past research has demonstrated that accepting false feedback can actually impact our decisions and behaviours, making them more congruent with the information provided. Whilst I do not explore such effects of false feedback acceptance with respect to choice blindness in the original experiments presented here, as discussed in chapter 3 experiencing choice blindness has been found to alter participant response (Johansson et al., 2014; Kusev et al., in preparation; Merckelbach et al., 2011). For example, this has been demonstrated using the choice blindness paradigm to switch the preferences for female faces (Johansson et al., 2014), with findings indicating that we are more likely to select a face we did not initially prefer as our preference, after being presented with it as false feedback and providing a justification for why we chose is. Similarly, the perceived intensity of psychological symptoms has been found to change, after being presented with an adjusted version of the initially reported intensity, in the direction of that adjustment (Merckelbach et al., 2011). In other words, after accepting altered feedback, participants are more likely to report heightened symptom severity if they see feedback suggesting increased intensity, and milder symptom intensity if the feedback if they see feedback suggesting lower intensity.

In the research presented in chapter 4 of this thesis, I examined whether the Barnum effect can induce a similar change, and failed to find an analogous effect. To my knowledge, there have been two studies in the past which tried to investigate this before, with a different conclusion (Halperin & Snyder, 1979; Sakamoto et al., 2000). Halperin and Snyder (1979) found that that giving participants feedback that suggests a high propensity to change resulted in more positive outcomes of snake phobia treatment. Sakamoto et al. (2000) reported that giving people feedback that suggests higher trait extraversion resulted in more conversation with a confederate, and better impression of the confederate by the individual and vice versa. Both studies, however, suffered from a range of limitations. First, both used female only samples, a demographic shown to accept feedback more readily (compared to males; Layne, 1998). Second, neither used an appropriate control group, of real or neutral feedback, making it difficult to establish the changes induced by feedback type alone. Lastly, in both studies the feedback was delivered by the experimenter as prose, which could have result in demand characteristics, as well as motivational messages implicitly contained in the feedback.

My decision to carry out another experiment exploring consequences of the Barnum effect was motivated by these limitations, trying to provide a more controlled, albeit less ecologically valid procedure, using a gender balanced sample for the first time. Chapter 4 outlined two experiments: the first altering psychological profiles to assess their impact on participants' self-reported likelihood to volunteer for psychology experiments, and the second altering risk attitude profiles to assess their impact on preference for risky versus certain lotteries. Neither showed a relationship between the type of feedback provided and subsequent responses, regardless of whether participants rated the feedback as accurate or not. It is impossible to identify the precise reasons for why my results were not aligned with what would be predicted by Halperin and Snyder (1979) and Sakamoto and colleagues (2000). The only explanation we can rule out is gender since we detected no difference between male and female participants in feedback acceptance or self-reported behavioural propensities.

One possible explanation is the reduced susceptibility of the participants to potential experimenter effects, as for the first time the experiment was carried out online and did not require direct contact with people involved in the experiment. Another, is that the numeric feedback format used in the experiment may have reduced implicit encouragement cues that may be associated with prose. Alternatively, it is possible that the difference lies in the nature of behaviours measured, whether it is the difference in the domain tested or the dissociation between real and self-reported behaviour. All I can conclude, is that experiencing the Barnum effect does not always lead to change in behavioural propensity, and thus that accepting false feedback does not necessarily update our beliefs about the self. This finding is reassuring as it suggests that our behaviour is not always shaped by errors in our environment, yet poses a new and complex task of establishing when false feedback does or does not alter behaviour.

8.3 Mechanisms

In chapter 3, I discussed the mechanisms that might underlie the Barnum effect and choice blindness. Here I will revisit the discussion, focusing on memory failure, and self-perception theory, and ambiguity with regard to their ability to explain and predict when false feedback is accepted

One crucial question, which needs to be considered when explaining why false feedback acceptance may occur, is why we do not simply recall past actions to assess the accuracy of the feedback. In chapter 3 I outlined that for choice blindness it is puzzling why we cannot simply compare the outcome of a choice task, to the feedback presented in the same format. For the Barnum effect, however, since participants may not be able to directly envision the relationship between the task they complete and the feedback generated, the question is slightly different in that we need to understand why people do not use knowledge of their past behaviours to assess the likelihood of the information presented.

Johnson et al., (1985) proposed that people do exactly that, but because of the heuristics, or shortcuts, used to access our memories, the outcome is the erroneous acceptance of false information as accurate (Kahneman & Tversky, 1973). The researchers suggested that because we have a large memory store about ourselves, when faced with general enough feedback we are likely to find available behavioural evidence confirming any common trait in ourselves. Furthermore, biases in memory can provide a reasonable explanation for the higher level of Barnum effect for positive compared to negative feedback about the self, as people tend to pay more attention to positive feedback and remember it better, thus providing more instances that come to mind for positive information, even when those instances are not common (Sedikides et al., 2004). Whilst this theory is very strong in explaining the acceptance of common, or general traits, it is questionable to what extent it could be applied to specific, numeric feedback. For such feedback formats, in providing the scale of possibilities, and not verbal information, we automatically provide people with a comparison point which would be likely to draw attention to the fact that we are on one scale and not the other, which may result in memory retrieval for both scenarios. Presumably, this would prompt someone who receives contradictory feedback, to recall the instances to the contrary. The findings presented in chapter 4 provide further support that Barnum effect can occur for specific, numeric feedback, suggesting the validity of the availability heuristic should be re-visited and further empirical research may be necessary to understand whether a limited scale can influence acceptance of invalid information. On the other hand, we only observe high accuracy ratings, when the feedback provided is positive which is in line with what one would expect if the Barnum effect was a result of the availability heuristic.

Positive feedback being rated as more accurate than negative feedback not only provides support for the memory bias approach put forward by Johnson et al. (1985), it can also be taken to support the self-serving bias explanation (Macdonald & Standing, 2002) of the Barnum effect. The self-serving bias approach proposes that we are more likely to attribute positive traits to our own nature, whilst negative traits are attributed to external factors, such as the validity of the psychometric tool used to generate the personal description (see also Collins et al., 1977; Snyder & Shenkel, 1976). This approach, however, fails to explain why acceptance of negative feedback is often documented in past literature (e.g., Poškus, 2014; Wyman & Vyse, 2008), making the availability heuristic approach more comprehensive in explaining observed phenomena.

Neither the self-serving bias, nor the memory bias approach can sufficiently explain choice blindness as the memory for the encountered stimuli should not differ, and by the very nature of the feedback presented being the non-preferred alternative the feedback is never more positive than the actual choice. Whilst it is possible that different false feedback effects have different underlying mechanisms, it seems improbable that these are completely distinct. Or at the very least this is unlikely to be the most parsimonious explanation of the Barnum and choice blindness phenomena.

For choice blindness, the memory failure necessary for accepting false feedback is even more surprising, since we can remember complex autobiographical information yet fail to recall a simple choice we made just minutes before judging the validity of an outcome of that choice. Research suggests that only the source memory, or the choice made, appears to be unavailable yet the recognition of the stimuli in the task remain intact (Pärnamets et al., 2015; Sagana et al., 2014a). The findings reported in chapters 5 and 6 of this thesis also support the conclusion that a recognition memory limitation explanation of choice blindness is unlikely. First, if cognitive resources are already under strain, it would follow that increasing the amount of information to encode and store would lead to increased difficulty in monitoring such information and thus higher choice blindness. Yet, chapter 5 shows the opposite effect by demonstrating that increasing the number of alternatives increases the level of detection. Second, if memory constraint is responsible for choice blindness, using a better remembered stimulus with incongruent task instructions as feedback would increase the chance of detection as participants would be more likely to recall the various features of their choice. In preferential choice, the selected stimulus is more likely be remembered (McClelland et al., 1987), yet when we use the selected alternative with incongruent justification instructions (i.e., Why did you not select this alternative?) as the post-decision task in chapter 6 people are more likely to experience choice blindness compared to when they are presented with the unselected alternative with incongruent instructions.

Pärnamets et al. (2015) proposed that whilst recognition memory remains intact, the evaluative and reconstructive processes involved in accessing source memory (Johnson et al., 1993; Yonelinas, 1999) make it more susceptible to misremembering. Chapter 7 of this thesis does suggest, however, that the similarity of the information source is crucial in determining the level of detection exhibited by participants. More specifically, when instructed to justify a previously non-chosen alternative, participants are less likely to notice the mismatch if that alternative was previously encountered in a visually and cognitively similar context, such as a preceding binary choice. On the other hand, when the alternative presented is encountered in a different task such as evaluation of the item on individual basis, the proportion of people who detect the mismatch is substantially higher. This supports the notion that the assessment of which choice was previously made involves accessing some information from memory of the event, accompanied by situational judgements of event likelihood as proposed for source memory overall (Johnson et al., 1993).

Pärnamets et al. (2015) stipulated that the tendency to use external information to judge our own behavioural propensity is best described by self-perception theory (Bem, 1967). Self-perception theory states that when internal information is weak, individuals come to know their own attitudes by observing their own behaviour and environment. The application of self-perception theory to the Barnum effect and choice blindness is discussed in detail in chapter 2, with the general conclusion that self-perception can describe the possible mechanisms behind choice blindness and the Barnum effect. However, the theory lacks the specificity required to be rigorously tested using empirical data. Support for the application of self-perception theory comes from the very demonstration that false feedback is often accepted, whether in the context of the Barnum effect or choice blindness. The paradigms highlight that whilst in some instances, people use internal information to successfully reject false feedback about the self, whilst in others they use external cues to infer their own attitudes and behaviour which can lead to accepting false information (e.g., Johansson et al., 2005). By demonstrating false feedback acceptance across a wider range of situations than before (chapters 4 through 5), we also support the existence of processes akin to those described by self-perception theory. Further support that we use knowledge of past actions to infer our own attitudes and propensities comes from literature demonstrating that accepting false feedback can in turn lead to behavioural change (Halperin & Snyder, 1979; Johansson et al., 2014; Merckelbach et al., 2011; Sakamoto et al., 2000; see chapter 3 for discussion).

Chapter 4 of this thesis, however, presents findings that do not support this pattern, demonstrating that altering participants' personality or risk attitude profiles does not alter subsequent self-reported behavioural propensities for the related fields of willingness to volunteer for experiments, and preferences for risky versus certain lotteries, even when they rate them as accurate. This poses the question of why in some instances we update our beliefs about the self and act accordingly and in others we will fail to do so. It could be argued that in these particular domains the internal information is strong enough to override the external information received, yet this still doesn't explain how the feedback is accepted in the first place whilst the actions do not change. Without a better definition of strength of internal information, and an appropriate measure, self-perception theory can only have weak predictive power.

Another problem for the application of self-perception theory arises when we consider the variability in choice blindness exhibited across different domains and procedural variations, as this requires the use of internal cues to explain such differences. For example, what reasons would an observer have to reach a different conclusion about a person's behaviour when they are making a positively framed choice versus negatively framed choice? (Chapter 7). Logically the memory of past choices made should equally inform preferences and lack of preference when making an attitude attribution, yet this does not appear to be the case. Similarly, it does not explain why the number of alternatives presented would affect choice blindness (chapter 5). In order to explain this, we would need to consider the cognitive mechanisms concerned with making any judgement, whether it concerns the self or someone else, which still remain to be incorporated into a more comprehensive model.

The possibility that an observer would reach a different conclusion under different tasks can be captured in the task ambiguity. Recall that ambiguity plays an important role in both choice blindness and the Barnum effect (e.g., Forer, 1949; Merckelbach et al., 2011; Sagana et al., 2013; Sundberg, 1955). As defined in chapter 3, ambiguity refers to the imprecision of the task and its environment which allows the information to be interpreted in more than one way, be it a choice outcome or personal description (Sloman, Fernbach & Hagmayer, 2010). In other words, for false feedback effects this is our ability to explain both the real or false outcome presented. Ambiguity can be affected through different properties such as generality, similarity, and time constraints, and is thought to be, at least to some extent, related to the construct of subjective confidence (e.g., Ellsberg, 1961).

For the experiment on Barnum effect presented in chapter 4, I used feedback that is presented in a specific manner (numeric representation of a position on trait scale) to induce the Barnum effect, using validated measures; IPIP five factor personality measures (IPIP, 2006) and the Domain Specific Risk Attitude Test (Blais & Weber, 2006). Nonetheless, overall participants appeared to accept false feedback, suggesting that even for some of our most reliable psychometric tools that have been designed to minimise ambiguity, the Barnum effect still occurs. On closer examination, this only appears to be the case for socially favourable feedback, a pattern that cannot be explained by ambiguity alone. Furthermore, as the acceptance rate for positive feedback was even higher than for real feedback, it also appears that for this condition decrease in similarity does not result in enhanced detection. It seems that whilst minimising ambiguity eliminated acceptance of negative feedback, a different mechanism is likely responsible for acceptance of positive feedback.

On the other hand, where subjective judgement is involved, there is always some level of ambiguity, which may be the necessary minimum for people to accept false feedback, explaining why the Barnum effect can and does occur for measures that appear highly specific. Sagana et al. (2013, 2014b) propose that the same applies for choice blindness in preferential choice. Indeed, we find that some level of choice blindness can be observed across every experimental condition presented throughout chapters 5 through 7, with the lowest acceptance rate of 12%, when the image used as false feedback is completely new to the participant. The low level of familiarity with choosing between faces is also likely to have played a role in introducing ambiguity which allowed for choice blindness to be observed (discovered preference hypothesis, Plott, 1966; constructed preference hypothesis, Lichtenstein & Slovic, 2006; Slovic, 1995), however, since we only use female faces as the choice domain in the experiments presented here this is not empirically tested.

As discussed in the previous section we also find evidence of other factors related to ambiguity such as similarity, favourability of outcome, and framing of the task having an impact on the level of detection observed. Similarity and favourability were controlled in all three choice blindness experiments presented here, however, only one of the studies found a significant effect of these variables. As one would expect, detection was lower for stimuli that were visually similar and of a similar favourability level, however, this was only found to be the case in the experiment presented in chapter 5. We propose that the effect is not observed due to being precluded by other factors that influence choice blindness in chapters 6 and 7, namely the high ambiguity associated with the framing used in chapter 6 and the source of feedback being the choice blindness determinant in chapter 7. The sample size does not provide high enough power to examine the effects of similarity and favourability within the individual conditions of the two experiments.

The focus of chapters 6 and 7 was to determine whether the framing of the choice task (positive versus negative), and the interaction with the stimulus presented during feedback (completed novel, previously seen, previously evaluated, or a part of a choice task distinct to the one for which the feedback is being presented) impact choice blindness. Both studies reported a significant effect of their respective manipulated variables, both of which could be construed as changing the ambiguity of the task at hand. The manipulation of task framing demonstrated that the highest proportion of detection is reported when the choice task (please select the face you prefer.) and the justification task (why did you prefer this face?) are both framed positively, whereas when either, or both, tasks are framed negatively the level of detection observed is significantly lower. Research has found that people find it harder to process sentences containing negative particles (e.g., not) than positive, finding them harder to judge (e.g., Carpenter & Just, 1975), and therefore more ambiguous, perhaps as a result of rarely having to identify the least preferred alternative compared to most preferred alternative. Accordingly, it is likely that changing the framing of the choice blindness task from positive to negative decreases detection by introducing additional ambiguity.

For chapter 6, it is possible that when the participants have never seen the stimulus presented as false feedback for their choice, or encountered it in a different setting, they recognise that the item presented was not a part of the choice set and therefore are able to reject it. This switches the necessary judgement from subjective (could I have preferred that choice) to an objective one (did I see this face in the choice task) minimising ambiguity. However, when the stimulus was encountered in a similar task, a source of ambiguity is re-introduced by making it harder to identify whether the task in which the stimulus was encountered was the same or a different one.

Whilst ambiguity provides one possible explanation for observed effects, without a way to measure ambiguity this remains to be proven. Confidence has been put forward as a candidate for measuring ambiguity. However, research has shown mixed findings regarding the relationship of choice blindness and confidence. Similarly, here we fail to detect a significant relationship of confidence and proportion of trials detected (chapters 5-7).

In chapter 3 I propose that a more objective measure of ambiguity can be inferred from choice consistency across trials. The proportion of trials on which participants switch choices varies with what appears to be ambiguity of the task (e.g., Loomes & Sugden, 1998; Rieskamp et al., 2006), and the process of measuring choice consistency is very similar to choice blindness, with both procedures testing choices over two points in time, and both test if they elicit the same outcome across the trials – whether through the same selection for choice consistency, or recognition of the wrong outcome in the choice blindness paradigm. This makes it easy to apply what we know about choice consistency to choice blindness. However, in trying to use choice consistency to generate hypotheses about choice blindness I found mixed results. Whilst in chapter 5 my results are congruent with those previously reported in choice consistency studies (DeShazo & Fermo, 2002; Collins & Vossler, 2009), demonstrating that people are more likely to detect a mismatch in intention and outcome when choosing from three compared to two alternatives, just as people show more consistency across their choices when their choice set consists of more than two options. This supports the notion that the two paradigms reflect the same underlying property, which I have referred to as ambiguity in the current work. On the other hand, in chapter 6 I find that negative framing of the choice blindness paradigm leads to a decrease in detection, whereas past research suggests that consistency for negatively framed tasks is in-fact higher (Kogut, 2011). I propose that the differences in procedures used, such as variation in the number of alternatives presented to the participants as well as the number of choices selected, may account for this discrepancy (see chapter 6 for discussion), however, it does pose the possibility that choice blindness and choice consistency may, in fact, measure different properties. However, since negatively framed statements are likely to introduce more ambiguity to the task, I propose that the experiments presented in chapter 6 are better suited to reflect the effects of ambiguity compared to the procedure employed to study framing in the literature on consistency (Kogut, 2011).

Generally, incorporating ambiguity into self-perception theory provides us with a plausible account of the process undergone by participants in accepting false feedback. As discussed in chapter 3, predictions made by self-perception theory can also be explained by cognitive dissonance (when there is an inconsistency between cognitions people experience mental discomfort, which results in the adjustment of one or more of the cognitions – Festinger, 1957). Unfortunately, the research presented here does not allow us to differentiate which model is better suited to describing false feedback effects, however, as there is still no evidence to suggest people do experience psychological discomfort, self-perception theory provides the more parsimonious explanation of false feedback acceptance effects.

In this section, I have attempted to outline the contributions of my own research to how source memory (including availability heuristic and self-serving bias), self-perception theory (or cognitive dissonance) and ambiguity (including discovered preference hypothesis, or constructed preference theory) can explain, and predict false feedback effects. Overall my work provides further support for the plausibility of each of the mechanisms discussed, suggesting some element of every proposed approached are likely to play a role in bringing about choice blindness. This is not surprising, since none of the theories make contradictory prediction for the empirical question at hand, and in many ways the majority of the processes are complimentary, simply describing a different decision stage of every process. For instance, we can take self-perception theory as a skeleton of false feedback acceptance processes consisting of two-parts; assessing validity of the information based on internal cue strength, and using external information to judge the likelihood of the information provided as we would for any other individual. Memory strength can then be described as a determining factor of internal information, whilst ambiguity of the situation would be key in determining the judgement based on internal information. In this way, we can see that the approaches discussed are not in competition, and build on one another. To fully understand, why false feedback acceptance occurs, we would need to consider all the possible factors that can determine internal signals and external signals alike, which is outside the scope of this thesis. However, it does call for a development of such comprehensive, unified theory which can then be used to fine-tune our knowledge of human cognition, by creating testable hypotheses. For the moment, however, we lack such theory. (e.g., Gawronski & Strack, 2004).

Having discussed how the experimental work presented in chapters 4 through 7 can help us understand the factors that determine when this occurs and the possible underlying mechanisms, I will now turn to the last two sections of this thesis, which try to capture how the research presented here can contribute to wider academic literature (focusing on introspection, error detection and preferential choice) and the potential of applying such knowledge respectively.

8.4 Implications for Research

In chapter 2 of this thesis, I put forward three areas that appear to be closely intertwined with the false feedback acceptance effects discussed in this thesis; introspection, error detection and preferential choice. I will now briefly discuss how the findings reported in this thesis contribute to such phenomena.

The very nature of accepting false feedback that contradicts the information we recently provided about the self, questions our ability to be introspective, or access the information about our own attitudes, actions and what brought them about. Across the experiments presented here, we provide further evidence that false feedback is indeed often accepted, therefore questioning our introspective abilities. Nonetheless, the research also shows some hope for the knowledge of ourselves by demonstrating that some conditions can increases mismatch detections which in turn suggests higher introspective access. For example, chapters 4 and 5 show that if feedback provided is not advantageous and negatively affects our self-image we are likely to detect its inaccuracy. Furthermore, if the feedback encountered does not seem probable, for example, if we do not remember considering it (chapter 3), we also have a heightened ability to detect its falsehood.

This suggests that whilst we may not hold a precise description of our personality profiles, or preference order as a precise list, we must have the ability to access relevant past experiences and emotional associations. For example, in order to reject negative feedback, we need to access information that suggests ‘this is bad for me’, or when we reject feedback that was not a part of the choice, we must be able to compare to it to the knowledge of what we had seen in the past. Perhaps the research outlined does not call for labelling introspective information unreliable, but instead a reconsideration of what it means to be introspective.

This further has important implications for error detection. The support for the notion that people tend to accept false feedback can be viewed as a demonstration of failure to detect errors regardless of whether we treat that information as suggestive that we made an error or that some external factor has led to an error being made. The research, however, is also reassuring as it suggests that by defining tasks in a particular way can maximise our ability to detect mismatches between our intentions and the environment. For example, at this stage we can suggest that extending decision sets to three instead of two choices can increase our ability to monitor the choice that was made (chapter 5), similarly it appears that when dealing with choosing from a small set of alternatives we are better off phrasing the task in a positive manner (i.e., which one do you prefer; chapter 6), whilst past research indicates that when rejecting alternatives to narrow down the item pool asking people to reject the unwanted alternatives can result in a higher choice consistency (Kogut, 2011). Another positive finding comes from chapter 4 of this thesis, as it suggests that even when we fail to detect an error it does not mean that error is necessarily incorporated into our beliefs system, or that it will impact how we treat information at a later date.

It must be noted that in many instances it is unclear why we fail to detect errors, given that we have evolved specialised neural systems to do so (see Holroyd & Coles, 2002; Yeung et al., 2004). However, by clarifying circumstances in which error detection is high or low we are making a step in the right direction to determine how the human mind adapts to dealing with the imprecise and error-prone world. Perhaps now that we have built up some understanding using behavioural data, we can start introducing biological measures to decipher what this means. For instance, an EEG study measuring error-related negativity (Gehring, 1992) produced during the Barnum procedure or choice blindness can inform us whether not invalid information is detected as error at all, or whether this is simply not transferred to our consciousness. This can, in turn, be used to understand why some circumstances are more prone to error than others (as discussed in chapters 5 through 7) and help identify how to maximise error detection. Alternatively, monitoring the activity in the anterior cingulate cortex during the choice blindness task can tell us if a person is experiencing discomfort, or dissonance (van Veen, Krug, Schooler & Carter, 2009), and help us test the cognitive dissonance explanation of choice blindness. Whilst continuing behavioural research that identifies situational factors that can improve choice monitoring and is no doubt of value in academic and practical application settings alike, expanding the approach can aid the interpretation of the behavioural data and help formulate a comprehensive model of feedback acceptance.

Lastly, I would like to touch on the impact my findings may have on our understanding of preferential choice. In chapter 3, I discussed how understanding violations of rational choice theory is crucial to developing descriptive models of how the mind operates, which has been a challenge academics have been trying to tackle for the last half a century (e.g., Kahneman & Tversky, 1979). Choice blindness (Johansson et al., 2005) was one of the latest of such violations, demonstrating that we are unable to access a stable representation of our preference, or even to use a recently made choice to inform the task of justifying a decision (that we did not make). Beyond demonstrating that we fail to think like rational agents, choice blindness has also provided us with a way to measure choice stability, or the extent to which a stated preference is likely to remain the same regardless of external influences.

In the past, choice stability has been measured through choice consistency, or the elicitation of the same outcome across different points in time, with research often reporting that people fail to remain consistent across two choices on about 25% of trials (Camerer, 1989; Hey, 2001; Loomes & Sugden, 1998). Here I compared the detection of invalid feedback, to the rates of consistency across trials reported in past literature. Specifically, I investigated whether increasing the number of choice alternatives, as well as whether negative framing of the task, can improve choice stability demonstrated through the choice blindness task, as would be hypothesised from past literature using choice consistency to measure stability (DeShazo & Fermo, 2002; Collins & Vossler, 2009, Kogut, 2011). In comparing the rate of choice blindness for two and three-alternative tasks, I find that ternary choice is indeed more stable than binary, however, this effect only occurs when the alternatives presented are of different visual similarity and level of attractiveness. Nonetheless, since past research fails to control for such characteristics of the alternatives presented, it is possible that the same effect would be observed for choice consistency, suggesting that choice blindness and choice consistency are likely to reflect the same inherent characteristics of choice processes.

On the other hand, in exploring how framing impacts the level of choice blindness, I find that people are more likely to detect invalid feedback when both the choice and justification elements of the task are framed positively, contrary to what we would expect from Kogut's (2011) work on choice consistency. However, there are also a number of procedural differences that could have led to the discrepancy in conclusions, including the difference in number of alternatives presented and selected, and the fact that Kogut (2011) measured consistency of choices with previously stated opinions, and the current research attempts to measure the consistency between a choice and its direct outcome (for detailed discussion of the differences see chapter 6). Whether it is the procedural difference that led to this discrepancy, or that choice blindness and choice consistency do indeed reflect different cognitive mechanisms remains unclear, and requires further research. All I conclude is that as far as the choice blindness procedure is concerned, increasing the number of alternatives and framing the choice and judgement tasks in a positive manner can increase false feedback detection, and thus improve at least some form of choice stability.

Chapter 4 provides another interesting contribution to how we perceive preferential choice. Past research has shown false feedback acceptance can shape our preferences. For example, Johansson et al. (2014) demonstrated that accepting feedback that suggests the non-preferred alternative was, in fact, preferred results in higher likelihood of that alternative being selected in the future. Similarly, information about the self, such as high extraversion, has been shown to impact subsequent social interactions (Sakamoto et al., 2000). Accordingly, by integrating the two lines of research, we can hypothesise that presenting people with information about their own characteristics which are directly related to preferences, should impact actual subsequent preferences, as long as people accept the feedback presented as accurate. However, in the experiment presented in chapter 4, I find that risk preference as demonstrated in choosing between certain and risky lotteries is not affected by accepting altered information about one's own financial risk preference. This suggests that risk preferences are not as malleable, as one would hypothesise based on previous research, at least in some circumstances.

Overall, it seems clear that false feedback acceptance can help us decipher when choices are more stable, allowing us to more accurately describe and predict preferential choice behaviours. I have attempted to outline the small contribution my own work presents for the broader understanding, however, there is still a long way to go before we have a comprehensive theory of preferential choice. Personally, I feel that it may be time to pause and reflect on how the breadth of knowledge we have accumulated over the last century combines together and whether approaches from other disciplines, such as Barnum effect as a product of psychometric evaluation literature, can come together. Such a task is not, however, in the scope of this PhD thesis. Before concluding this thesis, I will now briefly discuss one last area to which the empirical work presented here is relevant, the potential for practical applications.

8.5 Relevance to application

Lastly, I would like to briefly discuss the practical implications of my work, and choice blindness more broadly. Like most academic research, the Barnum procedure and the choice blindness paradigm are usually carried out under strictly control experimental conditions with abstract choices of little consequence to the decision maker. Although this provides the rigour needed in academia, it makes it difficult to simply take the procedure and apply it to real world behavioural change, and yet I am often asked whether that is what I intend to do.

One question is whether acceptance of false feedback provides a "magic bullet" for behaviour change to replace 'bad' preferences with 'good' ones, for example, if a person struggling to eat healthily can we simply tell them that they chose a healthy option (e.g., apple) as opposed to an unhealthy option (e.g., chocolate) when ordering their lunch. I think this is very unlikely. Firstly, as Somerville and McGowan (2016) demonstrated in their application of choice blindness to familiar chocolate brands, only a very small proportion of people experience choice blindness for items with which they have had ample experience. Second, comparing apples with chocolates requires using two discrete categories of items in inducing choice blindness, this is yet to be investigated directly, however, research into similarity deems this endeavour unlikely to be successful (see chapter 5). Last but not least, no research would advocate the use of false feedback in real life, as this would be unethical under almost all circumstances. In academic research, participants are informed of their rights, only participate if they consent to do so, often pre-decide if they are willing to take part in deception based research (see Prolific Academic sign up agreement), and receive a full debrief at the end of the experiment. The inability to follow these steps in real life could have many adverse consequences from perceived loss of control to loss of trust in society, not to mention it is a direct violation of people's right to choose.

Although the use of paradigms that utilise false feedback is unlikely in real life situations, the insights we gained from the research within the choice blindness domain can have many important implications. Questions such as 'Which decision types are prone to manipulation?' or 'How do we elicit the most stable choices?' can be useful in many different contexts. One area which has already become a focus of choice blindness research is eyewitness testimony. Sagana and colleagues (Sauerland et al., 2013; Sagana et al., 2013; Sagana et al., 2014b; Sagana, Sauerland & Merckelbach, 2016) have established that choice blindness does in-fact occur in field studies that closely resemble the process undergone by witnesses in real life, questioning the validity of eyewitness identification overall. The aim of the research, however, is not to destroy the legal system, but to understand how eyewitness decisions are made and how these can be improved. For example, does the fact that distortion of the decision made in an identification task (through choice blindness) can lead to strengthening an invalid identification later (Sagana et al., 2014a) suggest that we should avoid using multiple identification processes? The answers perhaps are not as straight forward as the questions involved given the practical limitations of changing procedure, but in the very least the research based information can identify which decisions may be invalid.

The research presented in the current thesis suggests that people are more prone to the inability to monitor their choices when the alternatives presented are similar versus dissimilar, when there are two options to choose from compared to three, and when they are asked to reject as opposed to select their choice. I propose that it is in-fact this error-prone procedure that is most conducive to our legal system, as if the suspect presented within the possible alternatives is identified correctly within an error prone scenario it is more likely that the choice was made based on established knowledge. This however, is only applicable when only one suspect is presented within a line-up and the incorrect nature of choosing any other alternative can be established with certainty.

The eyewitness procedure is outside the scope of this thesis and my vision of how knowledge of what type of circumstances are prone to error should be used is nothing more than a personal hypothesis (see Sagana, 2015 for discussion). However, whilst the robustness of eyewitness testimony remains an area of concern, research that helps identify what makes decisions prone to error, including choice blindness and the Barnum effect, will remain relevant. In fact, this statement applies to any field that involves assessing the validity of information or making choices with large consequences, especially when others may have a vested interest in the outcome.

False feedback acceptance research also holds important implications for clinical diagnoses and treatment. Consider work by Halperin and Snyder (1979) which demonstrates that feedback that suggests high susceptibility to change can increase susceptibility to therapy. If this is really the case, perhaps the benefits of using false feedback can outweigh the ethical concerns associated with its use. On the other hand, research in chapter 4 suggests that changing responses with the use of false feedback is harder than anticipated and cautions us about the need for research validation before considering any form of applications. Alternatively, false feedback can also play an interesting role in the understanding of clinical symptoms as opposed to their direct treatment. For example, Aardema et al.'s (2014) findings that susceptibility to choice blindness is linked with traits associated with obsessive compulsive disorder, as well as scores on the schizotypal and depression scales. Whilst the knowledge itself doesn't constitute application, perhaps with better understanding when choice blindness occurs and how it can be minimised we can begin to develop more accurate diagnoses and treatments in turn.

Overall, the Barnum effect and choice blindness might not be directly transferrable to the real world, however, they do have important implications for how we can minimise the influence of error or manipulations on the decisions that we make. Every study that reveals something new about how we treat invalid information, can in turn, be considered a contribution to real life problems.

* * *

In summary, throughout this thesis, I attempted to outline the importance and characteristics of false feedback acceptance phenomena, namely the Barnum effect, or the tendency to accept false feedback about one's own personality, and choice blindness, or the tendency to accept false feedback about one's choices. In the introductory chapters, I presented the case for relevance of such work to our understanding of introspective ability, error detection and preferential choice, as well as cognition more generally, and went on to outline what we know about false feedback acceptance. More specifically, discussion was dedicated to establishing that task domain, similarity and favourability of feedback, and factors that are known to contribute to task ambiguity, such as time and problem frame, are likely to contribute to the likelihood of false feedback acceptance. In the original empirical work presented in this thesis, I provided further evidence that false feedback about the self is often accepted as accurate, for both the Barnum effect and choice blindness, and that the likelihood of such acceptance can indeed be altered by varying the task at hand. For the first time, I demonstrated that choice blindness can be achieved using alternatives outside of the task as feedback, and that the likelihood of detecting invalid feedback is decreased when using negatively framed task instructions, and three as opposed to two alternative tasks. I further reported that whilst false feedback acceptance has been found to impact subsequent behaviours in the past, using the Barnum effect to achieve such change is not always successful. The research outlined has enhanced the knowledge of when and why false feedback acceptance occurs and the potential subsequent effects, contributing to our understanding of how self-monitoring can be improved more generally. This is a small step to creating a comprehensive picture of human cognition, and as proposed throughout this work the benefits that can be gained from studying false feedback acceptance are far from exhausted and hold a lot of potential for future academic work.

References

Aardema, F., Johansson, P., Hall, L., Paradisis, S. M., Zidani, M., & Roberts, S. (2014). Choice blindness, confabulatory introspection, and obsessive-compulsive symptoms: A new area of investigation. *International Journal of Cognitive Therapy*, *7*(1), 83-102.

Abhyankar, P., O'Connor, D. B., & Lawton, R. (2008). The role of message framing in promoting MMR vaccination: Evidence of a loss-frame advantage. *Psychology, Health and Medicine*, *13*(1), 1-16.

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427-432.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, *3*(2), 111-150.

Alhakami, A. S., & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis*, *14*(6), 1085-1096.

Amalberti, R. (2013). Human Error at the Centre of the Debate on Safety. In *Navigating Safety* (pp. 19-52). Springer Netherlands.

Andersen, P., & Nordvik, H. (2002). Possible Barnum Effect in the Five Factor Model: Do Respondents Accept Random Neo Personality Inventory–Revised Scores as Their Actual Trait Profile? *Psychological Reports*, *90*(2), 539-545

Ariely, D. (2008). Are we in control of our own decisions? TED: EG The Entertainment Gathering Conference. Monterey, California.

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

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, *26*(102), 121-127.

Ayllon, T., & Azrin, N. (1968). The token economy: A motivational system for therapy and rehabilitation.

Bachrach, A. J., & Pattishall, E. G., Jr. (1960). An experiment in universal and personal validation. *Psychiatry*, *23*(3), 267-270.

Baillargeon, J., & Danis, C. (1984). Barnum meets the computer: A critical test. *Journal of Personality Assessment*, *48*(4), 415-419.

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

Bernstein, N. (1967). *The Coordination of Motor Function and Locomotion.* New York: Pergamon.

Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus on-line processing: Implications for attitude strength. *Journal of Experimental Social Psychology*, *42*(5), 646-653.

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1).

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. *Contributions to Probability and Statistics*, *2*, 97-132.

Boot, W. R., Brockmole, J. R., & Simons, D. J. (2005). Attention capture is modulated in dual-task situations. *Psychonomic Bulletin & Review*, *12*(4), 662-668.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). *Salience and Consumer Choice* (No. w17947). National Bureau of Economic Research.

Brasil-Neto, J. P., Cohen, L. G., Panizza, M., Nilsson, J., Roth, B. J., & Hallett, M. (1992). Optimal focal transcranial magnetic activation of the human motor cortex: effects of coil orientation, shape of the induced current pulse, and stimulus intensity. *Journal of Clinical Neurophysiology*, *9*(1), 132-136.

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

Brown, T. C., Kingsley, D., Peterson, G. L., Flores, N. E., Clarke, A., & Birjulin, A. (2008). Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement. *Journal of Public Economics*, *92*(7), 1595-1606.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, *2*(1), 61-104.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45.

Carruthers, P. (2011). *The opacity of mind: an integrative theory of self-knowledge*. OUP Oxford.

Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological*, *39*(7), 621-640.

Chater, N., Johansson, P., & Hall, L. (2011). The non-existence of risk attitude. *Frontiers in psychology*, *2*.

Chen, M. K., & Risen, J. L. (2010). How choice affects and reflects preferences: revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, *99*(4), 573.

Cheung, T. T. L., Junghans, A. F., Dijksterhuis, G. B., Kroese, F., Johansson, P., Hall, L., & De Ridder, D. T. D. (2016). Consumers' choice-blindness to ingredient information. *Appetite*, *106*, 2-12.

Cochran, K. J., Greenspan, R. L., Bogart, D. F., & Loftus, E. F. (2016). Memory blindness: Altered memory reports lead to distortion in eyewitness memory. *Memory & cognition*, *44*(5), 717-726.

Cioffi, D., & Garner, R. (1996). On doing the decision: Effects of active versus passive choice on commitment and self-perception. *Personality and Social Psychology Bulletin*, *22*(2), 133-147.

Collins, J. P., & Vossler, C. A. (2009). Incentive compatibility tests of choice experiment value elicitation questions. *Journal of Environmental Economics and Management*, *58*(2), 226-235.

Collins, R. W., Dmitruk, V. M., & Ranney, J. J. (1977). Personal validation: Some empirical and ethical considerations. *Journal of Consulting and Clinical Psychology, 45*, 70-77.

Costa, P. T., & McCrae, R. R. (1985). The NEO personality inventory.

Costa, P. T., & McCrae, R. R. (1992). Personality Inventory (Short Form): Neuroticism, Extraversion and Openness (NEO: Revised NEO Personality Inventory [NEOPI-R] and the NEO Five-Factor Inventory [NEO-FFI]): Professional Manual. Odessa, FL: Psychological Assessment Resources. *Psychological Assessment Resources, Odessa, FL*.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349.

Deaux, K. (1985). Sex and gender. *Annual review of psychology*, *36*(1), 49-81.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*(1), 123-143.

Dickson, D. H., & Kelly, I. W. (1985). The 'Barnum Effect' in personality assessment: A review of the literature. *Psychological Reports*, *57*(2), 367-382.

Dmitruk, V. M., Collins, R. W., & Clinger, D. L. (1973). The" Barnum effect" and acceptance of negative personal evaluation. *Journal of Consulting and Clinical Psychology*, *41*(2), 192.

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). MINDSPACE: influencing behaviour for public policy.

Dywan, J. (1995). The illusion of familiarity: an alternative to the report-criterion account of hypnotic recall. *International Journal of Clinical and Experimental Hypnosis*, *43*(2), 194-211.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643-669.

Erismann, T., & Kohler, I. (1953). Upright vision through inverting spectacles. *Penn. State College: Psychological Cinema Register*, (2070).

Fazio, R. H., & Zanna, M. P. (1978). Attitudinal qualities relating to the strength of the attitude-behavior relationship. *Journal of Experimental Social Psychology*, *14*(4), 398-408.

Festinger, L. (1957). *A Theory of Cognitive Dissonance.* Stanford, CA: Stanford University Press.

Fichten, C. S., & Sunerton, B. (1983). Popular horoscopes and the "Barnum effect". *The Journal of Psychology*, *114*(1), 123-134.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400-409.

Forer, B. R. (1949). The fallacy of personal validation: a classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology*, *44*(1), 118.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*(11), 1133-1140.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195.

Furnham, A. (1989). Personality and the acceptance of diagnostic feedback. *Personality and Individual Differences*, *10*(11), 1121-1133.

Furnham, A., & Schofield, S. (1987). Accepting personality test feedback: A review of the Barnum effect. *Current Psychology*, *6*(2), 162-178.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*(4), 535-542.

Gehring, W. J. (1992). *The error-related negativity: Evidence for a neural mechanism for error-related processing* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Gehring, W. J., Coles, M. G., Meyer, D. E., & Donchin, E. (1995). A brain potential manifestation of error-related processing. *Electroencephalography and Clinical Neurophysiology-Supplements only*, *44*, 261-272.

Ghaferi, A. A., Birkmeyer, J. D., & Dimick, J. B. (2009). Variation in hospital mortality associated with inpatient surgery. *New England Journal of Medicine*, *361*(14), 1368-1375.

Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*(5), 778.

Glick, P., Gottesman, D., & Jolton, J. (1989). The fault is not in the stars: Susceptibility of skeptics and believers in astrology to the Barnum effect. *Personality and Social Psychology Bulletin*, *15*(4), 572-583.

Godlove, D. C., Emeric, E. E., Segovis, C. M., Young, M. S., Schall, J. D., & Woodman, G. F. (2011). Event-related potentials elicited by errors during the stop-signal task. I. Macaque monkeys. *The Journal of Neuroscience*, *31*(44), 15640-15649.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public domain personality measures. *Journal of Research in Personality*, *40*, 84-96.

Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, *39*(2), 317-329.

Greene, R. L. (1977). Student acceptance of generalized personality interpretations: A reexamination. *Journal of Consulting and Clinical Psychology*, *45*(5), 965.

Greene, R. L., Harris, M. E., & Macon, R. S. (1979). Another look at personal validation. *Journal of Personality Assessment*, *43*(4), 419-423.

Guastello, S. J., & Rieke, M. L. (1990). The Barnum effect and validity of computer-based test interpretations: The Human Resource Development Report. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*(2), 186.

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), 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.

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 and voter intentions. *PLoS One*, *8*(4), e60554.

Halperin, K., & Snyder, C. (1979). Effects of enhanced psychological test feedback on treatment outcome: Therapeutic implications of the Barnum effect. *Journal of Consulting and Clinical Psychology*, *47*(1), 140.

Halperin, K., Snyder, C., Shenkel, R., & Houston, B. (1976). Effects of source status and message favorability on acceptance of personality feedback. *Journal of Applied Psychology, 61*, 85-88.

Hanoch, Y., Johnson, J. G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment: An application to risk-taking behavior. *Psychological Science*, *17*(4), 300-304.

Harmon-Jones, E., & Mills, J. (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association.

Harrison, J. D., Young, J. M., Butow, P., Salkeld, G., & Solomon, M. J. (2005). Is it worth the risk? A systematic review of instruments that measure risk propensity for use in the health setting. *Social Science & Medicine*, *60*(6), 1385-1396.

Haynes, G. A. (2009). Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology & Marketing*, *26*(3), 204-212.

Heider, F. (1958). *The psychology of Interpersonal Relations*. Psychology Press.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, *51*(6), 1106.

Herriges, J. A., & Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, *30*(1), 112-131.

Hey, J. D. (2001). Does repetition improve consistency?. *Experimental Economics*, *4*(1), 5-54.

Hoeffler, S., & Ariely, D. (1999). Constructing stable preferences: A look into dimensions of experience and their impact on preference stability. *Journal of Consumer Psychology*, *8*(2), 113-139.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679.

Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Research*, *1105*(1), 93-101.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica*, *17*(66), 159-174.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 90-98.

International Personality Item Pool. (2006). *IPIP: A scientific collaboratory for the development of advanced measures of personality and other* *individual differences.* Retrieved September 17, 2006 from http://ipip.ori.org/ipip/

Izuma, K., & Murayama, K. (2013). Choice-induced preference change in the free-choice paradigm: a critical methodological review. *Frontiers in Psychology*, *4*, 41.

Jacoby, J. (2000). Is It Rational to Assume Consumer Rationality-Some Consumer Psychological Perspective on Rational Choice Theory. *Roger Williams UL Rev.*, *6*, 81.

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, *82*(6), 477-482.

Johansson, P., Hall, L., & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia*, *51*(2), 142-155.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116-119.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15*(4), 673-692.

Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2014). Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it!. *Journal of Behavioral Decision Making*, *27*(3), 281-289.

Johnson, J. T., Cain, L. M., Falke, T. L., Hayman, J., & Perillo, E. (1985). The" Barnum effect" revisited: Cognitive and motivational factors in the acceptance of personality descriptions. *Journal of Personality and Social Psychology*, *49*(5), 1378.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*(4), 337-342.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Kintz, B. L., Delprato, D. J., Mettee, D. R., Persons, C. E., & Schappe, R. H. (1965). The experimenter effect. *Psychological Bulletin*, *63*(4), 223.

Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific-fear questionnaires. *Behavior Therapy*, *5*(3), 401-409.

Kogut, T. (2011). Choosing what I want or keeping what I should: The effect of decision strategy on choice consistency. *Organizational Behavior and Human Decision Processes*, *116*(1), 129-139.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80.

Lauriola, M., Levin, I. P., & Hart, S. S. (2007). Common and distinct factors in decision making under ambiguity and risk: A psychometric study of individual differences. *Organizational Behavior and Human Decision Processes*, *104*(2), 130-149.

Layne, C. (1979). The Barnum effect: Rationality versus gullibility? *Journal of Consulting and Clinical Psychology*, *47*(1), 219.

Layne, C. (1998). Gender and the Barnum effect: A reinterpretation of Piper-Terry and Downey's results. *Psychological Reports*, *83*(2), 608-610.

Liberman, N., & Förster, J. (2006). Inferences from decision difficulty. *Journal of Experimental Social Psychology*, *42*(3), 290-301.

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition*, *5*(6), 696-699.

Loftus, E. F. (1997). Creating false memories. *Scientific American*, *277*, 70-75.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361-366.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, 581-598.

Macdonald, D. J., & Standing, L. G. (2002). Does self-serving bias cancel the Barnum Effect?. *Social Behavior and Personality: an International Journal*, *30*(6), 625-630.

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport*, *12*(9), 1851-1855.

McClelland, G. H., Stewart, B. E., Judd, C. M., & Bourne, L. E. (1987). Effects of choice task on attribute memory. *Organizational Behavior and Human Decision Processes*, *40*(2), 235-254.

McLaughlin, O., & Somerville, J. (2013). Choice blindness in financial decision making. *Judgment and Decision Making*, *8*(5), 561-572.

Meehl, P. E. (1956). Wanted-a good cook-book. *American Psychologist*,*11*(6), 263.

Merckelbach, H., Jelicic, M., & Pieters, M. (2011). Misinformation increases symptom reporting: a test–retest study. *JRSM short reports*,*2* (10), 75.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task:

Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788-798.

Mitsuda, T., & Glaholt, M. G. (2014). Gaze bias during visual preference judgements: Effects of stimulus category and decision instructions. *Visual Cognition*, *22*(1), 11-29.

Moser, J. S., & Simons, R. F. (2009). The neural consequences of flip-flopping: The feedback-related negativity and salience of reward prediction. *Psychophysiology*, *46*(2), 313-320.

Mosher, D. L. (1965). Approval motive and acceptance of "fake" personality test interpretations which differ in favorability. *Psychological Reports*, *17*(2), 395-402.

Myers, I. B. (1962). *The Myers-Briggs Type Indicator* (pp. 1-5). Palo Alto, CA: Consulting Psychologists Press.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator* (Vol. 3). Palo Alto, CA: Consulting Psychologists Press.

Nawas, M. M. (1971). Standardized scheduled desensitization: Some unstable results and an improved program. *Behaviour Research and Therapy*, *9*(1), 35-38.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151-166.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an anti-saccade task. *Psychophysiology*, *38*(5), 752-760.

Nisbett, R. E., & Bellows, N. (1977). Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, *35*(9), 613.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*(1), 1.

Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, *44*(3), 347-356.

O'Dell, J. W. (1972). PT Barnum explores the computer. *Journal of Consulting and Clinical Psychology*, *38*(2), 270

Olsen, S. B., Lundhede, T. H., Jacobsen, J. B., & Thorsen, B. J. (2011). Tough and easy choices: testing the influence of utility difference on stated certainty-in-choice in choice experiments. *Environmental and Resource Economics*, *49*(4), 491-510.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237-241

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776.

Orpen, C., & Jamotte, A. (1975). The acceptance of generalized personality interpretations. *The Journal of Social Psychology*, *96*(1), 147-148.

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*(1), 54.

Pandey, J. (2011). Source Memory. In *Encyclopaedia of Clinical Neuropsychology* (pp. 2325-2326). Springer New York.

Pärnamets, P., Hall, L., & Johansson, P. (2015). Memory distortions resulting from a choice blindness task. In *37th Annual Conference of the Cognitive Science Society: Mind, Technology, and Society* (pp. 1823-1828). Cognitive Science Society.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87-131.

Pedale, T., & Santangelo, V. (2015). Perceptual salience affects the contents of working memory during free-recollection of objects from natural scenes. *Frontiers in Human Neuroscience*, *9*, 60.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, *22* (2-3), 227-250.

Plott, C. R. (1966). Externalities and corrective taxes. *Economica*, 84-87.

Poškus, M. S. (2014). A new way of looking at the Barnum effect and its links to personality traits in groups receiving different types of personality feedback. *Psychology*, *50*(50), 95-105.

Poškus, M. S., & Žukauskienė, R. (2014). „Jūsų NEO išvada "pageidaujamumo bei apibendrintumo suvokimas: rekomendacijos Barnumo efekto tyrimams [Perception of favorability and generalness of the "Your NEO Summary": recommendations for Barnum effect studies]. *Psichologiniai tyrimai. Reikšmė visuomenei– iššūkis tyrėjui*, 23-24.

Posner, R. A. (1993). *The problems of jurisprudence*. Harvard University Press.

Prolific Academic (2016). Prolific Academic [software]. Available from http://prolific.ac

The psychological image collection at Stirling (PICS) (n.d.). University of Stirling Psychology Department. Retrieved July 24, 2016 from http://pics.psych.stir.ac.uk/

Qualtrics. (2016). Qualtrics [software]. Available from http://qualtrics.com

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, *1*(4), 563-581.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45-48.

Richards, W. S., & Merrens, M. R. (1971). Student evaluation of generalized personality interpretations as a function of method of assessment. *Journal of Clinical Psychology*, *27*(4), 457-459.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631-661.

Rizzo, A., Ferrante, D., & Bagnara, S. (1995). Handling human error. *Expertise and Technology: Cognition & Human-Computer Cooperation*, 195-212.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Rolfe, J., & Bennett, J. (2009). The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics*, *68*(4), 1140-1148.

Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage*, *53*(2), 694-706.

Rosen, G. M. (1975). Effects of source prestige on subjects' acceptance of the Barnum effect: Psychologist versus astrologer. *Journal of Consulting and Clinical Psychology*, *43*(1), 95.

Ruchsow, M., Spitzer, M., Grön, G., Grothe, J., & Kiefer, M. (2005). Error processing and impulsiveness in normals: evidence from event-related potentials. *Cognitive Brain Research*, *24*(2), 317-325.

Sagana, A. (2015). *A blind man's bluff: Choice blindness in eyewitness testimony.* (Doctoral dissertation). Retrieved from https://cris.maastrichtuniversity.nl/portal/files/1715835/guid-0ac07c43-c029-4d52-8f91-c8b3b6932d39-ASSET1.0

Sagana, A., Sauerland, M., & Merckelbach, H. (2013). Witnesses' blindness for their own facial recognition decisions: A field study. *Behavioral Sciences & The Law*, *31*(5), 624-636.

Sagana, A., Sauerland, M., & Merckelbach, H. (2014a). Memory impairment is not sufficient for choice blindness to occur. *Frontiers in psychology*, *5*.

Sagana, A., Sauerland, M., & Merckelbach, H. (2014b). 'This Is the Person You Selected': Eyewitnesses' Blindness for Their Own Facial Recognition Decisions. *Applied Cognitive Psychology*, *28*(5), 753-764.

Sagana, A., Sauerland, M., & Merckelbach, H. (2016). The effect of choice reversals on blindness for identification decisions. *Psychology, Crime & Law*, *22*(4), 303-314.

Sakamoto, A., Miura, S., Sakamoto, K., & Mori, T. (2000). Popular psychological tests and self-fulfilling prophecy: An experiment of Japanese female undergraduate students. *Asian Journal of Social Psychology*, *3*(2), 107-124.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61-71.

San Martín, R., Manes, F., Hurtado, E., Isla, P., & Ibañez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *Neuroimage*, *51*(3), 1194-1204.

Sarter, N. B., & Alexander, H. M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *The International Journal of Aviation Psychology*, *10*(2), 189-206.

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, *18* (2), 371-381.

Sauerland, M., Sagana, A., Otgaar, H., & Broers, N. J. (2014). Self-relevance does not moderate choice blindness in adolescents and children. *PloS one*, *9*(6), e98563.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225.

Schmidt, R. A., & White, J. L. (1972). Evidence for an error detection mechanism in motor skills: A test of Adams' closed-loop theory. *Journal of Motor Behavior*, *4*(3), 143-153.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, *94*(1), 168.

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532-546

Sedikides, C., Green, J. D., & Pinter, B. (2004). Self-protective memory. *The Self and Memory*, 161-179.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, *21*(4), 546-556.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*(1), 11-36.

Sharot, T., Fleming, S. M., Yu, X., Koster, R., & Dolan, R. J. (2012). Is choice-induced preference change long lasting?. *Psychological Science*, *23*(10), 1123-1129.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*(1), 161-176.

Skinner, B. F. (1938). *The behaviour of organisms: An experimental analysis*. D. Appleton-Century Company Incorporated.

Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, *115*(2), 268-281.

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

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology Monograph, 78*, 1–18.

Snyder, C. R., & Larson, G. R. (1972). A further look at student acceptance of general personality interpretations. *Journal of Consulting and Clinical Psychology*, *38*(3), 384.

Snyder, C. R., & Shenkel, R. J. (1976). Effects of" favorability," modality, and relevance on acceptance of general personality interpretations prior to and after receiving diagnostic feedback. *Journal of Consulting and Clinical Psychology*, *44*(1), 34.

Snyder, C. R., Shenkel, R. J., & Lowery, C. R. (1977). Acceptance of personality interpretations: the" Barnum Effect" and beyond. *Journal of Consulting and Clinical Psychology*, *45*(1), 104.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*(4), 526.

Somerville, J., & McGowan, F. (2016). Can chocolate cure blindness? Investigating the effect of preference strength and incentives on the incidence of Choice Blindness. *Journal of Behavioral and Experimental Economics*, *61*, 1-11.

Steenfeldt-Kristensen, C., & Thornton, I. M. (2013). Haptic choice blindness. *i-Perception*, *4*(3), 207-210.

Stratton, G. M. (1896). Some preliminary experiments on vision without inversion of the retinal image. *Psychological Review*, *3*(6), 611.

Sundberg, N. D. (1955). The acceptability of" fake" versus" bona fide" personality test interpretations. *The Journal of Abnormal and Social Psychology*, *50*(1), 145.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, *11*, 249-288.

Tormala, Z. L., Clarkson, J. J., & Petty, R. E. (2006). Resisting persuasion by the skin of one's teeth: the hidden success of resisted persuasive messages. *Journal of Personality and Social Psychology*, *91*(3), 423.

Tormala, Z. L., & Petty, R. E. (2004). Source credibility and attitude certainty: A metacognitive analysis of resistance to persuasion. *Journal of Consumer Psychology*, *14*(4), 427-442.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

Tversky, A. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*, 30.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, Probability, and Human Decision Making* (pp. 141-162). Springer Netherlands.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis* (pp. 107-129). Springer Berlin Heidelberg.

Ulrich, R. E., Stachnik, T. J., & Stainton, N. R. (1963). Student acceptance of generalized personality interpretations. *Psychological Reports*, *13*(3), 831-834.

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, *4*(4), 400.

Van Damme, I., & Smets, K. (2014). The power of emotion versus the power of suggestion: memory for emotional events in the misinformation paradigm. *Emotion*, *14*(2), 310.

Van Schaik, P., Kusev, P., & Juliusson, A. (2011). Human preferences and risky choices. *Frontiers in Psychology*, *2, 333*.

Van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, *12*(11), 1469-1474.

Von Neumann, J., & Morgenstern, O. (1944). Game theory and economic behavior. *Princeton, Princeton University*.

Weinman, G. (1982). The prophecy that never fails. *Sociological Inquiry*, *52*, 275-87.

White, P. A. (1988). Knowing more about what we can tell: ‘Introspective access' and causal report accuracy 10 years later. *British Journal of Psychology*, *79*(1), 13-45.

Wilson, T. D. (2002). *Strangers to ourselves: Self-insight and the adaptive unconscious*. Harvard University Press.

Wilson, T. D., & Bar-Anan, Y. (2008). The unseen mind. *Science*, *321*(5892), 1046-1047.

Wilson, T. D., & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology*, 118-131.

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*(1), 195-206.

Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *The American Economic Review*, *104*(5), 495-500.

Wyman, A. J., & Vyse, S. (2008). Science versus the stars: A double-blind test of the validity of the NEO five-factor inventory and computer-generated astrological natal charts. *The Journal of general psychology*, *135*(3), 287-300.

Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making*, *10*(3), 211-220.

Yaniv, I., Schul, Y., Raphaelli-Hirsch, R., & Maoz, I. (2002). Inclusive and exclusive modes of thinking: Studies of prediction, preference, and social perception during parliamentary elections. *Journal of Experimental Social Psychology*, *38*(4), 352-367.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2p2), 1.