DOCTOR OF PHILOSOPHY

The effects of static, dynamic and behavioural social cues on social utility and decision-making.

Shore, Danielle

Award date:
2012

Awarding institution:
Bangor University

The effects of static, dynamic and behavioural social cues on social utility and decision-making.

by

Danielle M. Shore

This thesis is submitted in partial fulfillment for the degree of Doctor of Philosophy.

Work completed as School of Psychology, Bangor University.

Acknowledgments

First and foremost I would like to thank Dr Erin Heerey for all her invaluable guidance and support. The work I have produced is a tribute to her excellent supervision throughout the course of my PhD. I would like to express my very great appreciation for all the time and energy invested in helping me develop as a researcher, and for making my PhD experience so enjoyable.

I would also like to express my gratitude to my second supervisor Prof Bob Rafal and my committee chair Dr John Parkinson for their guidance and constructive advice throughout my PhD. I also thank Dr John Parkinson for first suggesting that I should consider pursuing a PhD, and both John and Bob for supporting my application. Without their encouragement and advice I would not have had the courage to take this path.

My grateful thanks are also extended to lab colleagues both old and new for creating a happy, supportive and enjoyable working environment. Special thanks should be given to Thandiwe Gilder for her endless support, care, advice and empathizing.

I would like to express my very great appreciation to Nick Davies, Helen Morgan, Thandiwe Gilder, Jess Shirley and Markos Alexandrou for their invaluable help during the writing of this thesis. I would also like to thank those who have assisted with data collection, and the Bangor Psychology IT team for all their technical support and wizardry. I would also like to acknowledge the Economic and Research Council for my 1+3 studentship.

My particular thanks go to my wonderfully supportive friends and family, who have reassured and encouraged me throughout my PhD. A special mention of appreciation goes to the occupants of 3 Garth Terrace for keeping me motivated and well fed! I would also like to express gratitude to my parents and extended family for always being there to listen, console, motivate and advise. Finally, I would like to thank Markos Alexandrou for always being there to provide reassurance and encouragement.

Contents

List of tables

Chapter Three: **'Tis better to look trustworthy than to be good: The effect of social cues on social utility**

List of Figures

Chapter Two: The Value of Genuine and Polite Smiles

Chapter Three: **'Tis better to look trustworthy than to be good: The effect of social cues on social utility**

Chapter Four: Do social utility judgements influence attentional processing?

Abstract

Humans make many important decisions in social environments. Social judgements of interaction partners based on the cues they send likely influence those decisions, by shaping expectations about the utility or subjective desirability of interacting with those partners. Here I examine how subtle differences in social utility judgements shape the influence social cues have on the decision-making process, in a series of three broad questions.

The first of these relates to the intrinsic value of specific social cues. Anecdotally, cues such as smiles carry value as social rewards. Using a novel application of expected utility theory in an economic game, I show that relative to polite smiles, participants value genuine smiles to the extent that they are willing to forgo the chance to win money for the chance to see these social rewards. This suggests that genuine smiles have value as social cues and increase the utility of the interaction partners who display them.

In social interaction, people receive many different types of social cues. These include, static appearance-based cues such as facial trustworthiness, dynamic expressions such as smiles, and the social decisions an interaction partner makes. These cues sometimes conflict–for example, when an untrustworthy face smiles genuinely. My second question addresses how these cues interact during social decision-making. I investigated how conflicting cues modulate social utility in an investor-trustee game. Surprisingly, unfavourable appearance-based judgements biased investment behaviour long after people had learned the value of a face based on behaviour. Interestingly, a smile present at the time of feedback eliminated this bias. This suggests when more stable cues (appearance and behaviour) conflict, dynamic cues (expressions) are powerful modulators of social utility.

My final question considered one mechanism for understanding how social cues shape decision-making. Specifically, social cues may bias the allocation of attention during social interactions, thereby shaping the social information people acquire. I examined how appearance- and behaviour-based judgements altered the allocation of attention to social stimuli. Strong appearance-based judgements (positive and negative) enhanced stimulus

recognition but only positive behaviour-based judgements altered involuntary allocation of attention to stimuli. This suggests that behaviour-based social utility judgements alter pre-attentive processing, and consequently bias attention distribution.

Taken together, these findings suggest that appearance, facial expression and behaviour cues all contribute to social utility, and this utility guides decision-making in social environments.

Chapter One

General Introduction

Humans make hundreds of decisions everyday, ranging from simple choices such as where to go for lunch, to complex decisions such as whether to change jobs. Regardless of complexity, the goal is to make the most advantageous or beneficial decision. Beneficial decisions are those that lead to the attainment of a goal or a valuable outcome, for example the receipt of money or the development of a friendship. To achieve the best possible outcomes, each choice must be evaluated in terms of both the quality of possible outcomes and the probability with which each is likely to occur. Theoretically, the more accurate the evaluation, the better the course of action that will result. As rational decision-makers, humans should evaluate evidence precisely and objectively, without relying on emotions, heuristics or other strategies to direct their behaviour. However, a great deal of research suggests when people make decisions they do not do so rationally (Camerer, 2003; De Martino, Kumaran, Seymour & Dolan, 2006; Erev & Roth, 1998; Tversky & Kahneman, 1981). People are similarly irrational when they make decisions in the social environment, the focus of this work.

People spend large amounts of their time in social environments, frequently interacting with others. Social interaction is inherently important for daily tasks; to achieve targets at work one must interact with co-workers and superiors. At home, family life revolves around interaction and many people choose to spend their free time interacting with friends. Consequently, many decisions are made in social contexts. These range from simple decisions about who to talk to at a party before one's friend arrives, to whether to trust an investment advisor who recommends the purchase of certain stocks. Obviously, others shape the decisions people make within each social environment. Indeed, research suggests these choices are heavily dependent on the signals people receive from their interaction partners (Behrens, Hunt, Woolrich & Rushworth, 2008).

Rational and beneficial decision-making requires the integration of many pieces of information. Social environments contain information from a variety of sources; a large

proportion of this information comes from other individuals within the environment. In particular, social decisions require the integration of the social cues interaction partners send. This influence can be obvious, such as when a decision requires cooperation with a social partner. However, an interaction partner can influence decisions in more subtle ways. For example, one may be more likely to approach and introduce oneself to an attractive stranger rather than approach a less attractive person at a party, even though they have met the less attractive person once before. In both cases, rational decisions rely on accurate judgements of an interaction partner's likely behaviour. However, as this example highlights, a social partner's cues may influence decisions in an irrational way. For this reason, it is important to know how the different cues people receive from social partners influence social decisions individually and in combination.

Recent research has begun to investigate the complex world of social decision-making (Frith & Singer, 2008), focusing on several key elements such as trust and fairness. Decision-making contexts may be a constructive approach to understanding behaviour in social interactions because they offer ways of measuring people's reactions to social cues using behavioural assays. Indeed, many elements of social behaviour are similar to the choices people make in decision-making contexts, such as choosing between different possible responses, receiving feedback about the chosen response, and evaluating behaviour and outcomes. Two key theories that explain decision-making behaviour are expected utility theory (von Neumann & Morgenstern, 1947) and reinforcement learning theory (Sutton & Barto, 1998). Due to the complexity of social decision-making, it has typically been considered to be distinct from these more simple processes. However, it may be that social information has value and thus these theoretical models may be able to explain how an interaction partner's social cues shape decision-making behaviour.

Social Utility

Expected utility theory is one way to quantify the actual decisions people make in terms of both subjective and objective components. Utility theory posits that people make choices based on a combination of objective and subjective factors. The idea is that stimuli

in the environment have value and this value or utility comes from two distinct components. The first component is the objective value of the object, what it is worth in quantitative terms. The second component is an object's subjective value, how desirable the object is to the individual at that moment in time. For example, a chocolate bar and a bag of crisps may have the same objective value in terms of the energy they will provide, however if the person making the choice between them has just been eating chocolate then the crisps will be subjectively more desirable. Thus, the utility of an object is the combination of its objective value, plus the degree to which that item is desirable at the moment of the decision (von Neumann & Morgenstern, 1947).

Applied to decision-making, utility theory predicts that when making a decision between a set of objects or possible actions, a rational decision-maker will behave in a way that maximizes their expected utility. This means that a rational decision-maker can make a decision, which objectively is irrational, but due to the subjective desirability of the decision outcome, is actually rational. In the example above, if the choice was between a pack of chocolate bars and one bag of crisps, the pack of chocolate bars is the objectively more valuable choice but due to the high subjective value of the crisps they are the rational choice. In terms of utility, the value of the crisps is higher and thus choosing them maximizes expected utility.

This research applies utility theory concepts to social decisions. Social utility is the value of a social partner. It reflects both the degree to which interactions with a particular partner are likely to be valuable and how desirable the specific type of outcome is at the moment of evaluation. In the example of whom to approach at a party, an attractive individual is expected to be friendly, as previous research shows positive traits are associated with attractive faces (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992). Therefore, the interaction with that person is likely to be chosen, because his or her expected utility is high and meeting someone new who is attractive is subjectively desirable. In this way, social utility guides perceptions of, and behaviour toward interaction partners.

Recent research suggests that social utility has a great impact on behaviour, as expectations are important. Expectations change decision-making processes, for example when a social partner fails to match an expected utility estimate, financial outcomes are adversely affected (Sanfey, 2009; Wilson & Eckel, 2006). Interaction partners who violate expectations are better remembered than those who behave as expected (Chang & Sanfey, 2009). In order to make rational decisions, it is necessary to make accurate social utility valuations to avoid losses that occur when a partner falls short of expectations (Sanfey, 2009). Many factors shape an interaction partner's social utility. This research considers two of them: trust and social rewards.

Trust

Trust is important in all social environments and is a key consideration when selecting a mate, a baby sitter, a financial advisor, or a presidential candidate. In psychological terms, trust is one's level of confidence in a social partner's ability to facilitate the achievement of a specific goal or reward (Simpson, 2007). In order to make beneficial social decisions, it is essential to know, or to learn, who is and who is not trustworthy. Accordingly, researchers have argued that the ability to accurately assess trustworthiness is crucial to survival (Cosmides & Tooby 2000).

People must often work with or rely on others, in order to accomplish goals. This means people must trust in a social partner to cooperate (Rempel, Holmes & Zanna, 1985) or provide sound guidance (Biele, Rieskamp & Gonzalez, 2009). In these situations trustworthiness is both objectively valuable and subjectively desirable. Placing trust in a social partner allows them to influence decision-making processes (Biele et al., 2009). The degree to which a social partner is believed to be trustworthy, makes a substantial contribution to their expected utility and thus affects how willing people are to cooperate or invest with that partner (King-Casas et al., 2005; van t' Wout & Sanfey, 2008), and who they choose to approach or avoid (Adolphs, Tranel, & Damasio, 1998).

Misplacement of trust leads to disadvantageous outcomes. For example, following incorrect advice can lead people to make a less beneficial decision than they may have

made without the advice (Biele et al., 2009). The same is true of incorrectly deciding who can and who cannot be trusted (Schlicht, Shimojo, Camerer, Battaglia, & Nakayama, 2010). People provide many different signals of trustworthiness. Perceiving and correctly evaluating these cues enables accurate assessment of trustworthiness and thus beneficial decision-making. In this way trustworthiness contributes to social utility. Due to the inherent requirement of trust in many social decisions it is a highly desirable social characteristic, so the more trustworthy an individual, the higher their social utility.

Social Rewards

Another factor that may contribute to social utility is social rewards. As an individual becomes associated with positive outcomes or rewards, their utility or subjective desirability as a social partner increases. Social rewards take many different forms, for example, positive interactions are rewarding (Rilling et al., 2002), and individuals who often provide them should be desirable interaction partners. Reciprocity of trust (Singer, Kiebel, Winston, Dolan & Frith, 2004), positive behaviour or outcomes (Behrens et al., 2008; Izuma, Saito & Sadato, 2008) and positive expressions (Tsukiura & Cabeza, 2008) are also rewarding and thus should enhance social utility. High social utility therefore reflects that a person is both objectively valuable and subjectively desirable, such that they lead to rewards. Conversely, social punishments, such as negative behaviour or expressions (Averbeck & Duchaine, 2009, Fehr & Camerer, 2007) should decrease an individual's social utility.

Reinforcement Learning

Social utility may help to quantify the subjective components of social decisions. However, it does not explain how people learn about others, make predictions from cues, or update estimates of utility in the face of contradictory evidence. However, reinforcement learning is a theoretical approach that may do so. Reinforcement learning models characterize the ways in which reinforcements and punishments allow learners to adapt to particular environments (Schultz, 2006). Applied to social environments, these

models propose that people learn social information through the association of rewards (e.g., cooperation that leads to financial gain) or punishments (e.g., defection that leads to financial loss) with social actions (Behrens et al., 2008). Consequently, social partners' may acquire 'value' through association with good or bad social outcomes. This value reflects expectations about the utility or subjective desirability of interacting with a particular social partner based on a history of experienced social outcomes with that partner (Chang, Doll, van t'Wout, Frank & Sanfey, 2010).

Within these models, each possible decision or action is associated with an 'expected value'. The expected value is an estimate about how rewarding the outcome of a decision is anticipated to be, multiplied by the probability that the outcome will occur (Sutton & Barto, 1998). The expected value of a decision derives from the value of the associated stimuli. Each stimulus associated with a particular decision is assigned a value (Rushworth, Mars & Summerfield, 2009), the individual values then combine to give each possible choice a total expected value.

In social terms, an interaction partner's social cues likely provide important information about how he or she may behave and thus what to expect from an interaction. Behaviour can then be adjusted based on an interaction partners' cues and the value of the outcomes they predict. Therefore, according to this idea, the value of an interaction partner's cues should combine to create an expectation about how likely that person is to provide social rewards. This expectation, guides perceptions of, and behaviour toward interaction partners. Logically, humans aim to maximize rewards and so should choose the behaviour, or interaction partner, with the highest expected value (O'Doherty, 2004). However, in real-world environments research suggests that people do not maximize gains, and in fact make 'irrational' choices (Camerer, 2003; Erev & Roth, 1998).

After a choice is made, the chosen action is performed. The true outcome of the chosen action is then experienced, and can be evaluated. The decision is evaluated by calculating the difference between the value of the actual outcome and the expected outcome. When the actual value of an outcome differs from the expected value, a 'prediction error' occurs. An outcome that is better than anticipated creates a positive

prediction error; conversely a worse-than-expected outcome constitutes a negative prediction error (Sutton & Barto, 1998). The size and direction of a prediction error changes the decision-maker's estimate of the expected value of the decision, such that a positive prediction error leads to a higher expected value and negative prediction errors lead to lower expected values. Under reinforcement-learning models, decisions with better-than-expected outcomes will be more likely to be repeated, and those with worse-than-expected outcomes are less likely to be repeated (Cohen & Ranganath, 2007). For example, if an interaction partner's cues predict positive outcomes (e.g., a happy expression from a trustworthy face) but the experienced behaviour is negative (e.g., a failure to reciprocate trust), trust is less likely to be placed with that partner in the future. Thus, by experiencing the outcomes of their social actions, people learn about the social environment and adapt future behaviour accordingly (Behrens, Hunt, & Rushworth, 2009). Consequently, people are more likely to trust social partners when they have a history of providing sound advice (Behrens et al., 2008) or cooperating in a game (Tomlin et al., 2006). This is because those partners' expected utility has increased via a positive behavioural history. A current social utility estimate then guides decision-making with each social partner.

Examining social decision-making from the perspective of reward learning may provide insight about how an interaction partner's social cues shape choice behaviour. That is, the relationships between social cues and outcomes during social interactions may enable people to predict the likelihood of rewards, and thus assess the value of interaction partners. Some social cues may be better predictors of rewards than others. Moreover, the same social cue from an interaction partner may not always predict the same outcome or reward. In this way, social cues likely interact to influence both decision-making and behaviour in social environments.

The present research aims to understand how the information contained in social cues shapes decision-making processes. It focuses on three types of social cue, which are present in most social interactions: static appearance-based cues such as how trustworthy or attractive someone looks, dynamic expressive cues such as facial expressions, and

behavioural cues such as the reciprocity of trust. This research examines how these types of cues shape social decision-making, both individually and in combination, and examines one mechanism for understanding how these cues might achieve their joint effects.

Static Appearance-based Cues

In both reward learning and utility models objects have an expected value. In a social environment the first piece of information available about a social partner is their appearance. Therefore, it is likely that a social partner's expected value is initially inferred from their appearance. People often admonish one another for 'judging a book by its cover'. Nonetheless, as decision-making theories would predict, this is exactly what they do. When viewing faces, people make a range of judgements about the individuals they depict, for example how attractive (Little, Jones & DeBruine, 2011), competent (Todorov, Mandisodza, Goren & Hall, 2005), likeable (Willis & Todorov, 2006), powerful (Hassin & Trope, 2000) and trustworthy they are (Bar, Neta & Linz, 2006). Indeed Lavater's (1880) *'Essays on Physiognomy'* provide a guide to reading personality characteristics from facial features. Although Lavater's guide is unrealistic, as specific facial features are unlikely to signify the qualities of a personality, it accurately suggests that the face is an important source of information. An interaction partner's face displays a variety of signals from which social information can be inferred, from personal characteristics to social intentions (Rezleszcu, Duchaine, Olivola, Chater, 2012; Willis & Todorov, 2006). Indeed, face perception plays a critical role in social interactions (Haxby, Hoffman & Gobbini, 2002) and behaviour towards interaction partners (Zebrowitz & Collins, 1997).

In utility theory terms, the way someone looks likely informs the subjective desirability of an interaction with that person. Research shows that, first impressions are formed rapidly, reliably and relatively automatically (Bar, Neta & Linz, 2006; Berry & Brownlow, 1989; Hassin & Trope, 2000; Olsen & Marshuetz, 2005; Willis & Todorov, 2006). These 'appearance-based' social judgements influence a wide range of important social decisions from job offers (Cash & Kilcullen, 1985), to voting (Little, Burriss, Jones & Roberts, 2007; Olivola & Todorov, 2010), to risk taking (Schlicht et al., 2010) and legal

sentencing (Porter, ten Brinke & Gustaw, 2010). Although people infer a variety of traits from appearance, research suggests that faces are categorized on two primary dimensions, one of which is trustworthiness (Todorov, Said, Engell & Oosterhof, 2008). Trustworthiness, however, is highly correlated with attractiveness (Wilson & Eckel, 2006). Appearance-based trustworthiness judgements of static faces have a significant influence on decision-making processes (Schlicht et al., 2010; van t' Wout & Sanfey, 2008). In terms of making rational decisions, it is important to understand how an interaction partner's static cues shape decision-making.

The 'kernel of truth' hypothesis posits that some personality characteristics or information about an individual can be accurately read from the face (Bond, Berry & Omar, 1994). According to this hypothesis, using appearance-based judgements to guide decision-making is not only rational, but also likely to lead to advantageous decision-making. Studies supporting this idea have shown correlations between an individual's self-reported personality traits and others' ratings of that individual (e.g., Berry & Brownlow, 1989). However, other studies have found no correlation between facial judgements and behavioural measures such as choosing to deceive others (Hassin & Trope, 2000; Masip, & Garrido, 2001; Olivola & Todorov, 2010; Porter, England, Juodis, van Brinke & Wilson, 2008; Zebrowitz, Voinescu & Collins, 1996). This suggests that even if these judgements are sometimes accurate they may not be accurate under all circumstances.

Despite mixed evidence about the accuracy of appearance-based judgements, what does seem to be clear from this literature is that people make strong appearance-based social judgements and use them to inform decision-making (Oosterhof & Todorov, 2008; Little, Burriss, Jones & Roberts, 2007; Schlicht et al., 2010; van t' Wout & Sanfey, 2008). It is therefore important to understand the function of these judgements and the influence they have on social decision-making. One possible function of these judgements is to guide initial expectations of how desirable an interaction with a person may be. Thus they provide an initial utility estimate that can be used to guide decision-making until more, and more reliable information becomes available. In terms of rational decision-making, character judgements from faces should have limited influence on choices

because if these judgements are inaccurate they will result in decision-making failures, meaning that people do not achieve the best possible social outcomes.

One theory for why appearance based judgements affect behaviour so strongly proposes that an interaction partner's face provides information about their likely social behaviour (Cosmides, 1989; Cosmides & Tooby, 2000; Frith & Frith, 1999; Gibson, 1979). Thus, first impressions of an interaction partner may contribute to their social utility (e.g., Chang, Doll, van t'Wout, Frank & Sanfey, 2010). Specifically, when no additional information about an interaction partner is available, appearance may shape expectations about the desirability of interaction. According to this idea, social judgements may be tantamount to social utility. This explains why initial impressions bias behaviour and decision-making processes involving social partners (Chang et al., 2010; Delgado, Frank & Phelps, 2005; van t'Wout & Sanfey, 2008). For example, interaction partners who look attractive or trustworthy are believed to be more likely to provide positive experiences and social rewards (Wilson & Eckel, 2006). Thus, appearance-based judgements influence people's expectations about potential social partners, such that the utility of interacting with positively judged partners is high and the opposite is true of negatively judged partners.

In summary, appearance-based social judgements likely reflect the expected utility of interaction with a social partner, thereby accounting for their influence on decision-making. Reliance on appearance-based impressions of trustworthiness should diminish as other information about a person's behaviour becomes available, as first impressions may not allow people to maximize possible gains from the social environment. Rather, people's utility estimates should incorporate behavioural information, as behaviour is a more reliable indicator of an individuals' true social value. Accordingly, in real-world social encounters the additional cues people send enable more accurate appraisals of trustworthiness, which in turn influence behaviour (Frank, Gilovich & Regan, 1993).

Behavioural Cues

In social environments, people typically have access to behavioural information, as

well as appearance-based information about a social partner. Moreover, an individual's recent behaviour is likely the best indicator of that partner's true or objective social value (King-Casas et al., 2005). Consequently, receiving explicit information about an interaction partner's previous behaviour shapes decisions involving that partner (Delgado et al., 2005; Singer et al., 2004), as does the experience of an actual interaction history. For example, research shows that positive experience with an interaction partner, such as truthful (Bayliss & Tipper, 2006), consistent (Heerey & Velani, 2010), or prosocial behaviour (Ames & Johar, 2009) alters the partner's social utility, leading to more positive interpersonal evaluations, which are indicative of value.

Although social utility is initially derived from rapidly formed appearance-based judgements, behavioural experience with an interaction partner should modify those impressions to form an integrated percept about a social partner's utility (Chang et al., 2010). Essentially, initial impressions of a social partner's utility guide decision-making until one obtains more reliable behavioural evidence about a social partner. Decisions related to a partner should evolve to reflect changes in social utility based on experienced behaviour (Chang et al., 2010). Therefore, the degree to which an initial impression of trustworthiness influences decisions should decline as more behaviour is experienced.

Indeed, appearance-based impressions about people should not be used once behavioural information becomes available, as appearance may be misleading. Nonetheless, some research suggests that even after providing explicit knowledge about an interaction partner's behaviour (e.g., "this person has not behaved fairly") initial appearance-based judgements continue to shape decision-making (Delgado et al., 2005; Rezleszcu et al., 2012). Thus, initial appearance-based impressions of trustworthiness may bias expectations about and behaviour toward a social partner, even when information about that partner's reputation is apparent.

Taken together, this work suggests that both appearance- and information-based social judgements may influence social decision-making. If first impressions have a longstanding influence on social utility, despite access to more reliable information, this is likely to lead to reduced decision quality. However, in real world social environments,

dynamic expressive displays modify static or so-called “unfakeable” social cues, such as the appearance of a person’s face in a neutral pose. Because these displays may also carry social value, they may mitigate the information that static cues convey, enhancing or diminishing the impression of utility that an individual’s neutral face conveys.

Dynamic Expressive Cues

In real world social environments, unlike computer displays in the laboratory, neutral static faces are rarely seen. Instead faces are dynamic objects that display many different expressions. Facial expressions quickly communicate information (Kringlebach & Rolls, 2003), such as emotions (Blair, 2003; Ekman, 1993; Horstmann, 2003; Morris & Keltner, 2003) and intentions between people (Fridlund, 1991). Expressions change the interpretation of other peoples’ behaviour (Ames & Johar, 2009). Given the indication expressions provide about the probable behaviour of an interaction partner, they should contribute to social utility as they enable individuals to infer more about what should be expected from that individual.

A longstanding evolutionary argument suggests that facial expressions may provide evolutionary advantage and are therefore adaptive (Darwin, 1965/1872). For example, a brow lowered in anger may protect the eyes if a social confrontation turns into a physical fight. Expressions such as smiles may be similarly adaptive because they serve to facilitate contact between people, particularly in the mother-infant dyad (Bowlby, 1958). Like static appearance cues, facial expressions result in judgements of people’s personality characteristics and traits (Zebrowits & Montepare, 2008). Indeed, some research suggests that judgements made from static appearance cues reflect overgeneralization of features that look like emotional expressions (Oosterhof & Todorov, 2009; Zebrowits & Montepare, 2008).

As communicators of intentions, facial expressions may predict social outcomes (Fridlund, 1991; Hooker, Germine, Knight, & D'Esposito, 2006). For example, a smile from an interaction partner may lead a receiver to anticipate a positive social outcome whereas a frown indicates a negative one (Kringelbach & Rolls, 2003). If this is true then according

to a reward learning perspective facial expressions may have 'value' that depends on the value of the outcome they predict. This value may therefore influence the utility of the social partner displaying that expression (Gesn & Ickes, 1999).

Smiles are particularly important social cues. They occur frequently in face-to-face social exchanges (Heerey & Kring, 2007; Hess & Bourgeois, 2010) and serve to communicate both emotion and social intentions (Ekman, 1993; Fridlund, 1991). In addition, they have a strong impact on real-world behaviour. Displaying smiles can influences generosity of tipping (Tidd & Lochard, 1978), electoral candidate choice (Mullen et al., 1986), and sentencing decisions (LaFrance & Hecht, 1995). Previous research shows that smiling faces are judged as more trustworthy than unsmiling faces, and influence decision-making in economic games (Scharlemann, Eckel, Kacelnik, & Wilson, 2001). In fact, research has shown that different facial expressions have differential effects on behaviour. For example, happy expressions increase the probability that the preceding behaviour which caused the expression, will be repeated (Matthews & Wells, 1999), whereas sad expressions have been found to make the preceding behavioural response less likely (Blair, 1995). Together, this research suggests that facial expressions may function as outcome-predicting cues (Hooker et al., 2006).

The present research focuses on smiles, as they are among the most commonly seen facial expressions. In particular, this work compares two distinct types of smiles: genuine smiles of pleasure and polite smiles. Genuine smiles of pleasure are caused by the experience of positive emotion and involve the activation of the orbicularis oculi, the muscle that rings the eye (Ekman, Davidson & Friesen, 1990). Comparatively, polite smiles do not require positive emotional experience and do not activate the orbicularis oculi (Frank, Ekman & Friesen, 1993). If these smiles have value they may change the utility of interaction partners who present them. For example, anecdotally, genuine smiles are rewarding and may therefore add value to interaction partners, thus increasing social utility.

In summary, dynamic expressive cues such as smiles, may add to social utility in several ways. First, they indicate what behaviour is likely from an interaction partner, thus

influencing initial expected utility estimates. Second, due to their predictive qualities, they may influence subjective desirability, for example due to their association with rewarding or positive outcomes smiles may be subjectively desirable.

Interactions between Social Cues

Individually, static appearance-based cues, behaviour, and dynamic facial expressions all provide information about the value of an interaction partner. Given that the social world is both dynamic and complex it is unlikely, in real-world environments, that these cues are seen in isolation. Instead these cues interact with each other creating an integrated percept of an interaction partner's social utility. Accordingly, research shows that knowledge of previous social behaviour and appearance-based social cues combine to inform trait judgements (Quist, DeBruine, Little & Jones, 2012). Other research shows that facial expressions change how an individual's behaviour is interpreted (Ames & Johar, 2009) and also change appearance-based judgements (Otta, Folladore & Hoshino, 1996). Less is known about how these types of cues interact over time. A rational model of decision-making would posit that while dynamic cues may add to or subtract from behavioural utility, over time behaviour would be the main contributor to social utility. Further, initial expectations of utility should be shaped over time by the cues an individual presents. Over-reliance on a particular social cue could lead to biased social utility estimates, which may lead to irrational decisions and a failure to maximize gains. This is especially likely to be the case if the cues an interaction partner presents conflict in terms of the behaviour they predict.

Influence of Social Cues on Decision-making

Social cues do not function in isolation. Instead cues interact to shape the informational landscape of a social interaction. As discussed above, each individual cue type has an influence on decision-making behaviour. However, in social interactions all of these cues occur concurrently. Thus, it is the combined influence of cues that drives the utility of a social partner. However, the information provided by an individual's cues may

not always agree. If cues conflict rather than concur (e.g., an individual who smiles whilst abusing trust), how do people integrate them to formulate a unified estimate of another's social utility? In order to understand social decision-making, it is important to understand the contribution or weighting of different cues in utility estimates. In reality, some cues matter more than others, for example dynamic cues can mitigate the impact of bad behaviour or elevate the value of good (Ames & Johar, 2009). However, this is only true for a while. Logically, in the end what matters more is behaviour, especially when it is consistent. One way to understand how these cues interact to influence social decisions is by applying a reinforcement learning model to make predictions about how differences in a set of cues change behaviour (Behrens et al., 2008).

Another question that remains is by what mechanism social cues influence decision-making. The social information contained in the cues people see is valuable. People pay attention to others in order to acquire it (Klein, Shepherd & Platt, 2009). One idea is that certain cues may bias attention in social settings. If a particular cue draws attention when it occurs, this may shape the information people acquire and therefore use to inform subsequent decisions.

Attention

People's capacity to process information is limited (Desimone & Duncan, 1995). Due to this limitation, attentional processes selectively direct attention towards salient or goal relevant stimuli (Land & Heyhoe, 2001). Social environments are complex and moment-to-moment social partners provide many changeable appearance- and behaviour-based cues. Social interactions, especially those consisting of multiple partners, contain more information than people are able to process (Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010).

Faces are arguably the most important social stimuli in the human environment, and as such, receive preferential processing when compared to other types of stimuli (Palermo & Rhodes, 2007). Faces expressing emotions are particularly salient. Research shows that they are pre-attentively processed compared to other types of stimuli (De

Martino, Kalisch, Rees & Dolan, 2009; Fox et al., 2000). However, valuable stimuli are also pre-attentively processed (Anderson, Laurent & Yantis, 2011; Dux & Marois, 2009; De Martino, Kalisch, Rees & Dolan, 2009). If unpleasant, untrustworthy looking faces and pleasant, trustworthy looking faces are simultaneously present, which receives more attention? This is one mechanism by which social utility judgements may influence social decisions and behaviour. Specifically, if an interaction partner can capture attention, even for a short while, that partner can bias the information one receives during an interaction. Therefore, social utility judgements may bias decision-making by guiding attention toward or away from particular partners, consequently determining the information people acquire and use to make subsequent decisions.

Conclusions

This research uses a reinforcement learning perspective to investigate how different types of social information contribute to social utility. In turn, it examines how social utility judgements influence decision-making processes, and investigates one possible mechanism for this influence. By examining people's responses to static appearance-based cues, dynamic expressive cues, and behavioural cues such as the reciprocity of trust, this research examines several specific questions. First, how do dynamic expressive cues such as smiles, influence social utility? Second, how do people integrate different types of social cues to formulate a cohesive estimate of a social partner's utility? Third, how do social utility estimates based on these cues guide selective attention?

Chapter Two

The Value of Genuine and Polite Smiles

Danielle M. Shore and Erin A. Heerey

Paper published in Emotion, volume 11, 2011, pages 169-174.

Abstract

Humans show remarkable ability to adapt their social behaviour to suit the changing requirements of their interactions. An interaction partner's social cues, particularly facial expressions, likely play an important role in motivating and reinforcing this behavioural adaptation. Over four studies, we test a key aspect of this idea. Specifically, we ask how the reinforcement value of facial expressions compares to that of nonsocial feedback and to what degree two frequently occurring expressions (genuine and polite smiles) differ in reinforcement value. Our findings show that social feedback is preferred over nonsocial feedback and that genuine smiles are preferred over polite smiles. Based on a logistic model of our data, we show that both monetary and social values of stimuli contribute significantly to participants' decisions. Indeed, participants were willing to sacrifice the chance of a monetary reward to receive a genuine smile and produced inflated estimates of the value of genuinely smiling faces. These findings suggest that genuine smiles, and potentially other social cues, may be useful social reinforcers and therefore important in the control of social behaviour on a moment-to-moment basis during interaction.

Humans possess extraordinary ability to adapt their social behaviour to an interaction's changing demands (Lynch, 2007), suggesting the presence of a highly flexible system for navigating social interactions. One candidate mechanism for understanding social ability is reinforcement learning. According to recent accounts (e.g., Behrens, Hunt, & Rushworth, 2009), people extract information about the social environment based on the outcome values of social actions, e.g., deciding whether to trust a partner based on his/her history of providing sound advice (Behrens, Hunt, Woolrich, & Rushworth, 2008) or cooperating in a game (Tomlin et al., 2006). In nonsocial environments, reinforcement learning has received widespread attention as a model for human and animal behaviour (Dayan & Balleine, 2002; Schultz, 2006). This work shows that reinforcement learning systems are highly flexible (Tremblay & Schultz, 1999), allowing finely tuned adaptation to environmental contingencies (Schultz, Dayan, & Montague, 1997). Here, we extend this research by testing the idea that social cues, specifically smiles, carry reward value and may therefore reinforce social behaviour.

There are two ways in which facial expressions may be rewarding. First, they might possess innate emotional value (Ekman, 1992) or cause emotion in receivers (Geday, Gjedde, Boldsen, & Kupers, 2003). Second, they may predict social outcomes (Fridlund, 1991; Hooker, Germine, Knight, & D'Esposito, 2006). For example, interaction partners' smiles may lead receivers to anticipate positive social outcomes whereas frowns suggest otherwise (Kringelbach & Rolls, 2003). As with nonsocial cues, which acquire value depending on the outcomes they predict (O'Doherty, 2004; Schultz, 2004), facial expressions may acquire value via a history of cue-outcome pairings. Regardless of whether this value is innate or acquired, for clarity we call it "intrinsic."

The idea that facial expressions carry intrinsic value is important for understanding how they might shape behaviour in real-world interactions. Much of the literature has focused on the emotional qualities of expressions. A question that remains unanswered is whether facial expressions can be understood in terms of their ability to shape receiver behaviour. If facial expressions carry intrinsic value, they may have a more important role in guiding social behaviour than previously reported (Gesn & Ickes, 1999). Moreover, the

ability to understand facial expressions in terms of intrinsic values would help explain diverse findings showing, for example, that emotional faces modulate cognitive performance (Banich et al., 2009; Tsukiura & Cabeza, 2008). In four experiments, we test whether facial expressions carry intrinsic value, whether different expressions have different values, and how participants value social relative to monetary rewards.

Experiment One

In a contingency learning game, we compared the relative ability of social and nonsocial feedback (both associated with the same monetary reward) to shape choice behaviour. We predicted that if both types of feedback had equal intrinsic value, then participants' behaviour would simply reflect relative differences in the likelihood of receiving monetary reward. However, if one feedback type is intrinsically more rewarding than the other, behaviour should reflect both the relative likelihood of receiving monetary rewards and the added value of the intrinsically rewarding feedback.

Methods

Participants

Forty-two undergraduate psychology students (7 male; mean age=19.88, SD=1.15) completed the study for course credits and bonus money. Participants gave written informed consent before participating. The local ethics committee approved the study (likewise for Experiments 2, 3 and 4).

Procedure

Participants completed a "matching pennies" game with four computerized opponents, who provided rewards with different frequencies. A unique image identified each opponent (two photos of people, two of computers). The task had a learning phase followed by a test phase. In learning-phase trials, participants viewed a photo of one opponent in a neutral pose and selected "heads" or "tails" of a coin with a key press. Participants attempted "to choose the same side of the coin" as the opponent. After each choice, the opponent provided feedback (Figure 2.1A). Human opponents smiled to

indicate matches and frowned to indicate non-matches. Computer opponents displayed either green ticks (matches) or red crosses (non-matches). Match feedback was worth 2 pence and non-match feedback was worth 0 pence. Participants played each opponent 40 times, divided randomly among 4 blocks of 40 trials (160 learning trials).

Unbeknownst to participants, feedback occurred independently of their behaviour. Instead, opponents provided rewards at different rates (85%, 70%, 55% and 40%). For example, the 85% opponent provided reward feedback on 85% of trials and non-reward feedback on 15% of trials, regardless of behaviour. To compare the value of social and non-social feedback one of the higher-frequency reward contingencies (either 85% or 70%) was randomly assigned to a computer opponent, as was one of the lower-frequency contingencies (55% or 40%). The other two reward contingencies were assigned to human opponents. This ensured that, across all participants, human and computer opponents provided reinforcement with approximately equal probability. Contingencies remained the same across both learning and test phases.

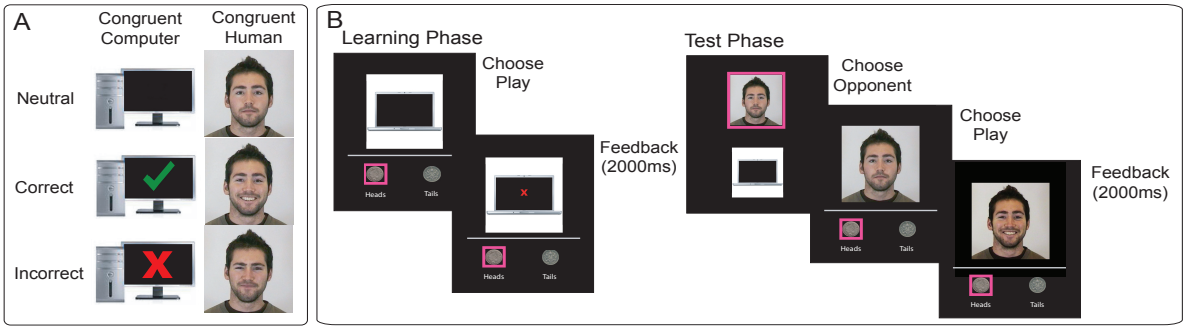

Figure 2.1. Experiment 1 stimuli and game. A) Example stimuli from the task, showing correct and incorrect feedback for both human and computer opponents. B) Example trials from both learning and test phases of a matching pennies game. In learning trials, participants saw one of four opponents, indicated whether they wished to choose "heads" or "tails" of a coin via a key press, and received feedback from the opponent. In test phase trials, participants first chose which opponent they wished to play from given pair of opponents, thereafter trials continued as in the learning phase.

In test-phase trials, participants chose which opponent they wished to play from a given pair (Figure 2.1B). Thereafter, trials proceeded as in the learning phase. There were six possible two-opponent pairs (opponents 1, 2; opponents 1, 3; etc.). Participants saw each pair 4 times in random order (24 test trials). Finally, participants ranked each opponent from 1 (most frequently rewarded) to 4 (least frequently rewarded), as an

explicit measure of reward-contingency knowledge. At the end of the task, participants received the bonus money they had earned (average=£2.50). The task was programmed using the Psychophysics Toolbox (Brainard, 1997) in Matlab (The MathWorks).

Results

Figure 2.2A shows the average number of times participants chose opponents during test-phase trials, according to reward probability. A repeated-measures ANOVA examining the effects of reward probability (85%, 70%, 55%, 40%) on choice behaviour showed that participants learned the contingencies, $F(3,41)=9.92$, $p<.001$; $\eta^2_p=.20$.

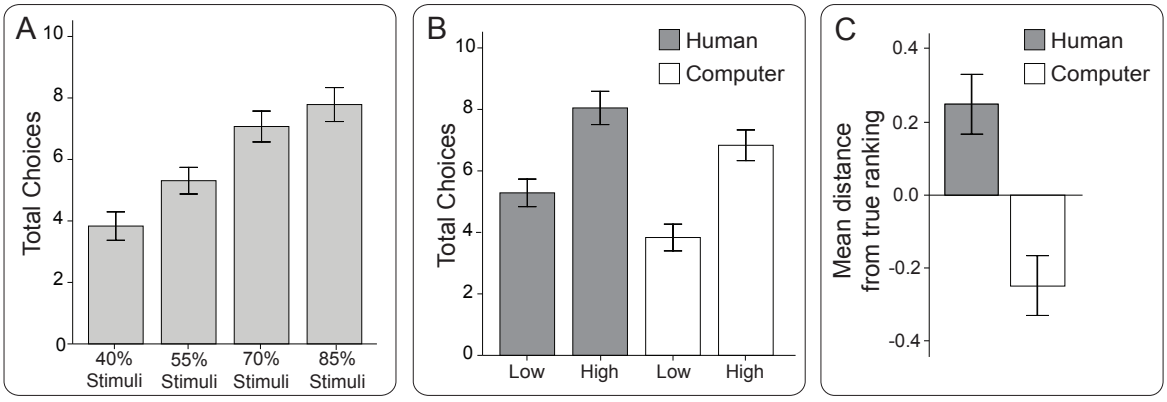

Figure 2.2. Experiment 1 results. A) Average number of times participants chose opponents of each reward probability (40%, 55%, 70%, and 85%). B) Average number of times participants chose human and computer opponents with high or low reward probabilities. C) Average rankings of computer and human opponents' reward frequency compared to opponents' true rankings (calculated by computing the difference between participants' ranking and opponents' 'true' ranks based on reward likelihood). Positive scores indicate that opponents were ranked as better than their true ranks. Error bars indicate +/- 1 SEM.

A paired-samples t-test confirmed there was no difference in the frequency of rewards from human versus computer opponents $t(41)=1.10$, $p=.28$, Cohen's $d=0.34$. However, a 2-factor (value: higher or lower reward likelihood; identity: human or computer) repeated-measures ANOVA on the choice data showed main effects of both value, $F(1,41)=22.39$, $p<.001$; $\eta^2_p=.35$, and identity, $F(1,41)=5.42$, $p=.03$; $\eta^2_p=.12$, suggesting that participants preferred social feedback, even though humans and computers provided monetary rewards equally frequently (Figure 2.2B) .

Finally, Figure 2.2C shows the average rank participants assigned to computer and human opponents, relative to opponents' actual ranks. A t-test showed that participants ranked human opponents as better and computers as worse than they actually were, $t(41)=3.04$, $p=.004$, Cohen's $d=0.95$.

Discussion

These results show that although participants learned contingencies regardless of opponent type, they treated human opponents as if they were more valuable than computer opponents, despite a lack of objective value differences. This finding may simply show that participants prefer human opponents; however, participants knew that all opponents were computer controlled, suggesting that this result cannot be explained by a simple preference to play human opponents (Aharoni & Fridlund, 2007; Moretti, Dragone, & di Pellegrino, 2009). However, to test whether a preference for human opponents or a preference for social feedback explains this result we conducted a control experiment in which a human opponent provided non-social feedback (ticks and crosses) and a computer opponent provided social feedback (smiles and frowns).

Experiment Two

We predicted that if participants preferred social feedback they would choose opponents that provided social feedback whether they were depicted as a human or a computer. If however participants preferred human opponents, they should choose to play human over computer opponents regardless of the type of feedback they provide.

Methods

Participants

Thirty-four undergraduate psychology students (15 male; mean age=19.50, SD=1.71) completed the study for course credits and bonus money (average £3). One participant was excluded for being an outlier, showing a preference for social feedback that was 2.75 SDs below the sample mean (N=33).

Procedure

Participants completed a "matching pennies" game with two human and two computer opponents as in Experiment 1. To differentiate between human opponents and social feedback one human and one computer opponent provided social feedback (smiles and frowns) and one human and one computer opponent displayed non-social feedback (ticks and crosses). Therefore two of the opponents (1 human and 1 computer) provided feedback congruent with the identity of the opponent as in Experiment 1. The remaining opponents provided incongruent feedback such that the human opponent indicated matches and non-matches with a green tick or a red cross superimposed over the neutral face respectively (Figure 2.3). The incongruent computer opponent displayed an image of a genuinely smiling face on its screen to indicate a match and a frowning face to indicate a non-match (photographs of human faces were edited to fit the computer screen in the image). The value of feedback was the same as Experiment 1 (match=2 pence and non-match=0 pence) and all opponents rewarded at a rate of 70%.

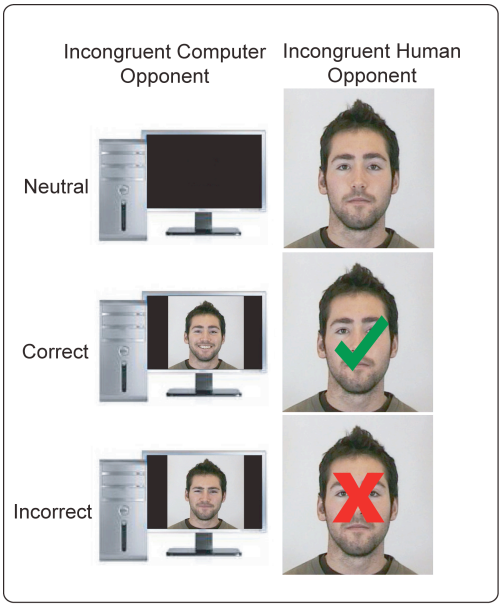

Figure 2.3. Experiment 2 stimuli. Example stimuli for the incongruent feedback opponents; incongruent computers provided social feedback and incongruent humans provided non-social feedback.

Participants played each opponent 20 times in random order (80 learning trials). On test-phase trials, as in Experiment 1, participants chose which opponent they wished

to play from a pair of opponents. All six possible opponent-pairings were tested. Participants completed 60 test trials, ten per pairing, in random order.

Results

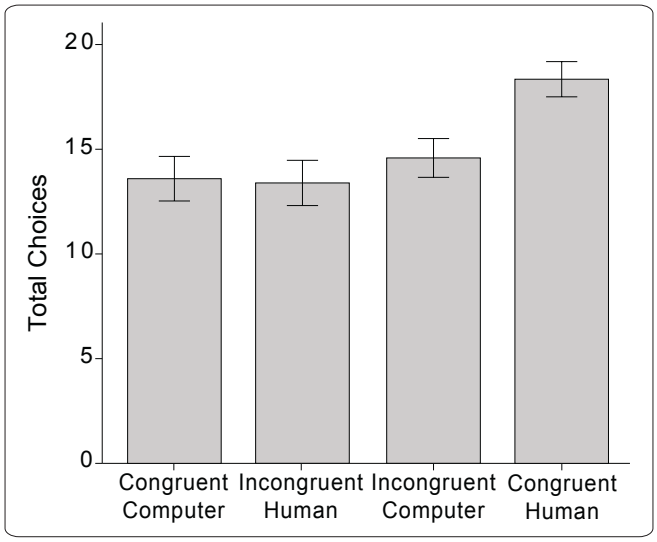

Figure 2.4. Experiment 2 results. Average number of times participants chose opponents of each feedback type. Congruent computer and incongruent human provided non-social feedback and congruent human and incongruent computer provided social feedback. Error bars indicate +/- 1 SEM.

Figure 2.4 shows the average number of times each opponent type was chosen. A 2-factor (opponent type: human or computer; feedback congruency: congruent or incongruent) repeated-measures ANOVA on choice behaviour showed a main effect of opponent type, $F(1,31)=4.07$, $p=.05$, $\eta^2_p=.12$ such that participants had a general preference to play human opponents. This was driven by choices of the congruent human opponent, who was the most preferred opponent (M=18.34, SD=4.76). There was no main effect of feedback congruency, $F(1,31)=2.36$, $p=.14$, $\eta^2_p=.07$, suggesting congruency did not drive participants choices. A significant opponent-type by feedback-congruency interaction showed participants preferred social feedback $F(1,31)=5.70$, $p=.023$, $\eta^2_p=.16$, such that participants chose the incongruent computer (social feedback, M=14.34, SD=5.25) more than the incongruent human (M=13.72, SD=6.20) and the congruent computer (M=13.59, SD=6.02; both nonsocial feedback). Together, these results suggest

that participants preferred social feedback to nonsocial feedback, even when this feedback came from a computer opponent.

Discussion

These results show that participants' choice behaviour in Experiment 1 and 2 is explained by a preference for social feedback rather than human opponents. This suggests that although social and nonsocial feedback carried the same monetary reward, social feedback was intrinsically more valuable to participants.

Experiment Three

In Experiment 2 we showed that participants had a preference for social feedback, suggesting social feedback is valued higher than non-social. In Experiment 1 and 2, opponents always displayed genuine smiles, meaning that participants' preference for social feedback might be specific to genuine smiles. Genuine smiles are distinguishable from non-genuine smiles based on the action of orbicularis oculi (Frank, Ekman, & Friesen, 1993). Genuine smiles have social (Fridlund, 1991) and emotional meanings (Ekman, 1992) and may have evoked positive feelings in our participants (Surakka & Hietanen, 1998), thereby inflating preferences for social feedback opponents. To test whether genuine smiles are intrinsically rewarding, we compared genuine- with polite-smile feedback. We predicted that if genuine smiles were intrinsically more rewarding than polite, participants would prefer genuinely smiling opponents.

Methods

Participants

Thirty-six undergraduate psychology students (18 male; mean age=19.56, SD=1.90) completed the study for course credits and bonus money (average=£3).

Procedure

Participants completed the same matching pennies game as in Experiment 2; however, in this version each opponent was identified by a unique human image (two

male, two female). Two opponents (one male, one female; randomly assigned) always provided genuine smiles and the others always provided polite smiles. Each opponent provided rewards with a probability of 0.7, so that participants received reward feedback (genuine or polite smiles) on 70% of trials and non-reward feedback (frowns) on 30% of trials. Regardless of smile type, reward feedback was worth 2 pence. Learning and test-phase trials ran as in Experiment 2 (trial and block numbers were the same).

Stimuli

The smile stimuli used in this and the other experiments consisted of still images of actors displaying genuine and polite smiles. We created the stimuli by asking eight actors (4 male) to pose a variety of facial expressions (including neutral poses and frowns). Expressions were recorded with a high-definition, digital camcorder. Actors posed each expression 8 times.

To obtain genuine smiles, we induced positive emotion by asking our actors to imagine/re-experience a situation in which they felt happy and display their happiness as if they were sharing the experience with a good friend. To obtain polite smiles, we asked actors to produce polite smiles after seeing them demonstrated.

From each of these short films, we chose the first frame at which the expression was at its peak and created a still image from that frame. We then selected, for each actor, the 5 photos for each expression that most closely resembled a prototypical expression of that type. To validate the smile stimuli, we conducted a separate study in which 60 participants (21 male; mean age=20.53, SD=3.55) viewed each smile photo for 1 second and classified it, as quickly as possible, as either a genuine or polite smile. There were no differences in mean reaction times to the two smile-types (genuine=0.97 seconds, SD=0.10; polite=0.96 seconds, SD=0.09); $t(59)=-1.55$, $p=.12$, Cohen's $d=-0.40$). We calculated d' for each participant as a measure of ability to distinguish between the smiles (Wickens, 2002). Overall, participants showed good ability to discriminate the smiles (mean d'=1.59, SD=0.39; which, for an unbiased observer, is about 75% correct).

We then calculated the proportion of participants who correctly identified each smile in the set. Based on these results, we chose the most frequently correctly identified

genuine and polite smile for each actor to use in the present studies. All the smiles we used were correctly identified at rates significantly greater than chance (p-values<.01). The average proportion of correct identifications for the 8 genuine smiles was 0.86 (SD=0.02); for polite, it was 0.84 (SD=0.03). There were no smile-type differences in the rates at which participants identified these smiles, t(14)=1.01, p=.33, Cohen's d=0.54.

Results

Given that opponents provided rewards with equal frequency, we predicted that if genuine smiles had intrinsic value then participants would prefer genuinely to politely smiling opponents. A 2-factor (smile type; genuine or polite; opponent sex: male or female) repeated-measures ANOVA showed that participants chose genuinely smiling opponents more often than polite ($M_{Genuine}$=37.25, $SD_{Genuine}$=9.39, M_{Polite}=22.75, SD_{Polite}=9.39; F(1,35)=21.44, p< .001; η^2_p=.38). There was no main effect of sex and no interaction (p-values>.30). Figure 2.5 shows the average number of times participants chose opponents according to smile type.

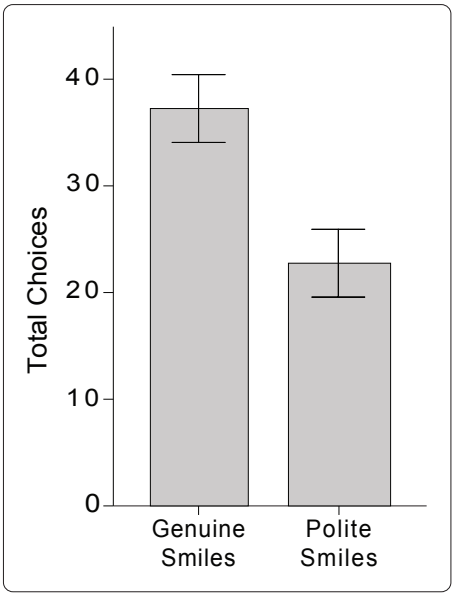

Figure 2.5. Experiment 3 results. Average number of times participants chose opponents according to smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Discussion

As anticipated, participants preferred genuinely to politely smiling opponents, suggesting that genuine smiles have greater intrinsic value than polite smiles. However, in this study, all opponents provided monetary rewards with equal probability, meaning that their expected monetary value was identical (the expected value of a stimulus is calculated as the probability of receiving a reward, multiplied by the value of the reward, Sutton & Barto, 1998). In order to unequivocally demonstrate that genuine smiles possess intrinsic value, Experiment 4 decoupled monetary value from smiles, allowing direct comparison of the relative values of money and smiles.

Experiment Four

We predicted that if genuine smiles had intrinsic reward value, participants would sacrifice the chance to gain a monetary reward in order to obtain a genuine smile. Measuring the degree to which smiles influenced participants' choices, allowed us to estimate the value of a smile in monetary terms.

Methods

Participants

Thirty-six undergraduate psychology students (14 male; mean age=20.58, SD=2.76) completed the study for course credits and bonus money (average=£3). We excluded one female participant for failing to follow task instructions (N=35).

Procedure

Participants completed the same matching pennies task as in Experiment 3, with two opponents providing genuine and two providing polite smile feedback. To compare the value of smiles to money, we altered the rates at which the opponents provided rewards. In this experiment, two opponents (one genuinely, one politely smiling; randomly assigned) provided rewards on 70% of trials. The other two opponents rewarded participants on 80% of trials. As before, both smile-types were worth 2 pence.

Because participants played only four opponents, we could not fully counterbalance reward probability, smile type and opponent sex. Therefore, participants saw opponents who were all of the same sex. Half the female participants played female opponents; likewise for male participants.

At the end of the task, participants ranked the opponents from 1 (most frequently rewarded) to 4 (least frequently rewarded), as an explicit measure of reward-contingency knowledge.

Results

We calculated the frequency with which participants chose each opponent in a given pair. When smile type was the same for both opponents in the test pair (e.g., the choice between the 80% and 70% politely smiling opponents), participants chose the opponent with the higher expected monetary value 58% of the time and significantly more often than predicted by chance, $t(34)=2.30$, $p=.03$, Cohen's $d=0.79$.

The degree to which genuine smiles carry intrinsic value should be evident in participants' preferences for genuinely over politely smiling opponents, when those opponents differ in expected monetary value. If genuine smiles have intrinsic value, then those smiles should increase an opponent's likelihood of being chosen. For example, an 80% genuinely smiling opponent should be chosen more often than an 80% politely smiling opponent. To test this, we calculated the proportion of times participants chose the 80% opponent when smile types were constant. We then subtracted this value from the proportion of times participants chose 80% opponents when smile types differed (Figure 2.6A). As anticipated, genuine smiles from the 80% opponent significantly increased choices of this opponent, relative to when smiles were matched, $t(34)=3.59$, $p=.001$; Cohen's $d=1.23$. Interestingly, when the 70% opponent displayed genuine smiles, participants showed a significant preference for that opponent, $t(34)=-2.51$, $p=.02$; Cohen's $d=-0.86$, suggesting that genuine smiles altered the desirability of the 70% opponent.

In Economics, the subjective desirability of a stimulus is called "utility". Here, the utility of both money and smiles can be modeled according to the degree to which they influence choice behaviour. This framework allows us to determine the utility of genuine smiles in comparison to the utility of money (for a formal description of expected utility maximization, see Von Neumann & Morgenstern, 1947). To do this, we applied a logistic model to the choice data using the logistic response function:

$$P_{OpponentA}=\exp(\theta)/(1+\exp(\theta))$$

where $P_{OpponentA}$ is the probability of choosing opponent A (the upper opponent in the test-phase choice pair, see Figure 1B) over opponent B, and θ is the difference in the opponents' utilities. We modeled θ as a linear function of opponents' monetary and social values, and the social- by monetary-value interaction.

$$\theta=\beta_1X_1 + \beta_2X_2 + \beta_3X_3$$

X_1 indicates the difference between opponents' expected monetary values (value of $Opponent_A$ minus value of $Opponent_B$; practically, this meant that if $Opponent_A$ had the larger expected monetary value on a given choice trial, X_1 was coded 1; if $Opponent_A$ was worse, X_1=-1; and if both opponents were equal in expected monetary value, X_1=0). X_2 is the difference in the opponent's social values (genuine or polite smile, coded in the same way as the monetary values, thereby placing the βs for money and smiles on the same scale). X_3 is the interaction term. The βs are the unstandardized logistic regression weightings for each variable–the degree to which money (β_1), smiles (β_2) and the money x smiles interaction (β_3) contribute to choice behaviour. The logistic regression was computed using an iteratively re-weighted least squares algorithm (O'Leary, 1990) to obtain the maximum likelihood estimate for each β. We estimated the βs for each participant independently, based on choices during the task. T-tests confirmed that both money, t(34)=2.32, p=.03, Cohen's d=0.80, and smiles significantly influenced participants' choices, t(34)=3.51, p=.001, Cohen's d=1.20. The interaction term was not significant, t(34)=0.61, p=.55, Cohen's d=0.21, suggesting that monetary and social rewards operated independently in this experimental design (Figure 2.6B). Surprisingly, the regression weight for smiles was 1.75 times higher than that for money, meaning that the difference

between genuinely and politely smiling opponents was 1.75 times more important in determining choices than a 10% difference in the probability of wining 2 pence. Based on this difference, we estimate that our participants would have chosen a genuinely smiling opponent with a reward probability of 62.5% (80%-17.5%) equally as often as a politely smiling opponent with an 80% reward probability. Put differently, a single genuine smile in this task had a utility equal to 0.35 pence.

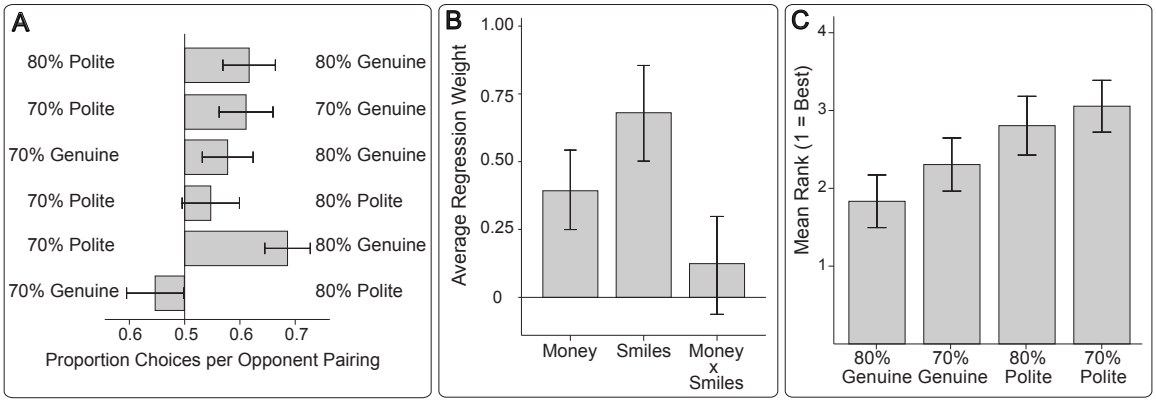

Figure 2.6. Experiment 4 results. A) Average proportion of choices of each opponent in a given pair. B) Average subjective contributions of opponents' monetary and social value to participants' choices. Subjective weightings are the estimated β weights from a logistic regression analysis. C) Mean explicit ranking of opponents' reward frequency depending on reward probability and smile type (1=best, 4=worst).

Figure 2.6C shows the average rank participants assigned to each opponent. A repeated-measures ANOVA showed a main effect of smile type, $F(3,102)=8.24$, $p<.001$; $\eta^2_p=.20$. Contrasts revealed that participants ranked genuinely smiling, 80% opponents highest, $F(1,34)=17.35$, $p<.001$; $\eta^2_p=.34$, and politely smiling 70% opponents lowest, $F(1,34)=15.20$, $p<.001$; $\eta^2_p=.31$. Consistent with the choice data, participants ranked 70% genuinely smiling opponents as better than 80% politely smiling opponents, although this difference was not statistically significant, $F(1,34)=2.03$, $p=.16$; $\eta^2_p=.06$. Taken together, these findings suggest that participants were willing to choose an opponent with lower expected monetary value, if that opponent offered a desirable social reward.

General Discussion

Our findings demonstrate that social stimuli, such as genuine smiles, carry intrinsic reward value, even when they are irrelevant to the task and not predictive of monetary

outcome. In Experiment 1, participants learned a reinforcement contingency from both social and non-social cues, and demonstrated preferences for social feedback. In Experiment 2, this preference for social feedback was confirmed. In Experiment 3, participants demonstrated a preference for opponents who provided genuine compared to polite smiles. In Experiment 4 genuine smiles altered opponents' utility such that genuinely smiling opponents were more desirable than expected value calculations would predict. Accordingly, participants ranked genuinely smiling opponents as having greater reward likelihood than politely smiling opponents.

Together, these results show that genuine smiles enhance stimulus utility. One way in which this might happen, is that smiles enhance the degree to which monetary rewards are incorporated (Averbeck & Duchaine, 2009). Alternatively, their utility may be added to that of other stimuli. Unfortunately, the present methods do not allow us to disentangle these explanations, nor do they allow us to determine whether a genuine smile's intrinsic value is innate or acquired.

As with all computer tasks, this task is only a proxy for the real social world. In order to show that learning mechanisms drive true social behaviour, these results must be replicated in the less constrained world of live social interaction. However, the fact that photographs of genuine smiles influenced behaviour, even in this artificial setting, suggests that they are likely to be powerful reinforcers in face-to-face interaction.

Learning to predict social outcomes from an interaction partner's cues is a beneficial skill (Behrens, et. al, 2009; Hampton, Bossaerts, & O'Doherty, 2008). Our findings demonstrate that genuine smiles have intrinsic reinforcement value. As such, these cues may guide social behaviour both on a moment-to-moment basis within an interaction and from one interaction to the next.

Chapter Three

'Tis better to look trustworthy than to be good: The effect of social cues on social utility

Danielle M. Shore and Erin A. Heerey

Abstract

An interaction partner's social cues provide important information about their potential behaviour and intentions. The different cues an interaction partner provides should combine to create an expectation about how likely that person is to provide social rewards. This expectation or social utility guides perceptions of and behaviour toward interaction partners. Over four studies, we investigate how appearance-based first impressions (static social cues), facial expressions (dynamic social cues) and interaction partner behaviour jointly affect social decision-making using an economic game. Participants played an investor-trustee game with social partners they had rated as trustworthy or untrustworthy. The partners played either fairly or unfairly with half playing consistently with participants' social expectations and half playing inconsistently. Partners also displayed expressions during the game (genuine and polite smiles). Results showed that social judgements and investment behaviour reflected opponent behaviour but differed depending on initial trustworthiness rating. This suggests that despite learning the behavioural value of a partner, initial impressions still tainted both social judgements and investment behaviour. Interestingly, displaying a smile at the time of feedback eliminated this bias. This suggests when stable cues (appearance and behaviour) conflict, dynamic cues (expressions) are powerful modulators of social utility. Indeed, by altering expectations about social partners, first impressions affect how partner behaviour is responded to. Together findings indicate that perceptions of social utility are based on the integration of multiple cues, however, these cues appear to interact, rather than combining in a linear fashion.

Trust is extremely important in the social world. Trustworthiness judgements inform social behaviour and are an essential aspect of successful social decision-making. Indeed, researchers have argued that the ability to assess trustworthiness accurately and use such judgements to guide behaviour is necessary for survival (Cosmides & Tooby, 2000). In accord, trustworthiness judgements influence a variety of social outcomes, including simple approach-avoid decisions (Adolphs, Tranel & Damasio, 1998), complex investment decisions (van t' Wout & Sanfey, 2008) and even voting decisions (Little, Roberts, Jones, & DeBruine, 2012).

People have many ways of signaling trustworthiness. The most readily available and most frequently studied of these cues is people's facial appearance. People make extremely rapid and highly reliable judgements of trustworthiness from photos of others' faces (Bar, Neta & Linz, 2006; Berry & Brownlow, 1989; Willis & Todorov, 2006). In similar vein, research shows that trustworthiness is one of two primary dimensions along which faces are categorized (Todorov, Said, Engell & Oosterhof, 2008), although trustworthiness is highly correlated with attractiveness (Wilson & Eckel, 2006). Despite the fact that appearance-based trustworthiness judgements inform social decision-making processes (Schlicht, Shimojo, Camerer, Battaglia & Nakayama 2010; van t' Wout & Sanfey, 2008), they are not always accurate (Olivola & Todorov, 2010; Porter, England, Juodis, van Brinke & Wilson 2008). It is therefore not surprising that the additional cues people send in real-world social encounters improve the accuracy of trustworthiness judgements and influence behaviour accordingly (Frank, Gilovich & Regan, 1993).

Although people readily judge trustworthiness from physical appearance, the best indicator of an interaction partner's trustworthiness is that person's previous behaviour (King-Casas et al., 2005). Research shows that information about how someone has behaved in the past influences subsequent decisions related to that individual (Delgado, Frank & Phelps, 2005; Singer, Kiebel, Winston, Dolan & Frith, 2004). Therefore, although first impressions of trustworthiness bias initial social decision-making, behavioural experience with a partner shifts those impressions to form an integrated percept about the degree to which a social partner can currently be trusted (Chang, Doll, van 't Wout,

Frank & Sanfey, 2010). Over time, therefore, an initial first impression of trustworthiness should gradually decline in the degree to which it influences behaviour towards an interaction partner.

Based on this reasoning, researchers have begun to apply models from reinforcement learning theory in order to explain how initial first impressions of trustworthiness change based on experience of a target's behaviour (e.g., Chang, et al., 2010). Such models propose that people learn 'social value' through the association of rewards (e.g., truthful behaviour that leads to financial gain) or punishments (e.g., untrustworthy behaviour that leads to financial loss) with social partners (Behrens, Hunt, Woolrich & Rushworth, 2008). This means that social partners may acquire 'value' as they become associated with good or bad social outcomes. This value reflects expectations about the utility or subjective desirability of interacting with a particular social partner based on one's history of experienced social outcomes with that partner (Chang et al., 2010). Thus, with no experience about actual behaviour, first impressions provide an initial expected utility estimate, which is updated through experience with a social partner.

Appearance-based stereotypes about faces should not be used once behavioural information becomes available, as they may not allow people to maximize possible gains from the social environment. Nonetheless, research shows that appearance-based judgements do shape decision-making, even when one has explicit knowledge about a partner's behaviour (Delgado et al., 2005; Resleszcu, Duchaine, Olivola, Chater, 2012). Thus, initial, appearance-based impressions of trustworthiness serve as a social 'prior' or belief, biasing expectations about and behaviour toward a partner long after an interaction history suggests the partner's true value. This appearance-based bias leads to irrational decision-making, as people fail to make the most advantageous decisions despite having the information available to do so.

In addition to static or "unfakeable" (Resleszcu et al., 2012) facial features and behaviour, people provide a third cue about their likelihood of behaving in a trustworthy fashion: facial expressions. Facial expressions both provide information about an interaction partner's intentions (Fridlund, 1991), and shape interpretations about the

behaviours people produce (Ames & Johar, 2009). Consistent with this idea research has shown that smiling faces are judged as more trustworthy than unsmiling faces, and that smiles influence decision-making in economic games (Scharlemann, Eckel, Kacelnik & Wilson, 2001; Winston, Strange, O'Doherty & Dolan, 2002). Research suggests that one mechanism by which smiles influence economic decisions is via the value that they add to social stimuli. Genuine smiles in particular carry reward value that can be measured in economic terms (Averbeck & Duchaine, 2009; Shore & Heerey, 2011). Smiles also make faces look more attractive (Otta, Folladore & Hoshino, 1996), which, given the high correlation between attractiveness and trustworthiness ratings, may serve to enhance perceptions of trustworthiness. Indeed, one weakness of much of the previous trustworthiness research is that participants view photographs of people in neutral poses (e.g., van t' Wout & Sanfey, 2008). In real life, however, faces are almost never 'neutral'. Smiles, especially those exchanged during one's first introduction to a new social partner, may therefore help to mitigate untrustworthy first impressions.

Here, we ask how first impressions judged from neutral faces interact with behavioural experience to create a joint percept of trustworthiness (Experiment 1), and how the presence of dynamic facial cues (in this case genuine and polite smiles) alters this effect (Experiments 2, 3, and 4). One excellent method of manipulating and measuring an interaction partner's trustworthiness is to use an iterated economic game in which participants invest multiple times with a series of partners. In this set of experiments, we use an 'investor-trustee' game in which participants play the role of the investor, investing some proportion of an endowment with an interaction partner on each trial. The invested amount is then multiplied by some factor (usually 3 or 4) and then the interaction partner (trustee) returns a portion of this matured investment to the participant (Berg, Dickhaut & McCabe, 1995). The return amount indicates a trustee's actual trustworthiness and the amount invested indicates how trustworthy the participant believes the trustee to be. In investor-trustee games, it is beneficial to invest heavily with trustworthy partners, as this will lead to financial gain. However, if a partner is untrustworthy, it is important to invest minimally, so as to minimize losses. Previous research shows that many factors affect

behaviour in investor-trustee games, such as facial trustworthiness (van t' Wout & Sanfey, 2008), previous behaviour (Chang et al., 2010; King-Casas et al., 2005), moral information (Delgado et al., 2005), and facial expressions (Scharlemann et al., 2001).

Experiment 1

Previous research suggests that people integrate information from their initial impressions and behavioural experiences with a partner to estimate that partner's social utility (Chang et al., 2010). This utility estimate, in turn, drives investment behaviour in an investor-trustee game. Experiment 1 examines the joint effects of appearance-based and behavioural experience-based information on social utility, under conditions in which these types of information agree and when they conflict. In particular, we examine how utility estimates shift when there is a shift in behaviour (e.g., a trustee who has played in a trustworthy fashion begins to play in an untrustworthy manner) and how appearance-based trustworthiness estimates influence people's ability to learn from behaviour. We predict that first impressions of trustworthiness will generate an initial expected value for interaction partners, which will initially influence participants' perceptions of interaction partners' social utility and thus their behaviour. Here, we test whether the influence of first impressions diminishes, as some research predicts (e.g., Chang et al., 2010), or whether it persists even after a partner's true value has been learned, as suggested by other work (e.g., Resleszcu et al., 2012).

Method

Participants

Forty undergraduate psychology students (11 male, mean age=20.02 SD=3.58) participated in the study for partial course credit and a small monetary bonus. All participants gave written informed consent and the University's Ethics Committee approved the study (likewise for Experiments 2, 3 and 4).

Procedure

Participants began the task by rating each of 14 grey-scale photos of actors' faces (either all male or all female faces; counterbalanced across participant sex) according to how happy, angry, attractive, friendly, outgoing and trustworthy they looked using a Likert scale ranging from 0 to 9 (0=not at all, 9=very much so). Actors wore neutral (no expression) poses, with eye-gaze directed toward the viewer and photos were presented in an elliptical window that closely cropped the face. Participants made all ratings for a face before rating the next face. The computer presented the faces in random order and the ratings occurred in random order for each face. Faces were presented centrally and remained visible until they had been rated on each of the traits. There was no time limit for responses.

Following the ratings, participants played a multi-round investor-trustee game with eight of the rated faces. Unbeknownst to participants, the computer selected faces based on their trustworthiness ratings, choosing the four highest- and four lowest-rated faces for each participant from the set of 14 faces. This procedure allowed us to create an individually calibrated set of high and low trustworthy faces for each participant.

Because we were interested in how investment strategies varied as a function of trustee trustworthiness, participants always played the role of the investor. On each trial, participants viewed a centrally presented photo of a trustee and received an endowment of 9 points. They decided how much of the endowment to invest (0 to 9 points) with the trustee using a key-press. Once they chose an investment amount, the investment matured (the computer tripled the invested amount) and the trustee made a return on the matured investment. Participants were instructed to pay attention to their returns and adopt an investment strategy that allowed them to earn as many points as possible, in order to maximize their bonus money at the end of the task. After each investment, participants received feedback showing the number of points the trustee had returned. There were 40 trials in each block (five per trustee) in random order. Over six blocks of trials, participants had the opportunity to invest with each trustee 30 times.

Returns in the game could be either fair (40% to 70% of tripled investment; this always resulted in a return amount equal to or greater than the original investment; Berg, et al., 1995) or unfair (0% to 30% of tripled investment; this always resulted in a return amount equal to or less than the original investment). The computer randomly selected the return amount on each trial from a uniform distribution. Two of the four trustees that had been rated as highly trustworthy (randomly assigned) returned investments fairly and two made unfair returns, likewise with the trustees that had been rated as untrustworthy. The feedback about these returns allowed participants to learn which trustees were actually trustworthy.

In order to examine how first impressions of trustworthiness interacted with learning, we included a manipulation in which some of the trustees switched their behaviour part way through the task. This meant that some trustees that had been playing fairly began to make unfair returns and vice versa for the unfairly playing trustees. Trustee behaviour was balanced across pre-rating condition such that among the high trustworthy trustees, there was one trustee (randomly assigned) that played fairly throughout the task, one that began fairly but switched to unfair play, one that began unfairly but switched to fair and one that played unfairly for the duration of the task. The same was true of the low-trustworthy pre-rated trustees. Switches occurred randomly between trials 12 and 16 (see table 3.1 for means and SDs of when switches occurred).

Table 3.1

Means and SDs of trial number when behavioural switch occurred by trustee type

Partner Type	Switch trial (mean)	Switch trial (SD)
High trustworthy fair switching to unfair	14.03	1.37
High trustworthy unfair switching to fair	13.80	1.59
Low trustworthy fair switching to unfair	14.45	1.85
Low trustworthy unfair switching to fair	13.93	1.59

In addition to the learning trials, participants completed two blocks in which they received no feedback about their investment returns. These blocks comprised the first and

last four investments with each trustee. These trials allowed us to obtain a baseline, pre-learning measure of investment behaviour with each trustee as well as a post-learning measure. The computer presented the trustees in random order in these trials.

Finally, participants re-rated the faces using the same method as in the pre-game rating. This allowed us to examine changes in ratings related to task behaviour. The task was programmed and presented in E-prime (Psychology Software Tools, Inc.).

Results

To ensure that the high- and low-rated faces did indeed differ on trustworthiness and to determine whether this difference was reflected in investment behaviour, we subjected the pre-game ratings and baseline (pre-learning) investments to repeated measures ANOVAs with trustworthiness rating (high or low) as the within-subjects factor. Results showed that participants viewed the high-rated faces as significantly more trustworthy than the low-rated faces, $F(1,39)=404.12$, $p<.001$, $\eta^2_p=.91$. Investment behaviour reflected participants' ratings. Specifically, participants invested more with faces rated as high on trustworthiness than with those rated as low in trustworthiness, $F(1,39)=169.49$, $p<.001$, $\eta^2_p=.81$.

To assess learning, the change in investment from pre-learning to the switch (pre-switch phase) and from the switch to post-learning (post-switch phase) were compared using a repeated measures ANOVA with trustworthiness rating (high or low), trustee behaviour (always fair; fair switching to unfair; unfair switching to fair; always unfair) and phase (pre-switch or post-switch) as within-subjects factors. Results show that participants learned trustee behaviour, increasing investments to fair trustees and decreasing investments to unfair trustees both before and after the switch, $F(3,117)=86.77$, $p<.001$, $\eta^2_p=.69$ (see Figure 3.1). Interestingly there was a significant interaction of trust rating and trustee behaviour, $F(3,117)=4.35$, $p=.006$, $\eta^2_p=.10$. Although participants learned trustee behaviour, the initial trust rating led to asymmetries in investments. Specifically, participants never invested as much with the low-rated fair playing trustees as they did with the high-rated fair playing trustees (see figure 3.1). This

finding suggests that although participants adapted their investment strategies to account for each trustee's current behaviour, their initial stereotypes about the trustees influenced investment strategies over and above learning.

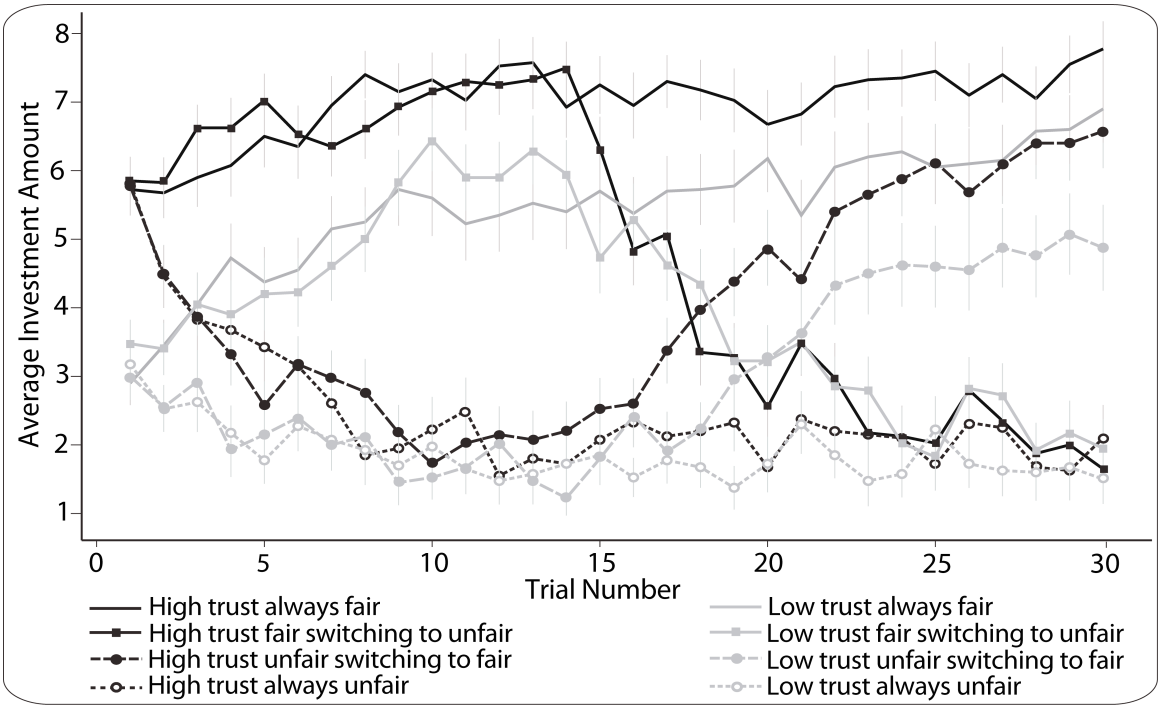

Figure 3.1. Experiment 1 investor-trustee game results. Participants' average investments on a trial-by-trial basis according to trustworthiness rating (high or low) and trustee behaviour (always fair; fair switching to unfair; unfair switching to fair; always unfair). Error bars indicate +/- 1 SEM.

To examine the degree to which current trustee behaviour (fair or unfair returns; +1, -1 coding), previous trustee behaviour (fair or unfair returns; +1, -1 coding), first impression of the trustee (high or low in trustworthiness; +1, -1 coding) and the interaction of these factors influenced participants' investment decisions, we modeled each participant's trial-by-trial investment behaviour with a linear regression model. One-sample t-tests against zero on participants' unstandardized regression weights showed that investment behaviour was driven by both first impressions, $t(39)=4.28, p<.001$, current trustee behaviour $t(39)=8.31, p<.001$ and previous trustee behaviour, $t(39)=6.44$, $p<.001$. Interestingly, there was also a significant interaction between behaviour and first impressions showing that participants treated a trustee's current behaviour differently depending on how that trustee had been rated prior to beginning the task and how they

had behaved in the past, t(39)=3.33,p=.002. Ultimately, this led to an asymmetry in investment behaviour such that participants invested less with fair but low-rated trustees than they did with fair but high-rated trustees. However, for unfair trustees, participants were unprejudiced by the initial trustworthiness ratings, as both high and low rated trustees were treated equally.

To assess how trustee behaviour changed investor behaviour we conducted a 2x4 repeated measures ANOVA on the post-learning investments with trustworthiness rating (high or low) and trustee behaviour (always fair; fair switching to unfair; unfair switching to fair; always unfair) as within-subjects factors. After learning, high rated trustees still received greater investments than did low rated trustees, $F(1,39)=13.21$, $p=.001$, $\eta^2_p=.25$. This occurred despite the fact that participants had learned each trustee's behaviour. Trustees that ended the game as fair received greater investments than those who ended the game as unfair, $F(3,117)=69.25$, $p<.001$, $\eta^2_p=.64$. As above, the interaction between initial impressions of trustworthiness and trustee behaviour was significant, $F(3,117)=3.95$, $p=.01$, $\eta^2_p=.09$, showing that participants' first impressions of trustworthiness biased investment such that participants never invested as much with low-rated trustees as they did with high-rated trustees, despite the behaviour of those trustees.

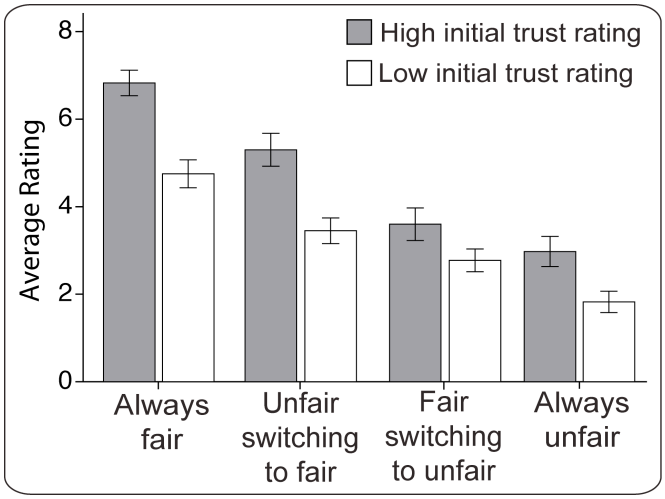

Figure 3.2. Experiment 1 rating results: Average post-game trustworthiness ratings according to trustworthiness rating (high or low) and trustee behaviour (always fair; fair switching to unfair; unfair switching to fair; always unfair). Error bars indicate +/- 1 SEM.

To assess how behaviour and initial impressions drove post game trustworthiness ratings, we analyzed post-game ratings using an ANCOVA model. We created dummy

coded variables for initial trustworthiness rating (1=high or 0=low) and behaviour (2=always fair, 1=fair switching to unfair or unfair switching to fair, 0=always unfair) and entered them into the model as fixed factors. Actual pre-game trustworthiness ratings were entered as a covariate. Figure 3.2 shows the average post-learning trustworthiness ratings according to trustee type. Unsurprisingly, results showed that post game ratings of trustworthiness depended on trustee behaviour $F(2,77)=61.94$, $p<.001$, $\eta^2_p=.62$. However, even after learning which players were fair and unfair, initial trustworthiness rating retained its influence, albeit at the trend level, $F(1,80)=3.59$, $p=.06$, $\eta^2_p=.04$. Thus, the original appearance-based trustworthiness judgement continued to influence ratings over and above the effects of actual trustworthiness experienced from behaviour.

Discussion

Together results from Experiment 1 show that when initial trustworthiness impression and experienced behaviour conflict, the ability to adapt behaviour depends on the nature of the conflict. Specifically, investments to fair-playing, high-rated trustees always surpassed those to fair-playing, low-rated trustees, despite clear evidence that participants learned how to invest with both types of trustees. The same was not true of the unfair playing trustees. When trustees played unfairly, investments did not differ depending on initial trustworthiness impression. Thus, a fair social partner's utility depends on both appearance-based and behavioural experience-based information, despite the fact that this leads to suboptimal behaviour with some trustees. These findings highlight that fact that first impressions bias social decision-making, leading to irrational choices.

Experiment 2

Experiment 1 results showed that when learning the social utility of an interaction partner, initial impressions of trustworthiness influenced investment behaviour. However, this effect was asymmetric: even though high-rated faces appeared to be just as bad as low-rated faces, the reverse was not true. Low rated faces, even when they offered

extremely good investment value, never seemed as trustworthy as high rated faces. This shows that 'unfakeable' facial characteristics or static cues have a strong effect on social utility, such that despite learning that a partner behaves fairly, investments are still biased by first impressions. Experiment 2 asks whether the addition of a common dynamic social cue, smiles, can modify the influence of the static cue on behaviour.

This experiment focused on two common and distinct smiles, genuine smiles of pleasure and polite smiles (for a review, see Ekman & Davidson, 1993; Ekman, Davidson & Friesen, 1990; Frank, Ekman & Friesen, 1993). Previous research suggests that smiles influence investment behaviour in trust games, such that smiling faces receive greater investments than neutral faces (Scharlemann et al., 2001). Smiling faces are viewed as more attractive (Lau, 1982) and more trustworthy than their non-smiling counterparts (Oosterhof & Todorov, 2009). One reason for this discrepancy is that smiles, particularly genuine smiles, have value that can be measured in economic terms, which means they may enhance the smiler's social utility (Shore & Heerey, 2011).

Here, participants interacted with smiling trustees, to determine whether the presence of these dynamic facial cues would overcome the bias against untrustworthy opponents. We predict that the addition of smiles will increase trustee utility, more so for untrustworthy than for trustworthy faces. Second, this effect will be larger for genuine compared to polite smiles. Based on these predictions, we anticipate that the asymmetry in investments to fair-playing, high- and low-rated trustees from Experiment 1 should disappear.

Method

Participants

Thirty undergraduate psychology students (17 male, mean age=20.47 SD=2.64) participated in the study for partial course credit and a small monetary bonus.

Stimuli

The neutral stimuli in this task were the same as those used in Experiment 1. However, Experiment 2 additionally included genuinely and politely smiling versions of

each face. Smile stimuli were created using the same methods as in Shore & Heerey (2011). All stimuli were presented in an elliptical window that closely cropped the face.

Procedure

Participants began the task by completing the same rating procedure as in Experiment 1, rating 14 faces with neutral expressions (all male or all female; counterbalanced across participant genders) on 6 attributes, the important one of which was trustworthiness.

As in Experiment 1, participants then played an investment game with eight of the faces, selected on the basis of their ratings (4 highest- and 4 lowest-rated on trustworthiness). Participants played four blocks of the investor-trustee game in which they made investments and received feedback about their returns. Trials of the investment task ran as in Experiment 1, with 40 trials in each learning block (5 per face; 30 learning trials per trustee in total). In this version of the task, the trustees behaved consistently and did not switch their behaviour at any point during the task. Two high-rated trustees returned fairly and two returned unfairly; likewise for the low-rated trustees. The fair trustees returned 40% to 70% of participants' investments whereas the unfair trustees returned 0% to 30% of investments.

Other than the fact that trustees behaved consistently, the key difference between this task and that in Experiment 1, was the inclusion of smiles. To enable us to assess how genuine and polite smiles modulate the interaction between first impressions and learned behaviour, the trustees displayed smiles throughout the investment rounds, rather than neutral poses. At the start of the task, the computer randomly assigned each face to display either genuine or polite smiles. Half the trustees displayed genuine smiles and the remainder displayed polite. Smiles were distributed across trustee conditions such that there was one fair-playing, high-rated trustee who smiled genuinely and one fair-playing high-rated trustee who smiled politely. There was also one genuinely smiling and one politely smiling high-rated trustee who played unfairly. The same distribution of smiles occurred across the low-rated trustees.

As in Experiment 1, we obtained pre- and post-learning measures of investment behaviour by asking participants to make investments to each trustee without receiving feedback. Participants made four investments per trustee in random order. To determine the degree to which smiles added to each trustee's social utility, participants made the first two of these investments to the trustee's neutrally posed face and the remaining two investments to the trustee's smiling face. The smiling faces were the same ones participants saw in the subsequent learning blocks. Participants completed the post-learning investments in similar fashion. However, they made the first two of the four investments with each trustee to the trustee's smiling face and the remainder to the trustee's neutral face. Participants finished the task by re-rating all 14 faces (in neutral poses).

Results

A repeated-measures ANOVA with trustworthiness rating (high or low) as the within-subjects factor confirmed that high- and low-rated faces differed on trustworthiness. High trust faces were rated as more trustworthy than the low-rated faces, $F(1,29)=295.34$, $p<.001$, $\eta^2_p=.91$. To assess the degree to which ratings and smiles altered investment behaviour, we subjected the pre-learning investments to a 2x2x2 repeated measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and face display (neutral or smile) as within-subjects factors. As anticipated, high rated trustees received greater investments than did low-rated trustees (see Table 3.2 for exact statistics). In addition, investments increased significantly when trustees displayed smiles compared to neutral expressions, although there was no main effect of smile type. There was a significant interaction between trustworthiness rating and face display, such that investments to low rated trustees increased more when they smiled than did investments to high rated trustees. As predicted, however, the three-way interaction was not significant. These results show that any type of smile increases the utility of a face when only appearance-based information is available, however, this effect is asymmetric such that low-trustworthy faces benefit more from this change than do high-trustworthy faces.

These results hint that, the addition of smiles to low-rated faces may serve to overcome the asymmetry in investment behaviour observed in Experiment 1.

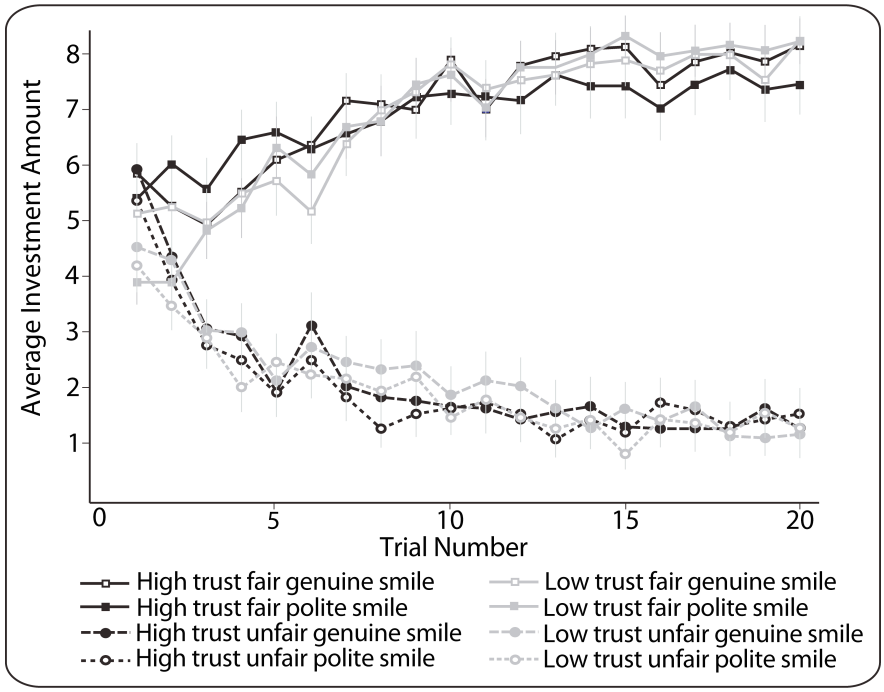

Figure 3.3. Experiment 2 investor-trustee game results. Participants' average investments on a trial-by-trial basis according to trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

To assess learning over time, the differences between pre- and post-learning investments (neutral face investments from no-feedback blocks) were compared using a repeated measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and trustee behaviour (fair or unfair) as within subject factors. As reported in Table 3.2 and shown in Figure 3.3, participants developed investment strategies that reflected trustee behaviour, increasing investments to fair trustees and decreasing investments to unfair trustees. There was also a main effect of trustworthiness rating. However, in contrast to Experiment 1 there was no interaction between trustee behaviour and trustworthiness rating. Interestingly, this shows that inclusion of smile displays eliminated the bias against low-rated, fair players. There was a significant 3-way interaction, suggesting that initial trustworthiness rating and smile type interacted to jointly influence behaviour. Specifically, for unfair trustees smiles only affected investments to high-rated partners, such that investments decreased more for genuinely

smiling compared to politely smiling trustees. For low-rated unfair trustees smiles did not affect investments. Conversely, for fair trustees smile type influenced investments to both low-rated and high-rated trustees. Specifically, for high-rated trustees genuine smiles led to a greater increase in investments than polite smiles as expected, however the opposite was true for low rated faces such that polite smiles led to greater increases in investments. This suggests that smiles have differential effects on social utility depending on behavioural and appearance-based trustworthiness.

A 2x2x2 repeated measure ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and trustee behaviour (fair or unfair) as within subject factors confirmed that initial trustworthiness rating did not bias investments, $F(1,29)=0.35$, $p=.56$, $\eta^2_p=.01$. Results also confirmed that participants' responses to the types of smiles they saw differed according to the interaction of first impression and behaviour $F(1,29)=5.47$, $p=.03$, $\eta^2_p=.16$.

To examine the degree to which trustee behaviour (fair or unfair returns; +1, -1 coding), first impression of the trustee (high or low in trustworthiness; +1, -1 coding), and smile type (genuine or polite; +1, -1 coding), and the interaction of these factors influenced participants' investment decisions, we modeled each participant's trial-by-trial investment behaviour with a linear regression model. One-sample t-tests against zero on participants' unstandardized regression weights showed that investment behaviour was driven by trustee behaviour alone, $t(29)=10.71$,$p<.001$. Neither first impressions, $t(29)=0.36$, $p=.72$, nor smile type, $t(29)=1.08$, $p=.29$, significantly affected investment behaviour. Importantly, there were no significant interactions between the terms showing that trustee ratings prior to the task did not change how participants treated trustee's behaviour (all p-values>.57). Essentially, the addition of smiles to faces led to equal investment behaviour regardless of initial trustworthiness ratings, such that unlike Experiment 1, participants invested the same with fair but low-rated trustees and fair but high-rated trustees. Likewise, participants' investments to all unfair trustees were similar.

Table 3.2

*Experiment 2 ANOVA results. Note that all ANOVAs are 2x2x2 repeated measures ANOVAs. Bold font and * indicates a significant result.*

Dependent Variable	Factor	df *numerator*	df *denominator*	F	p-value	Effect Size (η^2)
Pre-learning investments	Trust rating (high or low)	1	29	59.32	**<.001***	.67
	Smile type (genuine or polite)	1	29	1.64	.21	.05
	Face display (neutral or smile)	1	29	22.46	**<.001***	.44
	Trust rating x smile type	1	29	0.78	.39	.03
	Trust rating x face display	1	29	7.43	**.01***	.20
	Smile type x face display	1	29	0.80	.38	.03
	Trust rating x smile type x face display	1	29	0.005	.94	<.001
Learning difference pre- to post-investments	Trust rating (high or low)	1	29	23.95	**<.001***	.45
	Smile type (genuine or polite)	1	29	0.20	.66	.01
	Trustee behaviour (fair or unfair)	1	29	70.86	**<.001***	.71
	Trust rating x smile type	1	29	2.46	.13	.08
	Trust rating x trustee behaviour	1	29	0.41	.53	.01
	Smile type x trustee behaviour	1	29	0.03	.87	.001
	Trust rating x smile type x trustee behaviour	1	29	7.47	**.01***	.21

Figure 3.4 shows the post game trustworthiness ratings according to partner type. Post game trustworthiness ratings were analyzed with an ANCOVA model as in Experiment 1. Initial trustworthiness ratings were entered as a covariate and dummy coded variables were created for initial trustworthiness rating (1=high or 0=low) play

behaviour (1=fair, 0=unfair) and smile type (1=genuine, 0=polite) and entered into the model as fixed factors. Trustworthiness ratings only reflected experienced behaviour, $F(1, 29)=20.89$, $p<.001$, $\eta^2_p=.42$, initial trust rating ($p=.34$) and smile type ($p=.47$) did not influence ratings. This shows that adding smiles removed the influence of initial trustworthiness on trait judgements, such that unlike Experiment 1 judgements only reflected experienced behaviour and the effects of initial judgements did not persist.

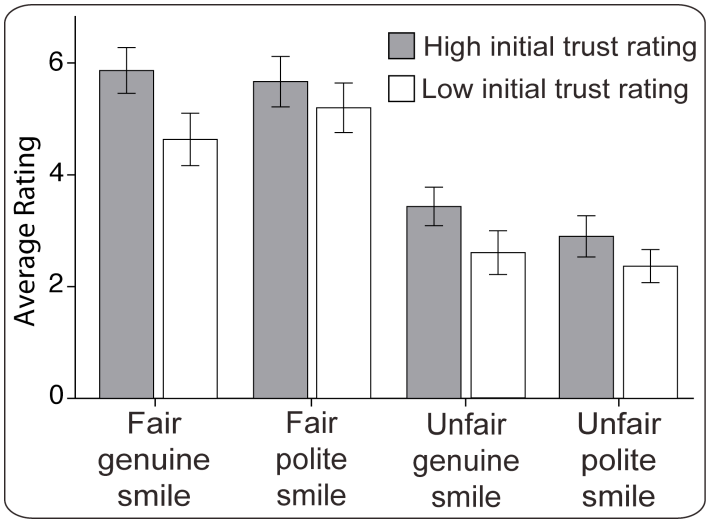

Figure 3.4. Experiment 2 rating results: Average post-game trustworthiness ratings according to initial trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Discussion

Smiles significantly modified investment behaviour such that the asymmetry observed in Experiment 1 disappeared. Specifically, when low-rated trustees smiled and played fairly, participants were willing to invest as much with these trustees as they did with fair-playing, high-rated trustees. This shows that the addition of a smile led participants to treat all fair-playing trustees equally, regardless of initial impressions of trust. This result suggests that smiles increased the utility of the fair-playing but low-rated trustees such that they became as valuable investment partners as the fair, high-rated trustees. Ratings data support this finding as they show that post-game ratings reflected only trustee behaviour and no longer depended on initial impressions of trustworthiness as in Experiment 1. Interestingly, participants' responses to smile type differed across trustees depending on both first impression and experienced behaviour. This suggests

that the value of a smile is not constant but may depend on factors in the social context, such as who provides the smile and the circumstances under which they do so. Overall, this shows that smiles alter the utility of an interaction partner, but that the specific effects of smiles depend both on the static features of facial appearance and on their behavioural context.

Experiment 3

Experiment 2 showed that the addition of dynamic social cues (genuine and polite smiles) changed trustees' social utility enabling rational decision-making. However, in Experiment 2, trustees displayed smiles throughout learning blocks. In live social interactions, expressions are transient, changing from moment-to-moment, to provide insight about a social partner's motivations or intentions (Blair, 2003; Fridlund, 1991). Therefore, Experiments 3 and 4 investigate whether the timing of a smile display is important. Previous research suggests that smiles add to the value of feedback, such that the value of the smile generalizes to the value of a financial gain increasing the perceived value of the reward (Shore & Heerey, 2011). To examine whether the effect of a smile stems from its integration with financial gains, participants in Experiment 3 only saw trustees smile when they provided feedback about participants' investment returns. Therefore, Experiment 3 tests whether the utility of a smile at feedback generalizes to subsequent investments.

Method

Participants

Thirty undergraduate psychology students (10 male, mean age=19.80, SD=2.11) participated in the study for course credit and a small monetary bonus.

Procedure

The study used exactly the same procedure as Experiment 2, with one difference in the investor-trustee game. In this version, participants viewed neutrally posed trustees when making investments and the trustees only displayed smiles when providing

feedback about investment returns. All other aspects of the procedure were identical to those described above.

Results

To verify that high- and low- rated faces differed on trustworthiness and initial investments reflected this we conducted two repeated measures ANOVAs with trustworthiness rating (high or low) as a within subject factor. Results confirmed that high-rated faces were rated as significantly more trustworthy than low-rated faces, $F(1,29)=263.81$, $p<.001$, $\eta^2_p=.90$, and received greater investments than those with low trustworthiness ratings, $F(1,29)=84.10$, $p<.001$, $\eta^2_p=.74$.

To confirm that smiles had a significant effect on investment behaviour, baseline investments were subjected to repeated measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and face display (neutral or smile) as within subject factors. Results showed that participants invested significantly more when with high rated trustees, and when a smile was displayed (see Table 3.3 for exact statistics). Moreover, low-trustworthy trustees tended to receive greater investments when they were smiling relative to neutral, whereas high trustworthy trustees did not, as indicated by the interaction of face display and trust rating. There was also a significant interaction of smile type and display, such that genuine smiles led to a bigger increase in investments than polite smiles when faces changed from neutral displays to smiles. This confirms the results of the Experiment 2 baseline investments showing that displaying a smile has a marked effect on investment behaviour.

To examine learning, the pre- and post-learning investments (from no-feedback blocks) were compared using a repeated-measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and trustee behaviour (fair or unfair) as within subject factors. Participants learned trustee behaviour changing their investment strategies accordingly, increasing investments to fair faces and decreasing investments to unfair faces. Importantly, as in Experiment 2, investment behaviour did not differ depending on initial trustworthiness rating when trustees behaved in the same manner.

Results showed no influence of smile type, suggesting that in this version smiles did not influence investment behaviour. Interestingly, unlike Experiment 2, the three-way interaction was not significant, suggesting that when smiles are only present during feedback their effects do not differ according to trustee behaviour and appearance-based judgements (see Figure 3.5).

Table 3.3

Experiment 3 ANOVA results. Note that all ANOVAs are 2x2x2 repeated measures ANOVAs. Bold font indicates a significant result.

ANOVA	Factor	df *numerator*	df *denominator*	F	p-value	Effect Size (η^2)
Pre-learning investments	Trust rating (high or low)	1	29	62.77	**<.001***	.68
	Smile type (genuine or polite)	1	29	17.14	**<.001***	.37
	Face display (neutral or smile)	1	29	17.14	**<.001***	.37
	Trust rating x smile type	1	29	3.53	.07*	.11
	Trust rating x face display	1	29	3.53	.07*	.11
	Smile type x face display	1	29	5.81	**.02***	.17
	Trust rating x smile type x face display	1	29	0.60	.45	.02
Learning difference pre- to post-investments	Trust rating (high or low)	1	29	37.50	**<.001***	.56
	Smile type (genuine or polite)	1	29	0.05	.82	.002
	Trustee behaviour (fair or unfair)	1	29	186.70	**<.001***	.87
	Trust rating x smile type	1	29	<.001	1	<.001
	Trust rating x trustee behaviour	1	29	0.01	.94	<.001
	Smile type x trustee behaviour	1	29	0.04	.84	.001
	Trust rating x smile type x trustee behaviour	1	29	0.60	.45	.02

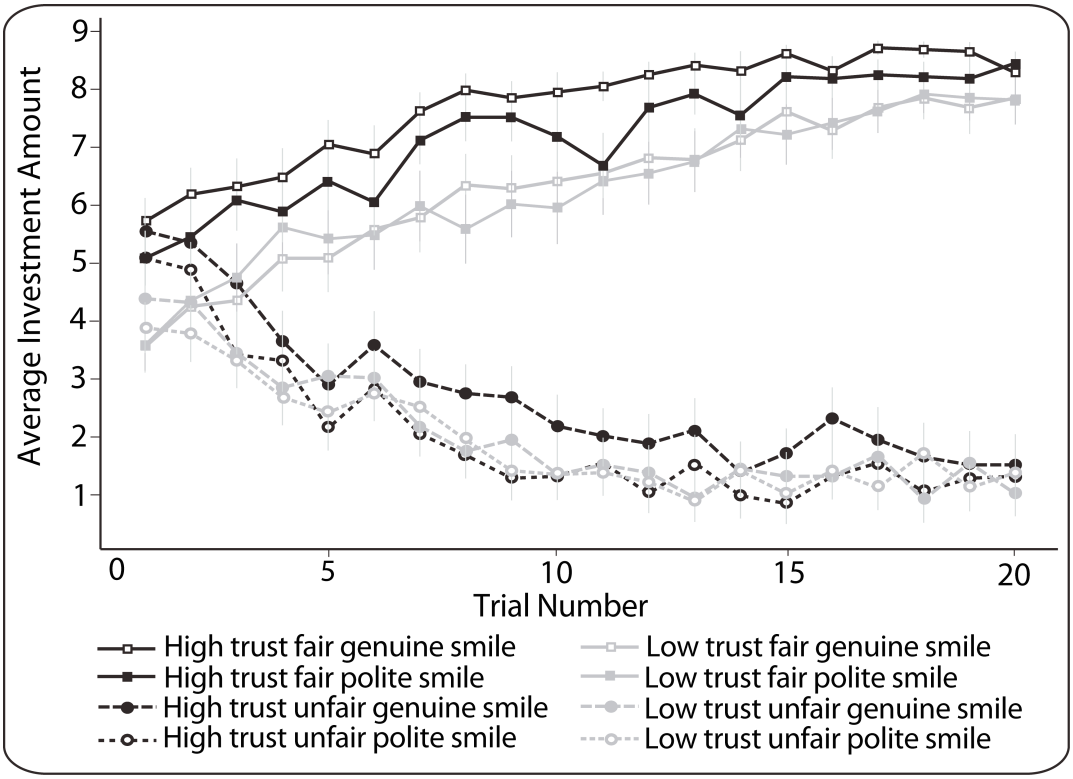

Figure 3.5. Experiment 3 investor-trustee game results. Participants' average investments on a trial-by-trial basis according to trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Post-game trust ratings were analyzed using an ANCOVA model with initial trustworthiness rating as a covariate. Dummy coded variables were created for trustee behaviour (1=fair, 0=unfair), initial trust rating (1=high, 0=low) and smile type (1=genuine 0=polite) and entered into the model as fixed factors. Trustworthiness ratings depended on trustee behaviour, $F(1,29)=35.10$, $p<.001$, $\eta^2_p=.55$, and the behaviour x smile type interaction, $F(1,29)=8.37$, $p=.007$, $\eta^2_p=.23$, such that all unfair partners received lower ratings than fair partners and polite smiles resulted in slightly higher ratings for fair playing faces than genuine smiles. Together these results show that trustee behaviour most strongly influenced trait ratings, but that smile type and behaviour interacted to influence ratings. This confirms that seeing smiles with feedback removes the first impression bias shown in Experiment 1 with smiles enhancing post-game ratings, even though the faces were neural at the time of the post-game trait ratings.

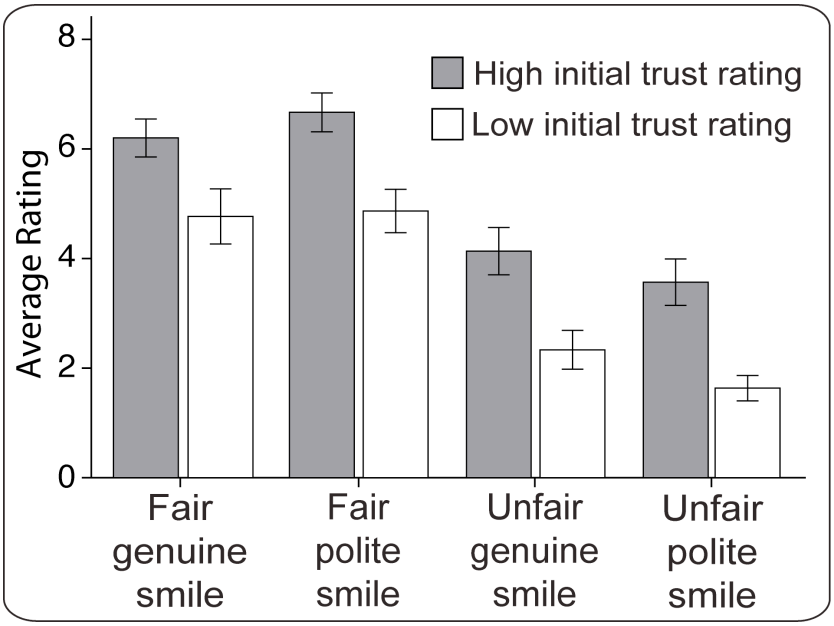

Figure 3.6. Experiment 3 rating results: Average post-game trustworthiness ratings according to initial trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Discussion

These results show that displaying a smile when providing investment return feedback negates the initial-impression investment-behaviour bias. This shows that displaying a smile during feedback increased the utility of the fair-playing but low-rated trustees leading participants to invest with all fair-playing trustees equally, despite appearance-based impressions of trust. This suggests that the influence of smiles on utility derives from its integration with financial gains. This supports the idea that smile value generalizes to feedback, thus increasing the perceived value of the feedback (Shore & Heerey, 2011). The increased feedback value subsequently increased the trustee utility. Rating behaviour confirmed investor-trustee results, such that first impressions of trustworthiness did not bias post-game ratings of trust. Unlike Experiment 2 the type of smile did not have different affects according to first impressions and trustee behaviour, suggesting that when present at feedback the added value of a smile does not differ according to trustee type.

Experiment 4

In Experiments 2 and 3 smiles increased social utility of interaction partners, eliminating the appearance-based bias found in Experiment 1. To test whether the effect of a smile display is contingent on timing, Experiment 4 examines the effects of smiles present at the time of investment. Therefore, the trustees in this experiment only displayed smiles at the time of investment and displayed neutral poses when participants received feedback about their decisions. Based on research showing that smiles are a good indicator of people's future intentions, such as the intent to reciprocate (Reed, Zeglen & Schmidt, 2011; Scharlemann et al., 2001), we test whether seeing a smile during investment increases social utility and eliminates the investment bias against low-rated faces. If the findings of Experiments 2 and 3 are due to smiles creating a general increase in partner utility then Experiment 4 should find smiles equal the value of trustees, thus removing initial appearance-based biases. However, it may be that the effects of smiles are due to the generalization of their value to the value of feedback. If this is the case then although smiles displayed at investment should increase social utility, it may not be enough to overcome initial appearance-based biases, as the value of feedback is not increased.

Method

Participants

Twenty-nine undergraduate psychology students (4 male, mean age=20.93 SD=3.45) participated in the study for course credit and a small monetary bonus.

Procedure

The study used exactly the same procedure as in Experiment 2, with one difference in the investor-trustee game. In this version, trustees only displayed smiles when participants chose how many points they wished to invest. When they provided feedback about participants' returns, the trustees wore neutral poses. Everything else (e.g., rating procedures, pre- and post-learning investments) remained the same.

Results

As in Experiments 1, 2, and 3, trustworthiness ratings and initial investments for low- and high-rated faces were significantly different (p-values<.001). To measure the effect of a smile on investment behaviour we conducted a 2x2x2 repeated measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and face display (neutral or smile) as within subject factors. Results replicated the Experiment 2 and 3 findings showing that participants invested more with high-rated than with low-rated trustees (see Table 3.4 for exact statistics). Similarly, adding a smile to a previously neutral face increased investments. There was no difference in investments to genuine or polite smiles. However, there was a trend-level interaction between trustworthiness rating and display, showing that although the type of smile did not influence investments, low rated trustees saw bigger investment increases when presenting a smile compare to high rated trustees (see Figure 3.7).

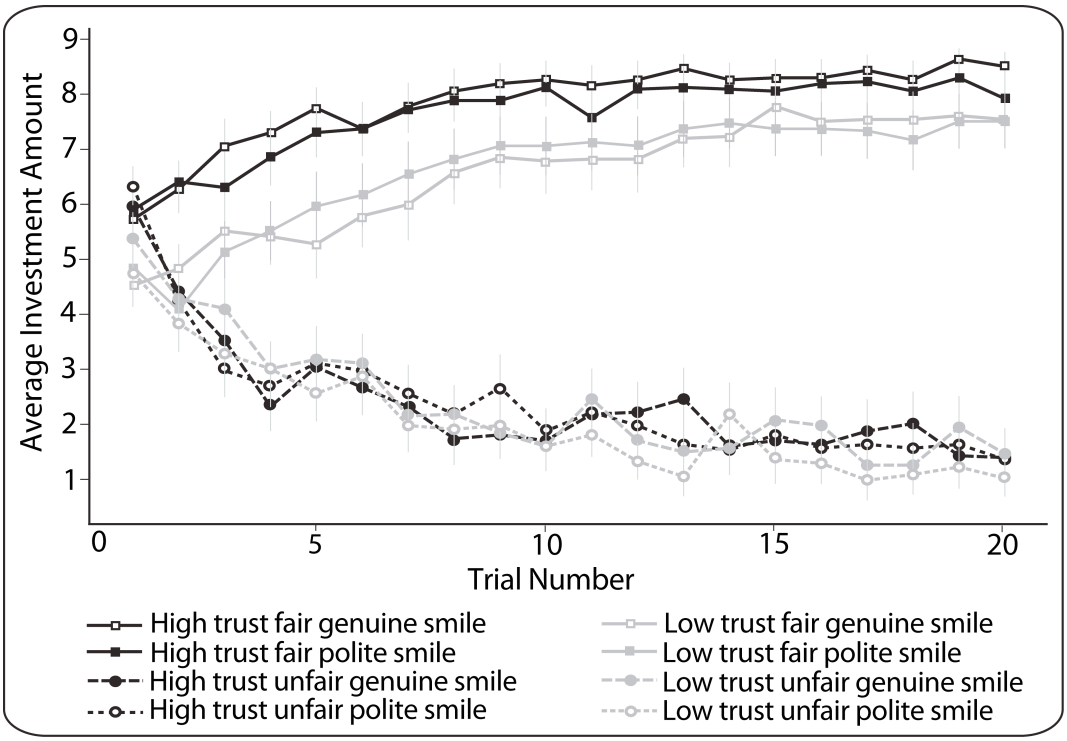

Figure 3.7. Experiment 4 investor-trustee game results. Participants' average investments on a trial-by-trial basis according to trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Table 3.4

Experiment 4 ANOVA results. Note that all ANOVAs are 2x2x2 repeated measures ANOVAs. Bold font indicates a significant result.

ANOVA	Factor	df *numerator*	df *denominator*	F	p-value	Effect Size (η^2)
Pre-learning investments	Trust rating (high or low)	1	28	42.64	**<.001***	.60
	Smile type (genuine or polite)	1	28	1.70	.20	.06
	Face display (neutral or smile)	1	28	28.19	**<.001***	.50
	Trust rating x smile type	1	28	0.01	.94	<.001
	Trust rating x face display	1	28	3.30	.08*	.11
	Smile type x face display	1	28	7.52	**.01***	.21
	Trust rating x smile type x face display	1	28	0.009	.92	<.001
Learning difference pre- to post-investments	Trust rating (high or low)	1	28	12.03	**.002***	.30
	Smile type (genuine or polite)	1	28	3.89	.06*	.12
	Trustee behaviour (fair or unfair)	1	28	151.83	**<.001***	.84
	Trust rating x smile type	1	28	0.068	.80	.002
	Trust rating x trustee behaviour	1	28	5.71	**.02***	.17
	Smile type x trustee behaviour	1	28	0.05	.83	.002
	Trust rating x smile type x trustee behaviour	1	28	1.26	.27	.04

To confirm that participants learned trustee behaviour, pre- and post-learning investments (no feedback blocks) were analyzed using a repeated-measures ANOVA with trustworthiness rating (high or low), smile type (genuine or polite) and trustee behaviour (fair or unfair) as within subject factors (Table 3.4). Participants learned trustee behaviour and adapted investment strategies accordingly. Interestingly, in contrast to Experiments 2

and 3, trustworthiness rating did affect investment behaviour, with participants investing more with high rated fair playing trustees than with low rated fair playing trustees. Thus, the asymmetry in investment behaviour suggests that participants allowed their initial prejudices to influence investment behaviour even after they learned the true value of the faces. Together this suggests that displaying smiles only at the time of investment does not overcome the initial trust rating bias[1].

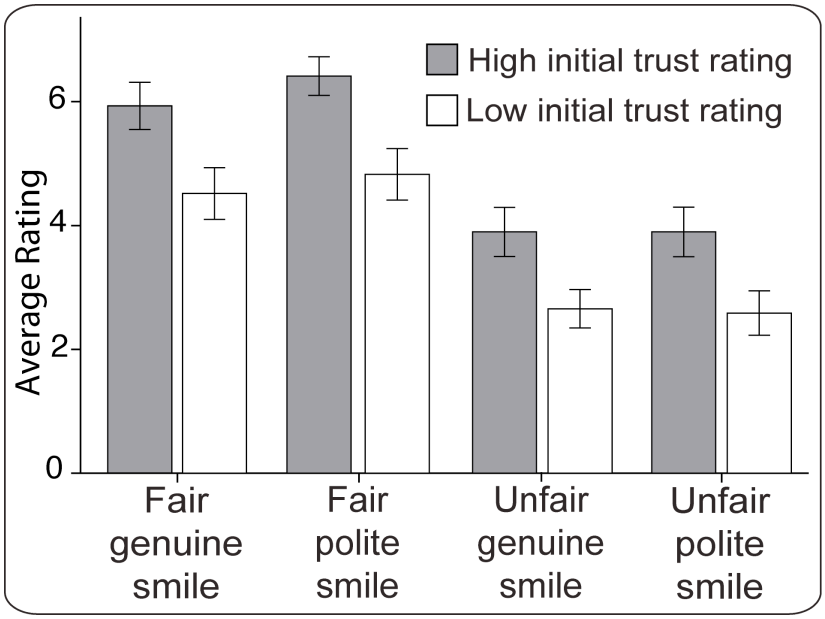

Figure 3.8. Experiment 4 rating results: Average post-game trustworthiness ratings according to initial trustworthiness rating (high or low), trustee behaviour (fair or unfair), and smile type (genuine or polite). Error bars indicate +/- 1 SEM.

Figure 3.8 shows post-game trust ratings according to partner type. We entered post-game trust ratings into an ANCOVA model with dummy coded variables for initial trustworthiness rating (1=high, 0=low), trustee behaviour (1=fair, 0=unfair), and smile

[1] To confirm that smile effects were contingent on timing we compared investment behavior towards fair plating trustees in experiment 3 (smiles at feedback only) and experiment 4 (smiles at time of investment only) using a repeated measures ANOVA. Trustworthiness rating (high or low) and smile type (genuine or polite) were within subject factors and smile timing (at investment or at feedback) was a between subject factor. Although there was no main effect of smile timing $F(1,57)=0.08$, $p=.77$, $\eta^2_p=.001$, there was a significant trustworthiness rating by smile timing interaction $F(1,57)=5.57$, $p=.02$, $\eta^2_p=.09$, such that differences between investments to high and low trust faces were significantly greater when the smile was presented during investment (experiment 4) compared to when the smile was present during feedback (experiment 3). This supports the conclusion that the affect of smiles depends on timing, such that smiles present during feedback eliminate appearance-based biases in investment behavior, but smiles presented at the time of investment do not.

type (1=genuine, 0=polite), as fixed factors. Actual pre-game trustworthiness ratings were entered as a covariate. Trustworthiness ratings were driven by trustee behaviour alone $F(1,28)=31.72$, $p<.001$, $\eta^2_p=.53$. This suggests that although not enough to change the bias in investment, smiles did reduce the influence of appearance on trust ratings.

Discussion

Findings show that a smile at the time of investment alone does not overcome the investment bias against trustees with low appearance-based first impressions found in Experiment 1, as does the presence of a smile throughout the trial or at the time of feedback alone. This suggests that the value of a smile lies in its integration with the value of feedback. Thus, although social utility value increased from the addition of a smile at investment it was not sufficient to eliminate the bias against low-rated trustees. The high investment values in this study (compared with investments made to neutral trustees), suggest increased trustee utility at the time of investment. However, when trustees returned to neutral displays during feedback, the perceived utility of the outcome decreased, such that without the added value of a smile, first-impressions of trustees biased participant behaviour. As the addition of a smile had a greater effect for low-rated trustees, the removal of a smile likely had a greater effect for low-rated trustees, meaning that any utility change is not sufficient to overcome the investment asymmetry we discovered in Experiment 1. This suggests that the timing of a smile display is important. Specifically, it appears that the generalization of smile value to the value of a financial gain is key to increasing the perceived social utility of an interaction partner. Smiles nonetheless appear to influence pre- and post-game investment behaviour and contribute to trait judgements.

General discussion

Our findings demonstrate that an interaction partner's social utility depends on a combination of the social cues they provide. When static appearance-based cues conflict with behavioural cues, initial impressions of trustworthiness are more important and have

a greater influence on investment behaviour than when those cues are consistent. This affect is asymmetrical, such that participants do not invest as much with low-rated partners who play fairly in comparison to their high-rated counterparts (Experiment 1). The addition of a dynamic appearance-based social cue, smiles, overcame this bias (Experiments 2 and 3), but only when the cue was present during the feedback phase, and not when the cue was only present at the time of investment (Experiment 4). Together these results suggest that when appearance and behaviour conflict, dynamic appearance-based cues (expressions) can modulate the effects of static or 'unfakeable' ones.

These findings support the idea that initial appearance-based impressions are akin to assigning a utility value to an interaction partner, which reinforcement learning processes update as new information becomes available (Chang, et al., 2010). Furthermore, these findings show that initial impressions of trustworthiness have an asymmetric effect on interaction partner utility, such that despite learning from behaviour, participants do not maximize their gains from low-rated partners. Changing the appearance of the face by adding a dynamic social cue, a smile, enables participants to maximize their gains from all partners, regardless of first impressions.

One explanation for these findings is that participants attend more to behaviours that confirm initial impressions (e.g., an unfair return from a trustee rated low in trustworthiness) and pay less attention to those that conflict (Delgado et al., 2005). According to this explanation, high-rated trustees that played unfairly should have received larger investments than low-rated unfair trustees. However, the results do not bear this suggestion out. The bias in investment behaviour discovered in this set of studies was asymmetrical, only fair playing trustees were treated differently depending on appearance, with low-rated trustees receiving smaller investments than high-rated trustees. The fact that this asymmetry was even present when the trustees shifted their behaviour suggests the strength of the effect.

The presence of a smile removed the bias in Experiment 2 and 3, when trustees displayed smiles throughout interaction or during feedback. This suggests that the addition of the smile to an untrustworthy looking trustee equalized social utility such that

low-rated trustees received the same investments as their high rated counterparts. However, only seeing smiles during investment (Experiment 4) did not eliminate the bias against low-rated faces. It may be that smiles affected social utility in this experiment, evidenced by the increased investments to smiling faces, but that when the smiles are not present faces decrease in utility. As low-rated faces benefit more from the addition of smiles, it is logical that they also see a greater reduction in utility when smiles are not present, hence the revival of the first impression bias in Experiment 4.

One explanation for the effect of smiles on investment is that the smiles changed the configuration of 'unfakeable' facial features and thus eliminated their influence on behaviour. Research suggests that appearance-based judgements may reflect the overgeneralization of features that look like emotional expressions (Adams, Nelson, Soto, Hess & Kleck, 2012; Oosterhof & Todorov, 2009; Zebrowits & Montepare, 2008). Therefore, displaying smiles all the time changed the appearance of the faces and thus the influence of appearance-based information. When displayed during feedback, smiles may have increased the value of the feedback, thus increasing social utility of trustees (Shore & Heerey, 2011). However, when smiles were only displayed at investment the appearance-based information did not change and feedback had no additional value, therefore initial-impressions retained their influence on behaviour.

Together results highlight the ways in which first impressions bias social utility, leading to irrational social decisions and behaviour. Experiments 2 and 3 showed that only with the addition of positively valued cues could negative first impressions be overcome. Indeed, many participants in Experiments 1 and 4 verbally reported their irrational decision strategies, during debriefing. Specifically, they mentioned that they knew which trustees returned fairly and unfairly but said they were unwilling to invest the highest amounts with faces that looked untrustworthy, even when they played fairly. This illustrates the importance of appearance-based expectations on both social decision-making and behaviour. This has implications for future social decision-making models, as they must account for the irrational influence of first impressions on behavioural choices, especially when there is conflict between appearance and behaviour.

The expected difference between genuine and polite smiles was only partially confirmed. In Experiment 2, genuine smiles led to higher investments but only for high-rated fair faces. Interestingly, for low-rated fair playing faces the opposite occurred, with politely smiling faces receiving higher investments than genuinely smiling ones. This suggests that the interpretation of a smile may differ, depending on the characteristics of the person providing the smile and the circumstances under which they do so. This suggests that the value of a smile is not fixed. That is, genuine smiles may not always be rewarding. Rather, the degree to which they are likely depends on a number of contextual factors. First impressions, therefore, have powerful effects on how both behavioural and expressive cues are valued as well as their contribution to an interaction partner's social utility.

Conclusions

The utility of an interaction partner shapes decision-making towards that social partner. This research shows that appearance, behaviour and facial expression cues all contribute to social utility. When cues conflict, initial impressions of trustworthiness have a strong influence on behaviour towards interaction partners despite accurate value learning. However, the presence of dynamic social cues, which change the configurations of facial features and the value of feedback, can prevent this irrational bias. This work illustrates the importance of understanding how different types of social cues are integrated to create social value. In real-world terms, first impressions really matter. Indeed appearance-based expectations have serious long-term consequences on the realization of potential, from academic achievement (Ritts, Patterson & Tubbs, 1992), to financial and occupational success as an adult (Baron, Markman & Bollinger, 2006; Frieze, Olson & Russell, 1991). This work confirms these findings but suggests that appearance-based biases can be overcome by positive facial expressions. Smiling, therefore, can enhance the impact of positive behaviour, and help to diminish appearance-based biases and negative stereotypes.

Chapter Four

Do social utility judgements influence attentional processing?

Danielle M. Shore and Erin A. Heerey

Paper submitted for publication.

Abstract

Research shows that social judgements influence decision-making in social environments. For example, judgements of an interaction partner's trustworthiness affect a variety of social behaviours and decisions. One mechanism by which they may influence social decisions is by biasing the automatic allocation of attention toward certain social partners, thereby shaping the social information people acquire. Using an attentional blink paradigm we investigate how appearance- and behaviour-based trustworthiness judgements alter the allocation of attention to social stimuli in a set of two studies. In the first, strong appearance-based judgements (positive and negative) enhanced stimulus recognizability but did not alter the size of the attentional blink, suggesting that appearance-based social judgements enhance face memory but do not affect pre-attentive processing. However, in the second study, in which judgements were based on behavioural experience rather than appearance, positive judgements enhanced pre-attentive processing of trustworthy faces. This suggests that a stimulus's potential benefits, rather than its disadvantages, shape the automatic distribution of attentional resources. These results have implications for understanding how appearance- and behaviour-based social cues shape attention distribution in social environments.

Humans make hundreds of decisions every day. Often, these choices are heavily dependent on the signals people receive from their interaction partners (Behrens, Hunt, Woolrich & Rushworth, 2008). One factor that shapes decision-making in social contexts is people's judgements of their interaction partners. These social judgements are important because they guide expectations about how a partner might behave (Cosmides, 1989; Cosmides & Tooby, 2000; Frith & Frith, 1999). Appearance-related social judgements are particularly influential (Willis & Todorov, 2006). For example, research shows that the degree to which an individual looks trustworthy influences a range of decisions including financial investments (van't Wout & Sanfey, 2008), the interpretation of verbal information (Hassin & Trope, 2000), wagering behaviour (Schlicht et al., 2010), legal decisions (Porter, ten Brinke & Gustaw, 2010), and voting (Olivola & Todorov, 2010).

Nonetheless, appearances are not always accurate (Porter, England, Juodis, ten Brinke & Wilson 2008). Research has therefore begun to examine how behavioural experience alters social judgements. This work shows that people's behaviour significantly alters others' judgements such that truthful, consistent, and prosocial behaviours lead to more positive interpersonal evaluations (Ames & Johar, 2009; Bayliss & Tipper, 2006; Heerey & Velani, 2010). Thus, these findings suggest that both appearance- and experience-based judgements influence social decisions.

Recently, research has begun to suggest that social judgements may be akin to economic value judgements (Chang, Doll, van't Wout, Frank & Sanfey, 2010) because they shape expectations about the utility or subjective desirability of interacting with a particular social partner. For example, the presence of social rewards such as genuine smiles increases stimulus utility and influences subsequent economic decisions (Shore & Heerey, 2011). Moreover, people assume that interaction partners who look attractive or trustworthy, or engage in prosocial behaviour will provide positive outcomes and other social rewards (Delgado, Frank & Phelps, 2005; Wilson & Eckel, 2006). This evidence therefore suggests that social cues lead to joint economic and social judgements of

interaction partners, which subsequently influence decisions by biasing people's expectations about those partners.

Biases based on social judgements may shape decision-making in more subtle ways as well. For example, research has shown that non-social reward cues change stimulus utility, and that this influences the pre-attentive processing of rewarded stimuli (Anderson, Laurent & Yantis, 2011; Dux & Marois, 2009; De Martino, Kalisch, Rees & Dolan, 2009). If social utility judgements influence neural processing in a similar fashion, we predict that social stimuli eliciting strong judgements should likewise capture attention, suggesting one mechanism by which those stimuli shape social decisions and behaviour. Specifically, the ability of an interaction partner to capture attention, even for a short while, may bias the information one gains during an interaction involving that partner. Therefore, social judgements may bias decision-making by guiding attention toward or away from particular partners, thereby determining the information people acquire and use to make subsequent decisions.

Here, we ask how appearance- (Experiment 1) and behaviour-based (Experiment 2) social utility judgements affect the allocation of attention to social stimuli. Understanding how such judgements shape the perception and attentional processing of stimuli provides an important clue about how social judgements influence decision-making processes. To measure differences between stimuli in terms of attention capture, we utilized an attentional blink (AB) paradigm (Raymond, Shapiro & Arnell, 1992; Raymond & O'Brien, 2009). The AB is an elegant way of measuring the degree to which different stimuli automatically capture attention. In AB tasks, participants must detect two visual stimuli presented at varying time points in a rapid stream of images. If the second stimulus occurs within 500ms of the first, it is often undetected (Chun & Potter, 1995), causing an impairment in perceptual encoding known as the attentional blink (Raymond et al., 1992). Interestingly, participants are less likely to miss a target presented within 500ms of another when the target is emotionally salient (Keil & Ihssen, 2004). If social judgements increase the motivational or emotional salience of social stimuli this should cause those

stimuli to capture attention, even when they are presented within the window of the attentional blink (i.e., within 500ms following another target).

Experiment One

Experiment 1 investigates whether appearance-based social utility judgements affect the recognition of faces when they appear within versus after the window of the attentional blink. In this experiment, we use judgements of trustworthiness, as this trait is judged quickly, reliably and automatically from physical appearance (Olsen & Marshuetz, 2005; Todorov, Pakrashi & Oosterhof, 2009; Berry & Brownlow, 1989; Engell, Haxby & Todorov, 2007; Todorov, Said, Engell & Oosterhof, 2008; Willis & Todorov, 2006). We predict that when stimuli appear outside the window of the attentional blink, recognition will be better for faces judged to be high or low in trustworthiness compared to average (medium) rated faces, because faces with more extreme ratings are thought to be more salient than average faces (Singer, Kiebel, Winston, Dolan & Frith, 2004; Winston, Strange, O'Doherty & Dolan, 2002).

For stimuli presented within the window of the attentional blink, however, the literature suggests two predictions. Given that valuable or positive stimuli reduce the attentional blink (e.g., Anderson Laurent & Yantis, 2011; Raymond & O'Brien, 2009), one might predict that trustworthy, but not untrustworthy or neutral faces would be preferentially processed and therefore reduce the attentional blink. Alternatively, based on research showing that less trustworthy faces are more likely to be remembered (e.g., Yamagashi, Tanida, Mashima, Shimoma & Kanazawa, 2003), one might anticipate a reduced attentional blink for faces that are low, rather than average or high in trustworthiness.

Method

Participants

Fifty five undergraduate psychology students (17 male, mean age=21.07 SD=3.97 participated in the study for partial course credit. All participants gave written informed

consent and the University's Ethics Committee approved the study (likewise for Experiment 2 below).

Stimuli

The task used three types of stimuli: faces, abstract images and image masks (see Figure 4.1). The face stimuli consisted of natural grey-scale images of actors' faces. Actors were photographed in neutral poses, with eye-gaze directed toward the viewer. Photos were presented in an elliptical window that closely cropped the face. There were 36 faces (18 male) in total, all of which were pre-rated for trustworthiness by an independent sample of 48 participants. Male and female faces were rated as equally trustworthy ($p=.15$). Based on these judgements, we split the faces into two sets. The first set consisted of 24 faces all rated as average in trustworthiness. These faces served as 'novel' faces in an attentional blink recognition task (see procedure). The second face set consisted of 12 faces rated as high in trustworthiness (4 faces), of average trustworthiness (4 faces) or low in trustworthiness (4 faces). Each level had 2 male and 2 female faces. A gender (male, female) and trust rating (high, average or low) ANOVA ensured that male and female faces were similarly trustworthy ($p=.28$) and that faces differed in trustworthiness across the rating categories ($p<.001$). High trustworthy faces ($M=5.78$ $SD=0.34$) were rated higher than average faces ($M=4.49$ $SD=0.18$), which in turn, were higher than low trustworthy faces ($M=2.93$ $SD=0.49$) on a 9-point Likert scale (1=Extremely untrustworthy; 9=Extremely trustworthy).

To test whether the natural face images differed in low-level image characteristics across levels of trustworthiness, we used a 2-dimensional discrete fast Fourier transform (in MATLAB) to decompose each image into its component frequencies to produce a power spectrum with 30 frequency bands (Diop, Alexandre, & Moisan, 2012). Using mixed-model ANOVA, we compared signal power within the frequency bands (frequency band served as the within-subjects variable). Face gender (male, female), and trustworthiness condition (low, average, high) were both between-subjects variables and frequency (30 levels) was the within-subjects factor. As with most natural images, there was greater low-frequency than high-frequency information ($p<.001$). There was also a

main effect of face gender such that female faces had more low-frequency information than did male faces ($p<.001$). Male and female faces did not differ in terms of power in the higher frequency bands, meaning that the face-gender by frequency band interaction was also significant ($p<.001$). Importantly, there were no differences in power spectra across trustworthiness conditions ($p=.56$), nor was there a frequency band x trustworthiness condition interaction ($p=.99$). The 3-way interaction was not significant ($p>.99$).

The remaining stimuli included 20 computer-generated, grey scale, abstract images made up of either circles or squares; and 20 face 'masks' made by splitting face images into 4x5 grids and randomly shuffling the pieces (see Figure 4.1).

Procedure

Participants began the task by rating each of the 12 pre-rated faces on a number of traits including trustworthiness. Participants' ratings were similar to those of the previous sample. As above, high trustworthy faces ($M=5.53$ $SD=1.21$) received higher ratings than average faces ($M=4.31$ $SD=1.29$), which received higher ratings than low trustworthy faces ($M=2.77$ $SD=1.36$; p-values$<.001$).

Following the rating procedure, participants completed a simple 1-back task that allowed them to become familiar with the 12 faces. Faces were presented in random order and participants made a key press response each time they saw a face appear twice in a row (faces appeared as 1-back targets with equal probability). Each face was presented for 750ms, with a 250ms blank screen between presentations. There were 56 presentations of each face (672 trials total) split over 3 blocks, with an average of 29 one-back trials per block. There were no differences in the frequency with which participants responded to each face when it was a 1-back target ($p=.72$).

Participants then completed an attentional blink (AB) task to assess whether appearance-based judgements of trustworthiness affected the recognition of faces when these were presented during or after the window of the attentional blink. Participants searched for two targets T1 (an abstract image) and T2 (a face) in a rapid serial visual presentation (RSVP; see Figure 4.1). On each trial, participants viewed the RSVP stream and answered a question about each target they had seen.

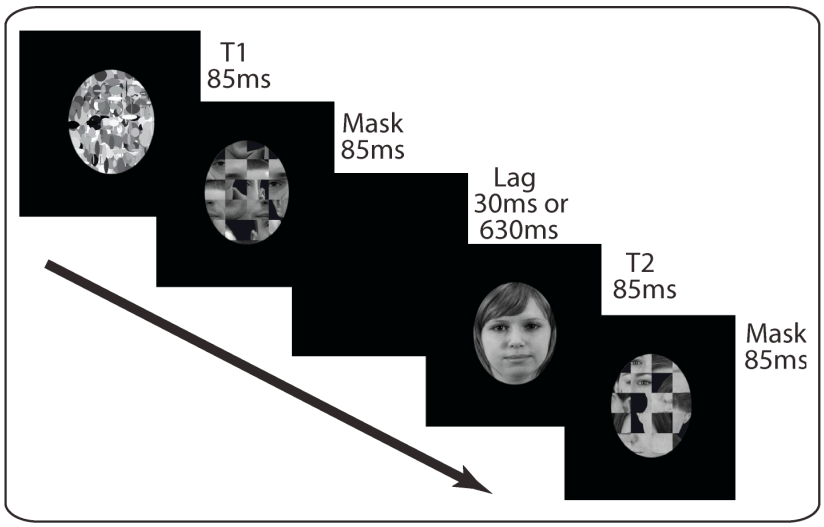

Figure 4.1. Trial timeline for the attentional blink task. Participants saw an RSVP stream consisting of two image-pairs separated by a blank screen of variable length.

Each trial started with a fixation cross for 1000ms followed by an RSVP stream consisting of 5 stimuli (T1, mask, blank screen, T2, mask). Each image was presented for 85ms except the blank screen, which had either a short (30ms) or a long (630ms) duration or lag. The two lag durations created two attentional conditions. When the lag was long (full attention condition), T2 was presented after processing of T1 was complete, meaning it appeared outside the window of the attentional blink and therefore received full attention. When the lag was short (reduced attention condition), T2 was presented within the window of the attentional blink prior to the completion of T1 processing, thus reducing the attention it received (Raymond et al., 1992).

After viewing the RSVP stream participants classified the T1 image as consisting of circles or squares and decided whether the T2 face was familiar (one of the 12 faces they had seen in the 1-back task) or unfamiliar (one of the 24 novel faces not present in the 1-back task). Participants answered each question with a key-press response. There was no time limit for responding. The next trial began immediately after participants responded to the second question.

Participants completed 240 long-lag (full-attention) trials and 240 short-lag (reduced-attention) trials in random order. In half the trials the T2 stimuli were familiar faces (from the 1-back task) and the other trials, they were novel faces from the set of 24 neutral/average-trustworthiness faces. Face novelty was balanced across long- and short-

lag trials. In the case of the familiar faces, which differed on ratings of trustworthiness, trustworthiness level (low, average and high) was balanced across the short- and long-lag conditions. Face gender was also balanced across conditions. T1 stimuli were equally likely to consist of circles or squares (randomly selected from the appropriate image pool), counterbalanced across lag, face novelty, face gender and face trustworthiness (familiar faces only) conditions. Mask images were randomly selected on each trial. The task was programmed and presented using E-prime (Psychology Software Tools, Inc., Pittsburgh, PA).

Data analysis

We classified the AB task responses as hits (correctly identifying familiar faces), misses (responding 'unfamiliar' to a familiar face), correct rejections (correctly identifying unfamiliar faces) or false alarms (responding 'familiar' to an unfamiliar face). False alarm and correct rejection rates were the same across all conditions as these were calculated from the novel faces. We then used signal detection theory (Green & Swets, 1966) to calculate d' for each participant as a measure of recognition performance for each face category (low-, average- and high-trustworthiness), across the long and short lag conditions. Only trials in which participants correctly identified the T1 stimulus were used in analyses. We applied Bonferroni's correction for multiple comparisons to all post-hoc comparisons (likewise for Experiment 2).

Results

A 2x2x3 repeated measures ANOVA with lag condition (short or long), face gender (male or female) and face type (high-, average- or low-trust) as within subjects factors and d' as the dependent variable, showed that there was no main effect of face gender, $F(1,54)=3.19$, $p=.08$, $\eta^2_p=.05$, nor did face gender interact with any other factors (all p-values>.08). Therefore, we collapsed across face gender for all analyses. As Figure 4.2 shows, a lag condition (short or long) by face type (high-, average-, or low-trustworthiness) repeated measures ANOVA showed that participants' recognition of T2 stimuli was better when attention was full (long lag) than when attention was reduced

(short lag), $F(1,54)=42.71$, $p<.001$, $\eta^2_p=.44$. T2 recognition also depended on face type, $F(2,108)=7.71$, $p=.001$, $\eta^2_p=.13$. Specifically, high- and low-trust faces were better recognized than average faces (p-values<.04), although high- and low-trust faces did not differ from one another in recognition (p=.19). There was no trust condition by lag interaction, $F(2,108)=1.20$, $p=.31$, $\eta^2_p=.02$.

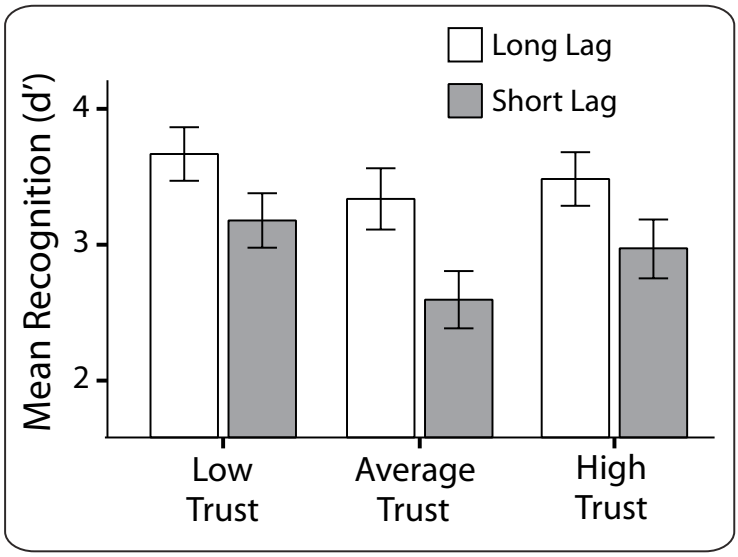

Figure 4.2. Experiment 1 results: Participants' average recognition (d') of faces in both full (long lag) and reduced attention (short lag) conditions according to face type (high-, average- or low-trustworthiness). Error bars indicate +/- 1 SEM.

A difference in recognition performance between attention conditions indicates the presence of an attentional blink (Raymond et al., 1992). To detect the presence of an AB, we used paired-samples t-tests to compare long- and short-lag trials for each face type (see Raymond & O'Brien, 2009). Interestingly participants showed an attentional blink for all three face types (p-values<.008), meaning that the recognition of T2 stimuli was significantly reduced in the short-lag compared to the long-lag trials. Together, these results suggest that strong appearance-based social judgements make stimuli more memorable but only in the long-lag condition. Therefore, appearance-based social judgements do not alter the size of the attentional blink.

Discussion

As predicted, in the full attention condition, salient faces (those judged to be high or low in trustworthiness) were better recognized than those judged to be average in trustworthiness. Contrary to prediction, however, this effect did not extend to the pre-attentive processing of low- or high-trustworthy faces. The finding, that faces eliciting more extreme social judgements (both positive and negative) are better recalled than those of more average ratings, supports the idea that such social utility judgements enhance memory for these stimuli (Singer et al., 2004). However, appearance-based social utility attributions are not strong enough to affect pre-attentive processing, as evidenced by their failure to attenuate the attentional blink. This suggests that insofar as appearance-based judgements about an interaction partner's trustworthiness affect social decision-making (e.g., Schlicht et al., 2010; van 't Wout & Sanfey, 2008), they do so via other channels than attentional biasing.

Experiment Two

Although appearance-based social judgements influence behaviour towards social partners, sound social decisions require a great deal more information than that conveyed by appearance alone. For example, it is unwise to make important decisions, such as choosing a babysitter, without any knowledge or experience of an individual's behaviour. Thus, when no other information is available, social decision-making may rely on appearance-based social judgements. However, as behavioural information becomes available decisions should shift to reflect actual experience, rather than appearance-based inferences, as behaviour is a more reliable indicator of a person's true social utility than appearance (Chang, et al., 2010; Rudoy & Paller, 2009; Singer, et al., 2004). Therefore, Experiment 2 examined social judgements based on an experienced history of behaviour.

To generate behaviour-based differences in trustworthiness, we used a multiple-round investor-trustee game (Berg, Dickhaut & McCabe, 1995) to familiarize participants with a set of faces. The trained faces served as stimuli in an AB task similar to that in Experiment 1. We predicted that behavioural experience with a partner would shape both

recognition and the allocation of attention in the AB paradigm. Specifically, we anticipated that in the full attention trials, faces with both positive and negative behavioural histories (e.g., fair and unfair players in the game) would be better recognized than faces that provided no behavioural feedback during the task (neutral faces). However, on reduced attention trials, behavioural history would bias attention, such that trustworthy faces would overcome the attentional blink (Anderson, Laurent & Yantis, 2011; Raymond & O'Brien, 2009).

Method

Participants

Forty-one undergraduate psychology students (14 male, mean age=21.10 SD=4.76) participated in the study for course credit and a small monetary bonus, dependent on their earnings in the investor-trustee game.

Stimuli

Experiment 2 used 36 face stimuli (18 males), pre-rated as similar and average in trustworthiness (these included the 24 average-trustworthiness faces from Experiment 1). Abstract images and masks were the same as in Experiment 1 (see Figure 4.1).

Procedure

In a multi-round investor-trustee game, participants played the role of the investor with 12 randomly selected faces (6 male) as trustees. At the start of each trial, participants received an endowment of 9 points. They then viewed a centrally presented image of the trustee and chose how much of their endowment they wished to invest. The investment matured (tripled), and a feedback display informed participants about the number of points the trustee had returned. Participants played 6 blocks of 50 trials (300 total; 25 trials per face in random order).

Four trustees were randomly assigned to play fairly (2 male), four unfairly (2 male) and four were neutral (2 male). Fair trustees always returned a randomly chosen amount between 40% and 70% of the matured investment and unfair trustees always returned between 0% and 30%. The exact percentage of return was randomly chosen on a round-

to-round basis from a uniform distribution. On average, fair players returned 52% of the matured investment (SD=0.05) and unfair players returned 13% (SD=0.07). Neutral trustees always displayed 'this is a no feedback trial' instead of an investment return. Trustees remained consistently fair, unfair or neutral across all task blocks.

After the investor-trustee game participants completed the same AB task as in Experiment 1, to assess whether trustee behaviour altered the allocation of attention. In this version of the AB task, the 'familiar' T2 stimuli were the 12 faces from the investor-trustee game. The novel T2 faces were the 24 average-trustworthiness face images that had not been included in the game.

Results

A 2x2x3 repeated measures ANOVA with lag condition (short or long), face gender (male or female) and face type (high-, average- or low-trust) as within subjects factors and d' as the dependent variable, showed that there was no main effect of face gender, $F(1,54)=3.19$, $p=.08$, $\eta^2_p=.06$, nor did face gender interact with any other factors (all p-values>.08). Therefore, we collapsed across face gender for all analyses. To assess learning in the investor-trustee game we calculated the average of the first five and last five investments to each face (early and late investments, respectively). A 2x3 repeated measures ANOVA with investment average as the dependent variable and investment time (early or late) and trustee behaviour (fair, neutral or unfair) as within-subjects variables, showed that participants learned trustee behaviour, $F(2,80)=80.92$, $p<.001$, $\eta^2_p=.67$, changing their investments significantly over time. Post-hoc comparisons showed increased investments to fair playing trustees, $t(40)=6.73$, $p<.001$, and decreased investments to unfair playing trustees $t(40)=-8.66$, $p<.001$. However, investments to neutral trustees decreased for the group as a whole, $t(40)=-4.69$, $p<.001$, but increased for some participants, suggesting that some participants' judgements of the 'neutral' trustees were not actually neutral.

Due to time constraints, participants did not rate the faces for trustworthiness after the task. However, data from a pilot sample of 28 participants showed that

investments to the neutral (no feedback) trustees strongly correlated with post-game ratings of those trustees (r=.55, p=.002). Participants who reduced their investments to the neutral trustees by more than 1.5 points from the first to the last investments rated them as negative (M=3.16, SD=0.86). Those who increased their investments by more than 1.5 points rated them as positive (M=5.25, SD=1.03). Participants whose investments did not change toward the neutral trustees continued to rate them as neutral (M=4.4, SD=1.65). This suggests that investment behaviour is a good proxy for ratings. Therefore, we classified neutral trustees on a participant-by-participant basis as fair, neutral or unfair depending on each participant's change in investment toward the trustee.[2] Using the differences between the early and late investments to measure change, we classified faces to whom participants increased their investments by >1.5 points as fair; faces to whom investments decreased >1.5 points as unfair; and faces to whom investments changed by less than 1.5 points as neutral.

To determine AB task results based on trustee values, we used the same signal detection approach as in Experiment 1. We calculated d' as a measure of the sensitivity of recognition performance for each face type (fair, neutral, and unfair) at each lag condition (short and long). Only trials in which participants correctly identified the T1 stimulus were used in analyses. A 2x2x3 repeated measures ANOVA with lag condition (short or long), face gender (male or female) and face type (fair, neutral, and unfair) as within subjects factors and d' as the dependent variable, showed that there was no main effect of face gender, $F(1,9)=3.67$, $p=.09$, $\eta^2_p=.29$, nor did face gender interact with any other factors (all p-values>.08). Therefore, we collapsed across face gender for all analyses.

An attention condition (full or reduced) by trustee behaviour (fair, neutral or unfair) repeated measures ANOVA showed that participants' recognition accuracy was

[2] Data from an independent sample of participants who played a similar investor-trustee game as the one used in this study showed that trustee behaviour significantly changed trustee trustworthiness ratings from pre-game to post-game. For faces randomly assigned to play fairly, participants increased their ratings of trustworthiness by 1.65 points (SD=1.90). For those who played unfairly, trustworthiness ratings decreased by 1.91 points (SD=1.66). On average, ratings of neutral (no feedback) faces remained the same as before the game, changing by 0.29 points (SD=0.83). For the fairly and unfairly playing faces, the change in trustworthiness rating correlated with participants' average return from those faces (fair trustees, r=.39, p=.002; unfair trustees, r=.49, p<.001).

generally better for full attention (long-lag) trials compared to reduced attention (short-lag) trials, $F(1,27)=16.22$, $p<.001$, $\eta^2_p=.38$. However, there was also a main effect of trustee behaviour, $F(2,54)=16.68$, $p<.001$, $\eta^2_p=.38$, such that fair trustees were recognized better than unfair or neutral trustees (p-values<.01). There were no differences in recognition of unfair and neutral partners (p=.14). This suggests that behaviour-based social utility attributions affect the recognition of social stimuli.

As predicted, paired samples t-tests confirmed the presence of a sizable attentional blink for both unfair and neutral trustees (both p-values<.01) but not for fair trustees (p=.238), suggesting that the photos of fair trustees were pre-attentively processed (see Figure 4.3). That is, the appearance of a fair trustee within a rapid series of images overcame the attentional blink, thereby facilitating perceptual recognition, even when attentional resources were otherwise engaged.

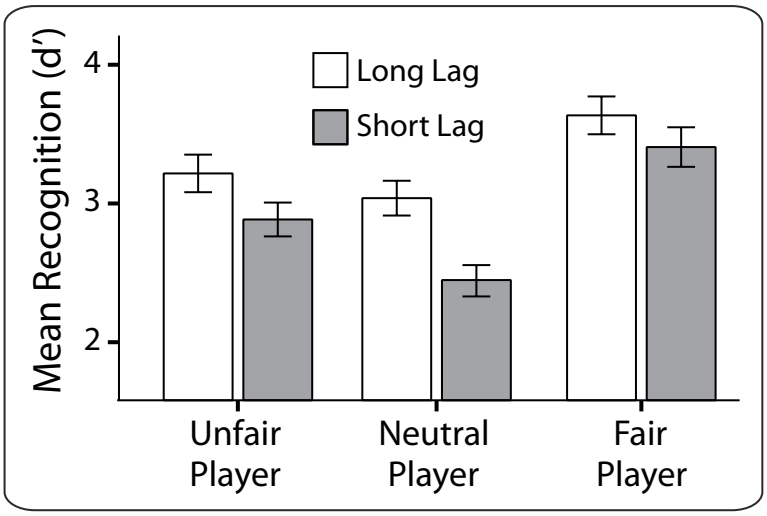

Figure 4.3. Experiment 2 results: Participants' average recognition (d') of faces in both full (long lag) and reduced attention (short lag) conditions according to return strategy (unfair, neutral or fair) in the investor-trustee task. Error bars indicate +/- 1 SEM.

Discussion

These results showed that judgements based on behavioural experience alter the involuntary allocation of attention to social stimuli. Specifically, interaction partners who acquired positive social values in the investor-trustee game were better recognized in both the full and reduced attention conditions, compared to partners with acquired

neutral and negative social values. Indeed, participants showed no attentional blink for the trustworthy faces. This result opposes the frequently expounded idea that people should allot more attention to untrustworthy individuals due to their potential to cause harm (Cosmides & Tooby, 2000). Rather, these results suggest that a stimulus's potential benefits alter pre-attentive processing thereby biasing attention towards trustworthy faces. These results have important implications for understanding how the signals people send shape attention distribution in a social environment.

General Discussion

Together, these results show two important things. First, people who elicit strong social impressions, both positive and negative and regardless of whether these judgements are based on appearance or behaviour, are more recognizable. Second, social utility judgements must be based on behavioural experience to alter pre-attentive processing. In Experiment 1, appearance-based social utility judgements, affected face recognition such that faces provoking stronger or more salient judgements (both positive and negative) were better recognized than those with average values. However, when constraints were placed on attention, appearance-based social utility judgements reduced the size of the attentional blink. Conversely when people made social utility judgements based on a behavioural history, these learned judgements led to better recognition of high utility faces, regardless of limits on attention. These results suggest that learned social utility, from cues integrated over time, shapes both interpersonal judgements and the pre-attentive processing of specific social stimuli.

One implication of this research is that appearance-related trait judgements are unlikely to markedly alter how people allocate attention and gather information in social settings. That is, although people may make strong appearance-based judgements (Hassin & Trope, 2000), these appear to affect social decisions at later time-points when cognition is conscious, rather than at pre-conscious processing stages. Because conscious cognition may be more easily modified by controllable mechanisms (Corbetta & Shulman, 2002),

people may have more cognitive control over this type of appearance-based social stereotype than previous research implies (Bargh & Williams, 2006).

Importantly, despite the fact that the faces in Experiment 1 differed slightly in physical characteristics, we do not believe that our results were driven by lower-level visual differences between these stimuli. Our image analysis only showed differences in low-level image characteristics between male and female faces, both of which were present in the task. However, male and female faces were equally well recognized and all faces were equally familiar. Thus, the recognition advantage for high- and low-trustworthy faces likely depends on social judgements, rather than on any physical differences in the faces themselves.

Interestingly, when social utility judgements were based on behavioural history, rather than appearance, it was positive judgements that shaped the allocation of attention. In this case, learned positive judgements both eliminated the attentional blink pre-attentively and enhanced recognition under the full-attention condition. This result is consistent with a value-based account of attentional processing (Anderson et al., 2011; Raymond & O'Brien, 2009) and contrasts with accounts predicting that untrustworthy faces should capture attention (e.g., Winston et al., 2002) appearance-based social utility judgements did not show early processing biases. One explanation for how this early processing bias may occur is that reinforcement learning processes may modulate selective attention and thus guide learning of perceptual characteristics (Roelfsema, Ooyen, & Watanabe, 2010). According to this model, positively valued faces receive more attention and thus their perceptual characteristics should be preferentially learned, compared to negatively valued faces.

These results lend support to the idea that it is goal-related or motivationally salient stimuli that capture attention rather than emotionally negative stimuli (Raymond & O'Brien, 2009). To maximise gains in the investor-trustee game it was equally important to learn which faces would give fair and unfair returns – and participants did indeed learn how to invest with both types of players. However in the AB task fair faces captured attention better, even under conditions in which attention was limited. This suggests that

even when facial trustworthiness is irrelevant to the task, the most salient faces are those with highest social utility. Judgements of social utility may therefore bias the allocation of attention in social environments toward high-utility individuals, but only if this value is acquired from behaviour. When social utility judgements depend on potentially inaccurate appearance-based data, they do not have the power to highjack attentional resources.

In Experiment 2, unfortunately, we cannot disentangle the social and monetary values of our stimuli. However, contrary to previous research (Raymond & O'Brien, 2009) we found that valence mattered in the full-attention as well as in the reduced attention conditions. This suggests that social utility based on behavioural experience is different from that based solely on financial value and changes perceptual recognition even when attention is fully available. To test this idea, we are currently investigating how trustworthiness based on other types of behavioural experience (e.g., deception and honesty) alters social judgements and pre-attentive processing to decouple social from financial value.

Social environments are complex and social partners provide many appearance- and behaviour-based cues that can change from moment-to-moment. Social interactions, especially those consisting of multiple partners, contain more information than people are able to process (Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010). Therefore, they must selectively choose whom they will attend and whom they will ignore at each moment in time. This research suggests that in such environments the allocation of attention will be biased towards interaction partners with whom one has a positive behavioural history. Ultimately, this means that some interaction partners will garner more attention than others – and that the cues other partners send may be ignored, leading to bias in the acquisition of social information and thereby influencing people's ability to make social decisions at a fundamental level.

Chapter Five
General Discussion

Generally, the work in this thesis examined the ways in which an interaction partner's social cues shape social decision-making, especially when those cues occur in combination. The broad aim of this work was an empirical investigation of the contributions different types of social cues (static, dynamic and behavioural) make to social utility, and to understand how social utility judgements influence decision-making processes. Together, this research suggests that static, dynamic, and behavioural social cues interact, rather than being combined in an additive or linear fashion to shape estimates of social utility. It also suggests that one fundamental mechanism by which judgements of social utility guide social decision-making processes is by shaping the allocation of attention in social environments.

Together, findings are consistent with the application of reinforcement learning models applied to social behaviour (Behrens et al., 2008) and suggest that biased social utility estimates may explain irrational social decisions under some circumstances. Specifically, this work shows that dynamic expressive cues, in this case smiles, have real reinforcement value, adjusting the utility of an interaction partner as measured in monetary terms. Indeed, we found that interaction partners who smile genuinely are treated as more economically valuable than they actually are. Interestingly, the value of these smiles can help to overcome conflict between static and behavioural cues. Specifically, we found that first impressions cause an irrational bias in behaviour towards interaction partners. However, the addition of a valuable dynamic cue overcomes this bias. Finally, one mechanism by which social utility estimates may influence social decision-making is by biasing attentional processing. In this pair of studies, results showed that strong appearance-based utility estimates (positive or negative) made faces more recognizable. However, behaviour-based utility estimates biased attention towards interaction partners with high, rather than low social utility.

The first set of experiments (Chapter 2) investigated whether smiles had intrinsic reward value and whether this value could be measured in economic terms. Using an

economic game and applying expected utility theory (von Neumann & Morgenstern, 1947), I compared the value of non-social rewards (money) and two social cues (genuine and polite smiles). Results showed first that social feedback is more valuable than non-social feedback, supporting the idea that social information is a valuable commodity (Klein et al., 2009) and may therefore be intrinsically rewarding (Matthews & Wells, 1990; Tsukiura & Cabeza, 2008). Second, results demonstrate that genuine smiles are more valuable than polite smiles. Specifically, although both monetary and social values contributed to participants' decisions, participants were willing to relinquish the chance to win money in order to receive a genuine smile. If genuine smiles did not enhance the subjective value or utility of the interaction partners displaying them, only the true monetary value of those partners would have driven choice behaviour. Interestingly, participants also explicitly over estimated the value of genuinely smiling faces. Thus, the value of social and non-social rewards appear to be additive. That is, genuine smiles simply add value to financial rewards, rather than interacting as in several other studies (Averbeck & Duchaine, 2009). Together the findings suggest that genuine smiles are social rewards and therefore increase the utility of the interaction partners who display them.

The second empirical chapter (Chapter 3) examined how people integrate different social cues to create estimates of social utility and how these estimates influence decision-making when they are consistent and when they conflict. I investigated how static appearance-based cues (first impressions of trustworthiness), dynamic expressive cues (smiles) and interaction partner behaviour (trustworthy behaviour) jointly affected social decision-making using an economic game. To learn how these cues interact, I manipulated their consistency to examine decision-making under the straightforward influences of these cues when they are consistent, as well as when they conflict. When cues are consistent, for example, a trustworthy face that reciprocates one's trust, findings were quite straightforward. That is, people were willing to make high investments to trustworthy, fair-playing faces and invested only minimally with unfair playing, untrustworthy looking faces. Surprisingly however, unfavourable appearance-based judgements continued to bias investment behaviour long after people had learned to trust

a social partner based on that partner's behaviour, supporting the idea that behaviour and appearance cues combine to inform expected utility judgements (Quist et al., 2012). Interestingly, a smile present at the time feedback eliminated this bias. This suggests that when static appearance-based cues and behavioural cues conflict, dynamic cues (expressions) can be powerful modulators of social utility. Results also show the timing of a smile is important. Specifically, consistent with Chapter 2, findings show it is the additional value that smiles afford to feedback that substantially increases social utility. Together, these findings show that stable or so-called 'unfakeable' (Rezleszcu et al., 2012) appearance-based cues can irrationally bias decisions. However, dynamic appearance-based cues (smiles) can enhance the impact of positive behaviour by eliminating appearance-based biases.

The third set of experiments (Chapter 4) examined how social cues shape decision-making by exploring one potential mechanism that explains how social utility judgements influence decisions. Specifically, different types of social judgements may bias the allocation of attention during social interactions, thereby influencing the information people acquire. Using an attentional blink task, I assessed how appearance- and behaviour-based social utility estimates changed attentional processing of social stimuli. Results showed that social partners with more extreme utility estimates (both positive and negative) based on appearance or behaviour were better recognized. However, only positive behaviour-based social utility estimates altered pre-attentive processing. In contrast with previous research that posits that untrustworthy partners should capture attention (Winston et al., 2002), this suggests that behaviour-based social utility estimates bias attention towards trustworthy interaction partners. Trustworthy partners have high social utility values compared to untrustworthy partners. Hence, this finding is consistent with value-based explanations of selective attention (Anderson et al., 2011; Raymond & O'Brien, 2009). This suggests that in social environments, interaction partners with whom one has a positive behavioural history will receive more immediate attention. In real-world social environments, the movements and cues of these partners may capture attention, drawing it away from other social partners. This will bias the acquisition of

social information by shifting attention toward some sources and away from others. Biases in the acquisition of social information may therefore profoundly influence the social decisions people make.

Together, this research suggests that an interaction partner's social utility has an important influence on how people perceive and interact with that partner. These findings confirm that trustworthiness is an important element of social utility (van t' Wout & Sanfey, 2008) and as such has a significant impact on decision-making (Kim, Choi & Jang, 2011). Findings also support the idea that the judgements people make about social partners based on their appearance create expectations about their likely behaviour (e.g., Gibson, 1979). In line with previous work, the present research found that beliefs about the trustworthiness of an interaction partner had strong influences on behaviour towards that partner (Chang et al., 2010; Delgado et al., 2005; Schlicht et al., 2010; van t' Wout & Sanfey, 2008). Further, this work demonstrates the full extent to which first impressions affect social behaviour; these results show that first impressions have an extremely powerful influence on decision-making, such that they continue to influence behaviour even after a wealth of contradictory behavioural experience. This suggests that appearance-based judgements of trustworthiness may not actually be as adaptive as many theories claim (e.g., Bond, Berry & Omar, 1994; Cosmides & Tooby 2000). Specifically, these irrational biases lead to suboptimal decision-making, serving to reduce participants' ability to maximize their financial gains.

Nonetheless, the degree to which biases based on appearance are irrational does depend on the accuracy of first impressions. If a first impression is accurate then it is highly adaptive (Cosmides & Tooby, 2000). However, a large body of work shows that on average, first impressions, at least of trustworthiness, may not be valid in terms of their ability to predict actual trustworthiness, even if they are reliable in a statistical sense (Hassin & Trope, 2000; Masip, & Garrido, 2001; Olivola & Todorov, 2010; Porter, England, Juodis, van Brinke & Wilson, 2008; Zebrowitz, Voinescu & Collins, 1996).

Reinforcement learning models would predict that expected values would be adapted over time, so first impressions should only influence decision-making while there

is little or no other information available (Harris & Fiske, 2010). As more behaviour is experienced, the degree to which first impressions influence behaviour should reduce (Chang et al., 2010). Following this model one would predict that by the end of the task in Chapter 3 untrustworthy looking, fair playing social partners should have received the same investments as their trustworthy looking counterparts (Chang et al., 2010). Instead however, first impressions continued to influence behaviour long after that influence should have waned. This continued influence can be explained by understanding that untrustworthy looking faces are subjectively less desirable, such that partners who look untrustworthy do not become as subjectively valuable as those that look trustworthy, even though their behaviour is the same.

An important implication of this research is that despite having a striking influence on behaviour, appearance-related utility judgements are unlikely to alter how people distribute their attention in social settings. Thus, they are unlikely to influence the information collected in a social environment. These findings stand in sharp contrast to theories suggesting that people ought to attend to potential threats (Fox et al., 2000; Hansen & Hansen, 1988; Öhman, Lunquvist & Esteves, 2001). Results from Chapter 4 demonstrated that although strong first impressions led to better recognition of faces they did not change attentional processing. However, trained value, based on behavioural experience, did shape attentional processing, such that fair but not unfair trustees captured attention. Behavioural experience may have resulted in fair faces becoming both subjectively and objectively more desirable, thus resulting in a greater social utility value.

Importantly, the different effects of first impressions and behavioural experience suggest that it is the type rather than the amount of experience with a social partner that shapes attention distribution. This suggests that objective value has a stronger influence on attentional processing. This is, of course, rational. Therefore, it may be that the way someone looks is a bonus or additional factor that adds to his or her social utility rather than a key component. In this way the contribution of appearance to utility may be enough to shape decision-making but not sufficient to change the attentional processing of stimuli. This is important because it shows that first impressions do not bias pre-

attentive processing, which might impair the ability to learn about social partner's objective value. Thus if an impression were inaccurate, it might lead to a failure in the ability to maximize social gains.

Previous research suggests that people have little cognitive control over the formation of appearance-based social stereotypes (Bargh & Williams, 2006). In accord with this work, the present results show that people do make strong, long-lasting appearance-based judgements (Hassin & Trope, 2000). However, these appear to affect social decisions at a more conscious level, rather than pre-attentively. People may therefore have more cognitive control over these adaptable conscious processes (Corbetta & Shulman, 2002), thereby enabling appearance-based influences to be overcome. However, as the results of Chapter 3 demonstrate, these biases are only overcome under certain conditions and continue to influence behaviour to the detriment of decision-makers.

From a reward learning perspective, one explanation for this asymmetry in investment behaviour may be that the relative subjective value of the trustworthy and untrustworthy faces differs. Selecting faces from participant's own personal distributions (highest and lowest rated) controlled this to a certain extent. However, it may be that 'good' faces are not as far above baseline (average trustworthiness) as 'bad' faces are below baseline. Therefore, it may be that the fair-playing untrustworthy faces must increase further in subjective terms to catch up with the trustworthy faces than the unfair but trustworthy faces have to drop to become equal to the unfair, untrustworthy faces. If this is the case, increasing the financial or objective value of fair, untrustworthy looking partners by increasing their returns may also have overcome the asymmetry.

An alternative explanation of these results is that although negative prediction error is likely worse than positive prediction error, expectancy violation in itself may be unpleasant. Previous research shows that expectations are important, and social partners that violate expectations are punished (Chang & Sanfey, 2009; Sanfey, 2009; Wilson & Eckel, 2006). Because appearance-based judgements serve as initial utility estimates this may explain their extended influence. Specifically, a bad surprise is likely worse than a

good surprise is good, but surprise alone may be unpleasant. Thus any violation of expectancy is bad, this explains both why unfair playing trustworthy faces quickly receive the same investments as their untrustworthy looking counterparts and why untrustworthy faces that play fairly do not become as good as trustworthy faces that behave fairly.

Interestingly, in this investor-trustee context, we found that in order to make participants' behaviour conform to reinforcement-learning model predictions, the subjective desirability of the stimuli had to be enhanced. Displaying positive facial expressions eliminated the bias caused by appearance-based first impressions, even though the only thing smiles changed was facial appearance – the financial value of the faces remained the same. This result is consistent with a utility theory explanation, such that smiles made the partners more desirable, thus increasing their utility value and consequently, participants' investments. Perceivers infer both intentions and future behaviour from faces (Rezleszcu et al., 2012; Willis & Todorov, 2006). These results suggest that although having a face perceived as low in trustworthiness is disadvantageous, displaying expressions that reflect positive intentions and emotions can help to overcome this initial disadvantage. This research therefore takes psychology one-step closer to understanding social decision-making in live social interaction. In the real social world, neutral faces are rarely seen. This work is the first, to my knowledge, that shows how dynamic cues modulate the effects of static and behavioural cues on decision-making processes.

This thesis highlights the importance of subjective desirability in social decision-making. Reinforcement learning may be able to explain these findings, but in order to do so, factors within the models need adjustment. Utility theory provides a way of incorporating objective value and subjective goodness. Thus, introducing a factor that attempts to quantify subjective value in a reinforcement-learning model may result in a better explanation of the results. If reinforcement learning models can include a metric for the social 'prior' created by first impressions or by the subjective value of an interaction partner, they may be able to account for Chapter 3 findings.

Taken together the results of Chapters 2 and 3 show how powerful smiles are, both as rewards and as a component of social utility. Indeed, smiles significantly increase the social utility of individuals who display them. This fits with a utility theory perspective of decision-making, as genuine smiles of pleasure are subjectively desirable. Previous results suggest that smiles enhance the extent to which monetary rewards are assimilated (Averbeck & Duchaine, 2009). However, Chapter 2 results suggest instead that, when used in a consistent social context, their utility may be added to that of other rewards such as money. Chapter 2 results therefore support the idea that smiles influence decision-making processes by biasing the evaluation of rewarding outcomes. Thus, participants overestimated the magnitude of the financial outcomes associated with genuinely smiling faces.

In Chapter 2 genuine smiles were distinctly more valuable than polite. In Chapter 3 however, it appeared that the value of a smile depended on the appearance of the face, such that genuine smiles added greater utility to partners that looked trustworthy but polite smiles added greater utility to untrustworthy looking partners. This suggests that the value of a smile, and other social cues, is not fixed but is context dependent and changes depending on the expected value of the individual who provides the smile and the circumstances under which they do so. For example, in Chapter 2 smiles always indicated wins – and did not indicate a partner's intentional behaviour. Therefore, they were never inconsistent with the context of that outcome. Comparatively, in Chapter 3, the differences in outcome context directly resulted from partner behaviour – partners either reciprocated fairly or did not. Receiving a genuine smile from a trustworthy fair face might be better than a polite smile from the same face. However, when a trustworthy face plays unfairly, the opposite is true. The context, in this case the type of face and the way it behaves, jointly affect the interpretation of the smile cue. This suggests that on their own, genuine smiles are more valuable than polite but when associated with intentional behaviour, the value of a smile changes, possibly because its interpretation is subjective. That is, the degree to which a genuine smile is rewarding hinges on contextual factors. Thus, in Chapter 2 the genuine smiles caused irrational decision-making because people

made choices that did not maximize financial gains (Gilovich, Griffin & Kahneman, 2002). Comparatively, in Chapter 3 the influence of smiles from the fair but untrustworthy looking faces led to rational decision-making, as they resulted in participants behaving in a manner that led to greater financial gains than when there were no smile displays. Together these studies suggest that smiles are both valuable and desirable, as they shape how people behave towards social partners.

Limitations

As with all experimental work, the studies reported here have a number of limitations. In particular, all the studies used laboratory-based experiments to investigate social behaviour. Although this means that the stimuli and methods were well controlled, these tasks grossly oversimplify social interaction and the decisions people make in social contexts. Thus, they are only a partial substitute for the real social world. However, given the complexity and flexibility of real-world social behaviour, it is nearly impossible to control the social cues participants see in face-to-face social interactions. Even when using confederates, every participant experiences variation in the social stimuli they receive, which reduces the ability to interpret data. In order to confirm these results, they should be tested in real world environments to see if they can be replicated in the less controlled world of live social interaction. Nonetheless, the findings from this collection of experiments support findings from applied real-world studies (e.g. Little, Burriss, Jones & Roberts, 2007; Olivola & Todorov, 2010). Therefore, the fact that we find such powerful effects of social utility on behaviour, even in this artificial environment, suggests that static, dynamic and behavioural social cues are likely to have strong influences on decision-making processes in live social interactions as well.

A related limitation of this work is that the highly controlled and specific stimuli used here may limit the generalizability of results. For example, appearance-based utility estimates are likely based on a variety if traits, rather than trustworthiness alone, although research suggests that trust is a good approximation of general social utility (Oosterhof & Todorov, 2008). This research also focused on smiles, as they are among the

most commonly seen facial expressions in general interactions (Heerey & Kring, 2007; Hess & Bourgeois, 2010). However, other facial expressions may not function in the same way, as they have different communicatory functions and thus different associated values. For example, fearful faces are indicative of a threat within the social setting. Therefore these cues may interact differently with behaviour and static appearance-based cues to influence social utility.

A third limitation is that although findings suggest that behaviour-based social utility judgements shape both decision-making and the allocation of attention, social and financial value cannot be dissociated in many of these studies. Therefore, this research is limited in terms of its applicability to the less explicitly financial values of social interactions in the real world. For example, friendships often involve elements that have no objective value but are nonetheless highly valuable in a subjective sense. How these cues contribute to decision-making in these more intimate contexts remains unknown. Regardless, these results provide some insight into how the value of a cue shapes its function in these settings. They additionally provide support for a theoretical approach that can be applied to non-financial settings to understand how cues may integrate to influence behaviour in interpersonal contexts.

Future directions

The findings in this body of work suggest a series of new questions. For example, given that contextual factors seem to play a large role in shaping the utility of a particular cue, one must ask both what these factors are and how the brain combines this information to calculate a perception of social utility. Based on the findings from Chapter 2, I am currently investigating how the brain integrates social and monetary rewards in decision-making processes. Using functional neuroimaging data, I am examining how the brain processes monetary and social rewards separately and then uses this information to generate a joint percept of an interaction partner's utility.

The results reported here suggest that social information has value. In relation to finding that first impressions bias behaviour asymmetrically, one might predict that this

asymmetry changes the ways in which people conceptualize the entire context of a social environment. For example, research shows that the distribution and volatility of rewards in the environment shapes the degree to which people are willing to explore it (Behrens, 2009). Perhaps the asymmetry in investments reflects people's willingness to take social risks with certain types of social partners. In investor-trustee tasks, social risk taking is simply investing a greater proportion of one's endowment. Manipulating the reward distributions (e.g., changing the proportion or likelihood of trustees' returns) might highlight the degree to which biases in social judgements shape decision-making by altering people's tendency to risk financial or other gains. Insights from this study would help explain whether the asymmetry in investment behaviour found in Chapter 3 resulted from biased estimates of social utility or from differences in the rates at which people learn from different types of feedback. I am also interested in how different social cues affect the degree to which individuals explore opportunities to learn from different opponents. For example, a cue that biases a social learner toward trusting an untrustworthy opponent may lead to greater flexibility in social behaviour and faster adaptation to changes in social reward contingencies. In a similar vein, given that the context within which a smile occurs appears to affect its interpretation, contextual factors may influence the rate at which people learn from them and the degree to which they are subjectively valuable. This question deserves exploration in a proper experimental setting.

Future work will also decouple social utility from financial value to examine the effects of social versus financial value on attention (and other aspects of cognition) and on the computation of social utility. This would determine whether attentional biases found in Chapter 4 are specific to the financial aspects of an individual's utility or apply to the social aspects as well. To test this idea, we are currently investigating how trustworthiness based on other types of behavioural experience (e.g., deception and honesty) alters social utility judgements and attentional processing. In other avenues, this work provokes interesting questions about how the integration of dynamic expressive cues and behavioural cues shape social attention. Research already shows that emotional expressions shape the attentional processing of face stimuli (De Martino et al., 2009; Fox

et al., 2000). However, other work shows that people remember the association of a person with a particular expression (Tsukiura & Cabeza, 2008). Therefore, I expect that facial expressions during the familiarization phase of the task would affect attention even when neutral faces are used in the AB task itself. Specifically, I would anticipate that faces associated with positive facial expressions would have significant greater effects on attention. Faces associated with genuine smiles should capture attention due to their high utility value. This finding would further support the idea that positive social utility values shape social attention.

As a corollary, this work also raises the question of how social attention influences the acquisition of social information. Although experienced social behaviour appears to shape social attention, it is important to investigate how much this bias changes the information processed within a social environment. One-way to examine this is to investigate what information is processed in a given social context and how this changes depending on the utility of other information in an environment.

Conclusions

Collectively, this research shows that social partners' characteristics, derived from both static and dynamic appearance cues as well as behaviour, heavily influence social decision-making processes. Together, results suggest that irrational decision-making can be attributed to the biased valuation and integration of social cues. In real world terms, this work suggests that first impressions are important, but that behaviour provides a context for interpreting them. Theoretically, reinforcement-learning models may be able to characterize and explain how humans make social decisions, but in order to achieve good fits between predicted and observed data, better measures for quantifying subjective value must be derived. In this way, the concept of social utility can improve the application of reinforcement learning to social stimuli in order that it better accounts for the inherently qualitative and subjective value of social stimuli. Utility theory also highlights the importance of personal preferences and contextual factors in real-world social decision-making. This research provides new insight into how social

information provided by an interaction partner influences social decision-making. In real-world environments social decisions reflect the integration of social information and material value.

References

Adams, R. B. Jr., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion, 26*, 431-441. doi:10.1080/02699931.2012.666502

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*, 470–474. doi:10.1038/30982

Aharoni, E., & Fridlund, A. J. (2007). Social reaction towards people vs. computers: How mere labels shape interactions. *Computers in Human Behaviour, 23,* 2175-2189. doi:10.1016/j.chb.2006.02.019

Ames, D. R., & Johar, G. V. (2009). I'll Know What You're Like When I See How You Feel: How and When Affective Displays Influence Behavior-Based Impressions. *Psychological Science, 20*, 586-593. doi:10.1111/j.1467-9280.2009.02330.x

Anderson, B. A., Laurent, P. A., & Yantis S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences, 108*, 10367-10371. doi:10.1073/pnas.1104047108

Averbeck, B. B., & Duchaine, B. (2009). Integration of social and utilitarian factors in decision making. *Emotion, 9*, 599-608. doi:10.1037/a0016509

Banich, M. T., Mackiewicz, K. L., Depue, B. E., Whitmer, A. J., Miller, G. A., & Heller, W. (2009). Cognitive control mechanisms, emotion and memory: a neural perspective with implications for psychopathology. *Neuroscience & Biobehavioral Reviews, 33*, 613-630. doi:10.1016/j.neurbiorev.2008.09.010

Bar, M., Neta, M., & Lintz, H. (2006). Very First Impressions. *Emotion, 6,* 269-278. doi:10.1037/1528-3542.6.2.269

Bargh, J. A., & Williams, E. L. (2006). The Automaticity of Social Life. *Current Directions in Psychological Science, 15*, 1-4. doi:10.1111/j.0963-7214.2006.00395.x

Baron, R. A., Markman, G. D., & Bollinger, M. (2006). Exporting Social Psychology: Effects of Attractiveness on Perceptions of Entrepreneurs, Their Ideas for New Products,

and Their Financial Success. *Journal of Applied Social Psychology, 36*, 467-492. doi:10.1111/j.0021-9029.2006.00015.x

Bayliss, A. P., & Tipper, S. P. (2006). Predictive Gaze Cues and Personality Judgments: Should Eye Trust You? *Psychological Science, 17*, 514-520. doi:10.1111/j.1467-9280.2006.01737.

Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behaviour. *Science, 324*, 1160-1164. doi:10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, F. S. (2008). Associative learning of social value. *Nature, 456*, 245-249. doi:10.1038/nature07538

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*, 122-142. doi:10.1006/game.1995.1027

Berry, D. S., & Brownlow, S. (1989). Were the Physiognomists Right? Personality Correlates of Facial Babyishness. *Personality and Social Psychology Bulletin, 15,* 266-279. doi:10.1177/0146167289152013

Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science, 33*, 206–242. doi:10.1111/j.1551-6709.2009.01010.x

Blair, R. J. R. (2003). Facial expression, their communicatory functions and neuro cognitive substrates. *Philosophical Transactions of Royal Society B Biological Sciences, 358*, 561-572. doi:10.1098/rstb.2002.1220

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition, 57*, 1-29. doi:10.1016/0010-0277(95)00676-P

Bond, C. F., Berry, D. S., & Omar, A. (1994). The Kernel of Truth in Judgments of Deceptiveness. *Basic and Applied Social Psychology, 15*, 523-534. doi:10.1207/s15324834basp1504_8

Bowlby, J. (1958). The Nature of the Child's Tie to his Mother. *International Journal of Psycho-Analysis, 39*, 350-373.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction.* Princeton, NJ: Russell Sage Foundation - Princeton University Press.

Cash, T. F., & Kilcullen, R. N. (1985). The aye of the beholder: susceptibility to sexism and beautyism in the evaluation of managerial applicants. *Journal of Applied Social Psychology 15,* 591–605. doi:10.1111/j.1559-1816.1985.tb00903.x

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A.G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*, 87-105. doi:10.1016/j.cogpsych.2010.03.001

Chang, L. J., & Sanfey, A. G. (2009). Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining. *Frontiers in Human Neuroscience, 3,* 36. doi:10.3389/neuro.08.036.2009

Chun, M., & Potter, M. C. (1995). A two-stage model for multiple target detection in serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 109–127. doi:10.1037/0096-1523.21.1.109

Cohen, M.X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *The Journal of Neuroscience, 27,* 371-378. doi:10.1523/JNEUROSCI.4421-06.2007

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201-215. doi:10.1038/nrn755

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Watson selection task. *Cognition, 31*, 187-276. doi:10.1016/0010-0277(89)90023-1

Cosmides, L. & Tooby, J. (2000). The cognitive neuroscience of social reasoning. In M. S. Gazzaniga (Eds.), *The New Cognitive Neurosciences, Second Edition (pp. 1259-1270)*. Cambridge, MA: MIT Press.

Darwin, C. (1965). *The expressions of the emotions in man and animals*. Chicago, IL: University of Chicago Press. (Original work published 1872).

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron, 36*, 285-298. doi:10.1016/S0896-6273(02)00963-7

Delgado, M. R., Frank, R. H., & Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8,* 1611-1618. doi:10.1038/nn1575

De Martino, B., Kalisch, R., Rees, G., & Dolan, R. J. (2009) Enhanced processing of threat stimuli under limited attentional resources. *Cerebral Cortex, 19*, 127–133. doi:10.1093/cercor/bhn062

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, Biases, and Rational Decision-Making in the Human Brain. *Science, 313*, 684-687. doi:10.1126/science.1128356

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews Neuroscience, 18*, 193-222. doi:10.1146/annurev.ne.18.030195.001205

Diop, E. H. S., Alexandre, R. & Moisan, L. (2012). Intrinsic nonlinear multiscale image decomposition: A 2D empirical mode decomposition-like tool. *Computer Vision and Image Understanding, 116,* 102-119, doi:10.1016/j.cviu.2011.09.003.

Dux, P. E., & Marois, R. (2009) How humans search for targets through time: A review of data and theory from the attentional blink. *Attention, Perception and Psychophysics, 71*, 1683-1700. doi:10.3758/APP.71.8.1683

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*, 109-128. doi:10.1037/0033-2909.110.1.109

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*, 550-553. doi:10.1037/0033-295X.99.3.550

Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist, 48,* 384–392. doi:10.1037/0003-066X.48.4.384

Ekman, P., & Davidson, R. J. (1993). Voluntary Smiling Changes Regional Brain Activity. *Psychological Science, 4*, 342-345. doi:10.1111/j.1467-9280.1993.tb00576.x

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology, 58,* 342-353. doi:10.1037//0022-3514.58.2.342.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience, 19*, 1508-1519. doi:10.1162/jocn.2007.19.9.1508

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibrium". American Economic Review, 88, 848-881. doi:10.1006/game.1999.0783

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*, 419-427. doi:10.1016/j.tics.2007.09.002

Feingold, A. (1992). Good looking people are not what we think. *Psychological Bulletin, 111*, 304-341. doi:10.1037/0033-2909.111.2.304

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117*, 319-331. doi:10.1016/j.cognition.2010.09.003

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61-92. doi:10.1080/026999300378996

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioural markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64*, 83-93. doi:10.1037/0022-3514.64.1.83

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology, 4*, 247–256. doi:10.1016/162-3095(93)90020

Fridlund, A. J. (1991). The sociality of solitary smiles: Effects of an implicit audience. *Journal of Personality and Social Psychology, 60*, 229-240. doi:10.1037/0022-3514.60.2.229

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology, 21,* 1039–1057. doi:10.1111/j.1559-1816.1991.tb00458.x

Frith, C. D., & Frith, U. (1999). Interacting Minds – A Biological Basis. *Science, 286*, 1692-1695. doi:10.1126/science.286.5445.1692

Frith, C. D., & Singer, T. (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B, 363*, 3875-3886. doi:10.1098/rstb.2008.156

Geday, J., Gjedde, A., Boldsen, A. S., & Kupers, R. (2003). Emotional valence modulates activity in the posterior fusiform gyrus and inferior medial prefrontal cortex in social perception. *Neuroimage, 18*, 675-684. doi:0.1016/S1053-8119(02)00038-1

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects*. Journal of Personality and Social Psychology, 77*, 746-761. doi:10.1037/0022-3514.77.4.746

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases: *The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences, 1*05, 6741-6746. doi:10.1073/pnas.0711099105

Hansen, C. H., & Hansen, R. D. (1988). Finding the Face in the Crowd: An Anger Superiority Effect. *Journal of Personality and Social Psychology, 54*, 917-924. doi:10.1037/0022-3514.54.6.917

Harris, L. T., & Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. Social neuroscience, 5, 76-91. doi:10.1080/17470910903135825

Hassin, R., & Trope, Y. (2000). Facing Faces: Studies on the Cognitive Aspects of Physiognomy. *Journal of Personality and Social Psychology, 78*, 837-852. doi:10.1037//0022-3514.78.5.837

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*, 59-67. doi:10.1016/S0006-3223(01)01330-0

Heerey, E. A., & Kring, A. M. (2007). Interpersonal Consequences of Social Anxiety. *Journal of Abnormal Psychology, 116*, 125-134. doi:10.1037/0021-843X.116.1.125

Heerey, E. A., & Velani, H. (2010). Implicit learning of social predictions. *Journal of Experimental Social Psychology, 46*, 577-581. doi:10.1016/j.jesp.2010.01.03

Hess, U., & Bourgeois, P. (2010). You smile–I smile: Emotion expression in social interaction. *Biological Psychology, 84*, 514–520. doi:10.1016/j.biopsycho.2009.11.001.

Hooker, C. I., Germine, L. T., Knight, R. T., & D'Esposito, M. (2006). Amygdala Response to Facial Expressions Reflects Emotional Learning. *Journal of Neuroscience, 26*, 8915-8922. doi:10.1523/jneurosci.3048-05.2006

Horstmann, G. (2003). What Do Facial Expression Convey: Feeling States, Behavioral Intentions, or Action Requests. *Emotion, 3,* 150-166. doi:10.1037/1528-3542.3.2.150

Izuma, K., Saito, D. N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi:10.1016/j.neuron.2008.03.020

Keil, A., & Ihssen, N. (2004). Identification Facilitation for Emotionally Arousing Verbs During the Attentional Blink. *Emotion, 4*, 23-35. doi:10.1037/1528-3542.4.1.23

Kim, H., Choi, M. J., & Jang, I. J. (2012). Lateral OFC Activity Predicts Decision Bias due to First Impressions during Ultimatum Games. *Journal of Cognitive Neuroscience, 24*, 428-439. doi:10.1162/jocn_a_00136

King-Casas, K., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science, 208*, 78-83. doi:10.1126/science.1108062

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social Attention and the Brain. Current Biology, 19, R958-R962. doi:10.1016/j.cub.2009.08.010

Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage, 20*, 1371-1383. doi:10.1016/S1053-8119(03)00393-8

LaFrance, M. & Hecht, M. A. (1995) Why smiles generate leniency. *Personality and Social Psychology Bulletin, 21,* 207-214. doi:10.1177/0146167295213002

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559–3566. doi:10.1016/S0042-6989(01)00102-X

Lau, S. (1982). The effect on smiling on person perception. *Journal of Social Psychology, 117*, 63-67. doi:10.1080/00224545.1982.9713408

Lavater, J. C. (1880). *Essays on physiognomy; for the promotion of the knowledge and the love of mankind.* William Teg: London. Eighteenth Century Collections Online. Gale Group.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior, 28*, 18–27. doi:10.1016/j.evolhumbehav.2006.09.002

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial Attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B, 366*, 1638-1659. doi:10.1098/rstb.2010.0404

Little, A. C., Roberts, S. C., Jones, B. C., & DeBruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in war and peace-time scenarios. *The Quarterly Journal of Experimental Psychology.* doi:10.1080/17470218.2012.677048

Lynch, K. D. (2007). Modeling role enactment: Linking role theory and social cognition. *Journal for the Theory of Social Behaviour, 37*, 379-399. doi: 10.1111/j.1468-5914.2007.00349.x

Masip, J., & Garrido, E. (2001). Is there a Kernel of Truth in Judgments of Deceptiveness? *Anales De Psicologia, 17*, 101-120.

Matthews, G., & Wells, A. (1999). The cognitive science of attention and emotion. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion.* New York: Wiley.

Moretti, L., Dragone, D., & di Pellegrino, G. (2009). Reward and social valuation deficits following ventromedial prefrontal damage. *Journal of Cognitive Neuroscience, 21*, 128-140. doi: 10.1162/jocn.2009.21011

Morris, M. W., & Keltner, D. (2003). How emotions work: The social functions of emotional expression in negotiations. *Research in Organizational Behavior, 22,* 1-50. doi:10.1016/S0191-3085(00)22002-9

Mullen, B., Futrell, D., Stairs, D., Tice, D. M., Dawson, K. E., Riordan, C. A., Kennedy, J. G., Baumeister, R. F., Radloff, C. E., Goethals, G. R. & Rosenfeld, P. (1986). Newscasters facial expressions and voting-behavior of viewers - can a smile elect a president?

Journal of Personality and Social Psychology, 51, 291-295. doi:10.1037/0022-3514.51.2.291

O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology, 14*, 769-776. doi:10.1016/j.conb.2004.10.016

Öhman, A., Lunqvist, D., & Esteves, F. (2001). The Face in the Crowd Revisited: A Threat Advantage With Schematic Stimuli. *Journal of Personality and Social Psychology, 80,* 381-396. doi:10.1037//0022-3514.80.3.381

O'Leary, D. P. (1990). Robust regression computation using iteratively reweighted least squares. *SIAM Journal of Matrix Analysis and Applications, 11*, 466-480. doi:10.1137/0611032

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior, 34*, 83-110. doi:10.1007/s10919-009-0082-1

Olsen, I. R., & Marshuetz, C. (2005). Facial Attractiveness Is Appraised in a Glance. *Emotion, 5*, 498-502. doi:10.1037/1528-3542.5.4.498

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105,* 11087-11092. doi:10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions from Faces. *Emotion, 9*, 128-133. doi:10.1037/a0014520

Otta E., Folladore A. F., Hoshino R. L. (1996). Reading a smiling face: messages conveyed by various forms of smiling. *Perceptual and Motor Skills, 82*, 1111–1121. doi:10.2466/pms.1996.82.3c.1111

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*, 75-92. doi:10.1016/j.neuropsychologia.2006.04.025

Porter, S., England, L., Juodis, M., van Brinke, l., & Wilson, K. (2008). Is the Face a Window to the Soul? Investigation of the Accuracy of Intuitive Judgments of the Trustworthiness of Human Faces. *Canadian Journal of Behavioural Science, 40,* 171 177. doi:10.1037/0008-400X.40.3.171

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decision: the impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 15*, 477-491. doi:10.1080/10683160902926141

Quist, M. C., DeBruine, L. M., Little, A. C., & Jones, B. C. (2012). Integrating social knowledge and physical cues when judging the attractiveness of potential mates. *Journal of Experimental Social Psychology, 48*, 770-773. doi:10.1016/j.jesp.2011.12.018

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992) Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink*? Journal of Experimental Psychology: Human Perception and Performance, 18*, 849-860. doi:10.1037/0096-1523.18.3.849

Raymond, J. E., & O'Brien, J. L. (2009) Selective Visual Attention and Motivation: The Consequences of Value Learning in an Attentional Blink Task. *Psychological Science, 20*, 981-988. doi:10.1111/j.1467-9280.2009.02391.x

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2011). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human* Behavior, 33, 200-209. doi:10.1016/j.evolhumbehav.2011.09.003

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*, 95–112.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior. *PLoS one, 7*, e34293. doi:10.1371/journal.pone.0034293

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation. Neuron, 35, 395-405. doi:10.1016/S0896-9273(02)00755-9

Ritts, V., Patterson, M. L., & Tubbs, M.E. (1992). Expectations, Impressions, and Judgments of Physically Attractive Students: A Review. *Review of Education Research, 62*, 413-426. doi:10.3102/00346543062004413

Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences, 1*4, 64-71. doi:10.1016/j.tics.2009.11.005

Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience, 3*, 1-6. doi:10.3389/neuro.09.016.2009

Rushworth, M. F. S., Mars, R. B., Summerfield, C. (2009). General mechanisms for making decisions? *Current Opinion in Neurobiology, 19*, 75-83. doi:10.1016/j.conb.2009.02.005

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617-640. doi:10.1016/S0167-4870(01)00059-9

Sanfey, A. G. (2009). Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses. *Mind & Society, 8*, 93-107. doi:10.1007/s11299-009-0053-6

Schlicht, E. J., Shimojo, S., Camerer, C. F., Battaglia, P., Nakayama, K. (2010). Human Wagering Behavior Depends on Opponents' Faces. *PLoS one, 5*, e11663. doi:10.1371/journal.pone.0011663

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology, 14*, 139-147. doi:10.1016/j.conb.2004.03.017

Schultz, W. (2006). Behavioural theories and the neurophysiologies of reward. *Annual Review Psychology*, 57, 8-115. doi:10.1146/annurev.psych.56.091103.070229

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593-1599. doi:10.1126/science.275.5306.1593

Shore, D.M., & Heerey, E.A. (2011). The value of genuine and polite smiles. *Emotion, 11*, 169-174. doi:10.1037/a0022601.

Simpson, J. A. (2007). Psychological Foundations of Trust. *Current Directions in Psychological Science, 16,* 264-268. doi:10.1111/j.1467-8721.2007.00517.x

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain Responses to the Acquired Moral Status of Faces. *Neuron, 41*, 653-662. doi:10.1016?S08966273(04)00014-5

Surakka, V., & Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology, 29*, 22-33. doi:10.1016/S0167-8760(97)00088-3

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press.

Tidd, K. L. & Lochard, J. S. (1978) Monetary significance of the affiliative smile: a case for reciprocal altruism. *Bulletin of the Psychonomic Society, 11*, 344-346.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of Competence from Faces Predict Election Outcomes. *Science, 308*, 1623-1626. doi:10.1126/science.1110589

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813–833. doi:10.1521/soco.2009.27.6.813

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evalutation of faces on social dimensions. *Trends in Cognitive Sciences, 12*, 455-460. doi:10.1016/j.tics.2008.10.001

Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science, 312*, 1047-1050. doi: 10.1126/science.1125596

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature, 398*, 704-708. doi:10.1038/19525

Tsukiura, T., & Cabeza, R. (2008). Orbitofrontal and hippocampal contributions to memory for face-name associations: the rewarding power of a smile. *Neuropsychologia, 46*, 2310-2319. doi:10.1016/j.neuropsychologia.2008.03.013

Tversky, A., & Kahneman, D. (1981). The Framing of Decision and the Psychology of Choice. *Science, 211*, 453-458. doi:10.1126/science.7455683

van 't Wout, M., & Sanfey, A.G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*, 796-803. doi:10.1016/j.cognition.2008.07.002

Vickery, T. J., & Jiang, Y. V. (2009). Inferior parietal lobule supports decision making under uncertainty in humans. *Cerebral Cortex, 19*, 916-925. doi:10.1093/cercor/bhn140

von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behaviour* (2nd ed.). Princeton, NJ: Princeton University Press.

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.

Willis, J. and Todorov, A. (2006). First impressions: making up your mind after 100 ms exposure to a face. *Psychological Science, 17*, 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Wilson, R. K., & Eckel, C. C. (2006). Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly, 59*, 189-202. doi:10.1177/106591290605900202

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluations of trustworthiness of faces. *Nature Neuroscience, 5*, 277-283. doi:10.1038/nn816

Yamagashi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover Evidence that cheaters may look different from co-operators. *Evolution and Human Behavior, 24*, 290-301. doi:10.1016/S1090-5138(03)00035-7

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate Social Perception at Zero Acquaintance: The Affordances of a Gibsonian Approach. *Personality and Social Psychology Review, 1,* 204-223. doi:10.1207/s15327957pspr0103_2

Zebrowits, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass, 2*, 1497–1517. doi:10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-Eyed" and "Crooked-Faced": Determinants of Perceived and Real Honesty Across the Life Span. *Personality and Social Psychology Bulletin, 22*, 1258-1269. doi:10.1177/01461672962212006