Title
The Intrinsic Value of Consensus

Author
Mistry, Prachi

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

The Intrinsic Value of Consensus

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Science

by

Prachi Mistry

Dissertation Committee
Assistant Professor Mimi Liljeholm, Chair
Assistant Professor Emre Neftci
Professor Michael Lee

2020

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my committee chair and advisor, Dr. Mimi Liljeholm. The research presented here would not be possible without her guidance, supervision and contribution.

I would also like to express my gratitude to my committee members, Dr. Michael Lee and Dr. Emre Neftci, for their helpful comments in the writing of this dissertation. Additionally, I would like to thank the School of Social Sciences for the Fellowship funding.

Finally, I would like to thank my family and friends for their support.

Vita

Prachi Mistry

2014	B.A. in Intensive Psychology, University of California, Santa Cruz
2014	B.S. in Cognitive Science, University of California, Santa Cruz
2014-19	Graduate Student Researcher, Learning and Decision Neuroscience Laboratory, University of California, Irvine
2014-19	Teaching Assistant, University of California, Irvine
2016	M.S. in Cognitive Neuroscience
2017	Research Assistant, U.S. Army Research Laboratory, Playa Vista, California
2020	Ph.D. in Cognitive Science with Cognitive Neuroscience concentration, University of California, Irvine

FIELD OF STUDY

Cognitive Neuroscience: Learning and Decision Making

PUBLICATIONS

Mistry, P., & Liljeholm, M. (2015). The Role of Outcome Divergence in Goal-Directed Choice. In *CogSci 2015 Proceedings*.

Mistry, P., & Liljeholm, M. (2016). Instrumental Divergence and the Value of Control. *Scientific reports*, *6*, 36295.

Liljeholm, M., Mistry, P., & Koh, S. (2018). The Influence of Schizotypal Traits on the Preference for High Instrumental Divergence. *Cog Sci 2018 Proceedings*.

Mistry, P., & Liljeholm, M. (2018). The Intrinsic Cost of Dissent. In *CogSci 2018 Proceedings*.

Mistry, P., & Liljeholm, M. (2019). The Expression and Transfer of Valence Associated with Social Conformity. *Scientific reports*, *9*(1), 1-12.

Abstract of the Dissertation

The Intrinsic Value of Consensus

By

Prachi Mistry

Doctor of Philosophy in Cognitive Science

University of California, Irvine 2020

Mimi Liljeholm, Chair

Consensus seeking – abandoning one's own judgment to align with a group majority – is a fundamental feature of human social interaction. Notably, such striving for majority affiliation often occurs in the absence of any apparent economic or social gain, suggesting that achieving consensus might have intrinsic value. The current work assessed the expression and transfer of valence associated with social conformity, and the relation between conformity and exploratory behavior. In the first two studies, using a paradigm in which participants assumed the role of a juror evaluating a series of misdemeanor criminal cases, we found that contexts that had been repeatedly paired with consensus decisions were rated as more likable, and selected more frequently in a two-alternative forced choice test, than were contexts paired with dissent from a unanimous majority. The second of these studies ruled out inferences about the accuracy of the majority opinion as the basis for such evaluative changes. A subsequent set of studies employed a simple gambling task, in which the decisions of ostensible previous gamblers were indicated below available options on each trial, to assess the trade-off between social and non-social

currencies, and the transfer of social valence to interpersonal stimuli. In spite of demonstrating near-perfect knowledge of objective reward probabilities, participants reliably preferred gambling options and previous gamblers associated with conformity over those associated with reward. Formally, we found that a reinforcement learner that treated conformity as a surrogate reward provided a better account of choice preferences than did a conventional model. Finally, we investigated the relationship between social conformity and a tendency to explore the environment for potentially greater, yet unknown, rewards. We found that the degree to which participants adjusted their ratings of subjective food preferences to match the aggregate ratings of ostensible previous participants was negatively correlated with exploratory behavior in a multi-armed bandit task. In summary, we provide evidence for a common value-scale for social and non-social currencies, an ability of conforming decisions to imbue concomitant stimuli with affective significance, and a negative relationship between social conformity and reward exploration. By characterizing conformity as reinforcement learning, the framework advanced in this dissertation provides a mechanism for how apparently inconsequential consensus decisions may be motivated by previously acquired valence, as well as for how that valence may in turn be transferred to concomitant stimuli, both contextual and interpersonal. This novel approach will bridge a critical gap, at neural and behavioral levels, between complex socio-cognitive representations and basic mechanisms of reward-based learning and decision-making.

Chapter 1 Introduction

Social animals must often come to a consensus with other members of their group when making collective decisions. Through agreeing with a majority, individual members are able to maintain in-group identity, retain access to group resources and avoid social punishments (Bond and Smith, 1996; Reyson & Branscombe, 2008). While consensus seeking is not surprising in such contexts, individuals also consistently engage in consensus seeking behavior in the absence of apparent reward (Sherif, 1935; Klucharev et al., 2009; Corriveau et al., 2009; Nook & Zaki, 2015; Sun & Yu, 2016). This suggests that the act of reaching consensus may have intrinsic value.

A possible basis of that value is reward learning; according to reinforcement learning theory, actions and stimuli that have a history of being paired with reward will acquire value in their own right (Sutton & Barto, 1981). Prior work has shown that consensus seeking behavior can lead to better monetary rewards and judgement accuracy (Harris et al., 2012; Toyokawa et al., 2014). Furthermore, the idea that reward learning mechanisms may drive consensus seeking behavior is supported by an overlap in neural substrates involved in both reward learning and processing of consensus information (Klucharev et al, 2009; Cambell-Meiklejohn et al., 2010; Zaki et al, 2011; Yu & Sun; 2013; Nook & Zaki, 2015). The current literature, however, does not directly address the possible intrinsic value of consensus. Therefore, the primary goal of this work was to examine the potential role of reward learning in consensus seeking behavior and better understand the cognitive mechanisms driving inconsequential consensus seeking.

1.1 Inconsequential Consensus Seeking

Consensus seeking behavior is often attributed to an attempt to either gain social reward (i.e. external validation and group belonging) or avoid social punishment (i.e. conflict and social isolation) (Bond and Smith, 1996; Reyson & Branscombe, 2008). Similarly, there is a clear incentive for conformity where consensus decisions maximize monetary rewards or accuracy. (Harris et al., 2012; Toyokawa et al., 2014; Gürçay et al., 2014) Individuals, however, consistently engage in consensus seeking behavior even in the absence of monetary or social reward; a phenomenon here referred to as inconsequential consensus seeking.

In a perceptual judgment task, Asch (1951) found that participants conformed to the consensus 32% of the time, despite the consensus judgment being visibly incorrect. Here participants were shown a card with a line on it and were then shown a card with three lines and asked to pick the line that matched the original. During the task there were 8 confederate subjects in the room who would verbally report their answer. Confederate subjects reported the correct answer for the first four trials and unanimously reported a preselected incorrect answer for the remaining eleven trials. Participants who completed the task in the absence of confederates chose the correct answer over 99% of the time, while those in the presence of confederates chose the correct answer 68% of the time. Critically, when present, confederates would not react to any of the answers given. Additionally, the participants did not know or speak to the confederates prior to the experiment. It is, however, still possible that the mere presence of the confederates created some degree of social pressure regardless of their lack of involvement with the participants.

Ruling out the mere presence of confederates as a source of social pressure, Sherif (1935) demonstrated that individuals conform to group judgments even when the group is absent. In Sherif's task, participants were initially asked to report the perceived movement of a stationary spot of light. When participants later repeated the task in groups, responses gradually converged to the group mean estimate over several trials. Critically, when subsequently again performing the task individually, the group mean estimate still influenced individual judgments despite the absence of any social motivation. Importantly, during the initial, individual, phase of Sherif's (1935) task, participants reported some movement (between 2 and 12 inches) in spite of the spot of light being stationary, suggesting that the stimulus properties were ambiguous. It is possible, therefore, that participants deferred to the group mean estimate due to high levels of uncertainty.

Individuals, however, conform even when there is no objectively correct answer. For example, Nook and Zaki (2015) found that participants conformed to a group norm when rating how much they wanted to eat specific food items in the absence of any social or monetary incentive. In this experiment, participants were shown 150 food items and were asked to rate on a scale of 1 to 8 how much they wanted to eat that particular food item. On each trial, after rating a particular food item, participants were shown the ostensible average rating of a large group of peers for that same food item. The group ratings were manipulated by the experimenters such that the group norm matched the participant's rating, was beneath the participant's rating, or was above the participant's rating. Subsequently, participants were again asked to indicate their preference for all 150 food items. Importantly, participants' compensation (monetary or course credit) was entirely independent of their judgments, and their knowledge of ostensible group averages was based solely on numerical displays with no exposure to, or information about, actual individuals.

In other words, no economic or social gain was contingent on reaching consensus. Regardless, participants significantly shifted their follow-up food ratings in the direction of the group norm.

An apparent preference for consensus is also revealed by treatment of individuals previously associated with consensus: When learning new words, children will trust an individual previously associated with consensus rather than a known dissenter. Corriveau et al. (2009) conducted a study in which three and four-year olds were shown unknown objects and a three-person consensus would supply a name for each object, while a lone dissenter supplied a different name. In subsequent trials two members of the consensus group would leave, while the dissenter and one member of the consensus group remained. Subsequently, across a novel set of objects and words, children preferred to endorse information given by the member of the original consensus and remained mistrustful of the dissenter. Like adults, children at this age also defer to an inaccurate consensus, despite being accurate when making independent judgments, suggesting that similar mechanisms might drive inconsequential consensus in children and adults. The fact that information provided by individuals associated with consensus is preferred over information given by a dissenter, in the absence of any other apparent reward, suggests that consensus may be intrinsically rewarding. This intrinsic value may have been acquired through reward learning.

1.2 Reinforcement Learning

Reinforcement learning is a well-documented phenomenon where an individual is more likely to engage in a behavior if that behavior has been previously rewarded (e.g., Hull, 1943; Rescorla & Wagner, 1972; Skinner, 1938). Specifically, when paired with unexpected reward, actions or stimuli acquire value based on the presence of a *reward prediction error* – the discrepancy between expected and obtained reward (Sutton & Barto, 1981). Formally, the *reward prediction error* can be derived using a model-free reinforcement learner. Here an agent estimates the value of each action (a) given the current state (s). Each time an action is taken within a particular state, the state-action value $Q(s,a)$ is updated based on the reward obtained in the following state $r(s')$ and the estimated value of the subsequent state and action, $Q(s',a')$. Specifically, the reward prediction error is defined as:

$$\delta_{RPE} = r(s') + Q(s', a') - Q(s, a) \quad \textbf{(1)}$$

and is used to update a state-action value as:

$$Q(s, a) = Q(s, a) + \alpha\delta_{RPE}, \quad \textbf{(2)}$$

where α is a learning rate (Sutton & Barto, 1998). Thus, based on reinforcement learning mechanisms, actions and stimuli that have a history of being paired with reward acquire value in their own right. The greater the value of an instrumental action based on its reinforcement history, the greater the probability that action has of being subsequently performed. Notably, any type of reward, including social, monetary and primary rewards, can reinforce, and thus increase the value of an action. While primary rewards (i.e. sleep, food, air, water and sex) have innate value and are essential for survival and reproduction, secondary rewards (e.g., money and social praise) gain value through pairings with primary rewards (Skinner, 1938).

According to reinforcement learning theory, if consensus has a history of being paired with reward, whether social, monetary or primary, it will itself acquire value. Previous work has demonstrated that consensus seeking behavior results in greater levels of social approval as well as greater monetary payoffs (Bond & Smith, 1996; Toyokawa et al., 2014). Moreover, consensus judgments yield superior memory retrieval performance and are often objectively more accurate (Sniezak & Henry, 1990; Weldon & Bellinger, 1997; Gürçay et al., 2014; Harris et al., 2012). Finally, an aggregate of disparate estimates across individuals in a group often outperforms that of any given individual suggesting that, even in the absence of a group majority, relying on the mean of the group might optimize accuracy and performance (Surowiecki, 2005). Therefore, consensus behavior may acquire value through a reinforcement history increasing the probability of actions leading to consensus being performed, even when there is no immediate opportunity to earn reward.

The notion that reward learning mechanisms may drive consensus seeking behavior is further supported by an overlap in neural substrates involved in both reward learning and the processing of consensus information, particularly with respect to the discrepancy between individual and group judgements. Brain regions frequently implicated in the coding and assessment of reward, including the ventromedial prefrontal (vmPFC), orbitofrontal cortex (OFC), and ventral striatum (VS) (e.g., O'Doherty et al., 2004; Hare et al., 2008; Oyama et al., 2010; Asaad & Eskandar, 2011), have also been identified in the processing of social information, particularly with respect to the discrepancy between individual and group judgments (e.g., Zaki et al., 2011; Nook & Zaki, 2015). For example, Nook and Zaki (2015) found that ventral striatum (VS) activity was greater when a participant's judgment about the likability of a particular food was in agreement with a mean group judgment than when in disagreement, and this VS activity also predicted subsequent conformity with group judgments during follow-up ratings. Moreover, during follow-up ratings, ventromedial prefrontal cortex (vmPFC) activity scaled with the signed difference between an individual's original judgment and the group mean, suggesting that these areas may have contributed to adjustments towards the group mean. Similarly, using facial attractiveness ratings, Zaki et al. (2011) found significantly greater activity in the VS and orbitofrontal cortex (OFC) when participants were re-rating faces for which a group had previously provided an attractiveness rating that was higher than the participant's original rating, as compared to a lower rating. Thus, structures frequently implicated in the anticipation and receipt of reward (e.g., Kim et al., 2010; O'Doherty et al., 2006) appear to be involved in the processing of, and adjustment relative to, information provided by a consensus; this suggests that the same neural mechanisms may be driving both processes.

The fronto-striatal areas discussed above are part of the mesolimbic and mesocortial dopaminergic system, which originates in the ventral tegmental area and innervates the ventral striatum and prefrontal cortex. Within this system, the phasic firing of dopamine neurons has been implicated in processing rewarding (see Schultz et al., 1997 for review). Pessiglione et al. (2006) found that individuals treated with drugs that enhance dopaminergic function were more likely to select rewarding actions than those who were treated with drugs that reduce dopaminergic activity. Likewise, in modifying extracellular concentration of dopamine, Campbell-Meiklejohn et al. (2012) were able to influence the degree of conformity individuals exhibited. When rating trustworthiness of faces, participants who were administered a dose of Methylphenidate, a reuptake inhibitor that increases extracellular levels of dopamine levels in the striatum, exhibited twice the conformity as those who had received a placebo pill. These psychopharmacological findings suggest that the dopaminergic system may be causally involved in both reward learning and consensus seeking behavior.

The behavioral and neural evidence, however, does not directly address the possible intrinsic value of consensus. Specifically, it is problematic to conclude that consensus seeking is rewarding simply because individuals conform. Convergence towards group means may arise from an effort to approximate accuracy through the minimization of expectation violations rather than the value of consensus itself. Moreover, the apparent involvement of brain regions frequently implicated in reward processing does not warrant the reverse inference that conforming decisions have a hedonic component. Those same neural regions respond to valence neutral but surprising, or otherwise salient, stimuli (Horvitz, 2000; Zink et al., 2003; Jensen et al., 2007; Cooper & Knutson, 2008). Additionally, the neural activity shown in previous studies of conformity seems to reflect an error monitoring system rather than the rewarding aspect of consensus. For example, Zaki et al., found that during subsequent re-exposure to rated stimuli the activity in the VS, as well as in the medial orbitofrontal cortex, scaled with the signed difference between the participant's rating and the group norm. Such signed signals may reflect retrieval of previously experienced divergence from the group or error-adjusted stimulus values but are inconsistent with a reinforcement signal encoding the hedonic valence of majority alignment. If consensus has intrinsic value, there should be an increase in response to stimuli paired with the positive hedonics of conforming decisions and decrease in response to stimuli associated with the aversiveness of dissent. The approach discussed in Aims 1 and 2 directly examines the intrinsic value of consensus and allows us to evaluate reward learning as a possible mechanism driving inconsequential consensus through the transfer of valence to concomitant stimuli.

1.3 Summary

In summary, evidence that consensus on average yields greater reward suggests that a reinforcement learning mechanism might be responsible for consensus behavior in the absence of apparent immediate reward, a possibility that is further strengthened by the overlap in neural substrates mediating both reward learning and consensus decision making. Notably, if the same prediction error mechanism that is used to adjust reward expectations is also engaged during evaluation of social information, a similarly implicit adjustment (i.e., error-reduction mechanism) might drive an individuals' behavior towards consensus decisions. In the following chapters we begin with an investigation of the intrinsic value of consensus through its ability to induce conditioned reinforcement by assessing the degree to which neutral stimuli paired with consensus acquire motivational significance and behavioral sensitivity to such stimuli and examine how information about accuracy can influence consensus seeking behavior. We then perform an assessment of the trade-off between consensus and monetary rewards. Finally, we examine how individual differences in consensus seeking may relate to differences in reward processing.

Chapter 2 The Intrinsic Cost of Dissent

2.1 Abstract

Consensus seeking is a fundamental feature of human social interaction. Notably, such striving for majority affiliation often occurs in the absence of any apparent economic or social gain, suggesting that achieving consensus might have intrinsic value. Here, we examine the affective properties of consensus decisions by assessing the transfer of valence to concomitant stimuli. Specifically, in two studies, we show that contexts repeatedly paired with consensus decisions are rated as more likable and selected more frequently in a two-alternative forced choice test, than are contexts repeatedly paired with dissent from a unanimous majority. In the second study, we rule out inferences about the accuracy of the majority opinion as the basis for such evaluative changes. Our results suggest that an intrinsic value of consensus, or cost of dissent, may motivate and reinforce social conformity. (Published in *CogSci 2018 Conference Proceedings*)

2.2 Introduction

Individuals consistently conform to a group majority in the absence of any social or monetary reward, which suggests that the act of reaching a consensus may be itself be rewarding (e.g., Sherif, 1936; Klucharev et al., 2009; Corriveau et al., 2009; Chen et al., 2013; Nook & Zaki, 2015). According to reinforcement learning theory, actions and stimuli that have a history of being paired with reward will acquire value in their own right (Sutton & Barto, 1990). Thus, if consensus has a history of being paired with reward, it should itself acquire value. Prior studies have shown that relying on a group majority often yields superior memory retrieval, judgment accuracy and monetary payoffs (Harris et al., 2012; Gürçay et al., 2014; Toyokawa et al., 2014). Furthermore, by agreeing with a majority, individuals are able to avoid social rejection and retain access to group resources (Bond & Smith, 1996; Reysen & Branscombe, 2008). Taken together, this suggests that consensus decisions may often lead to better outcomes. Also consistent with the notion that consensus decisions serve as a positive reinforcement signal, recent neuroimaging work has demonstrated an overlap between neural substrates mediating conformity and those involved in processing reward (Klucharev et al., 2009; Zaki et al., 2011; Campbell-Meiklejohn et al., 2012; Yu & Sun, 2013; Nook & Zaki, 2015).

Still it is problematic to conclude that consensus-seeking decisions are rewarding from the existing behavioral and neural evidence. Error-based adjustments towards a reference, such as a majority opinion, need not be associated with valence but may simply reflect an effort to approximate accuracy by minimizing expectation violations. Moreover, the apparent involvement of brain regions frequently implicated in reward processing does not warrant the reverse inference that consensus-seeking decisions have a hedonic component since those same regions also respond to valence neutral salient stimuli. There is a clear need, thus, for studies that employ independent measures of the valence associated with consensus and dissent.

Some social psychology studies have used evaluative measures to assess emotional constructs associated with dissent from group opinions. For example, Matz and Wood (2005) used an emotion measure to assess dissonance discomfort, negative self-evaluation and positive feelings associated with agreeing or disagreeing with a group of ostensible peers in a mock jury. They found that participants who disagreed with the group experienced significantly greater dissonance discomfort than those who agreed, especially if they believed that they would be required to discuss their opinions or reach consensus with other jury members. No such effects were found for measures of negative self-evaluation and positive feelings; however, in a subsequent study, positive feelings increased, and negative self-evaluation decreased when participants were given the opportunity to achieve consensus by persuading others or joining a more congenial group. This and related work suggest that some form of valence does accompany decisions made relative to a group norm. However, lacking a formal framework of reward-based behavior, the approach is poorly suited to quantify hedonic aspects of social conformity.

To address these limitations, we have developed a novel paradigm that tests the hypothesis that social conformity has intrinsic value by assessing the degree to which that value is transferred to contextual stimuli. A fundamental property of stimuli that possess intrinsic value is their ability to transfer that valence to arbitrary neutral stimuli – a phenomenon known as conditioned reinforcement. Formally, this value transfer can be estimated using Temporal Difference (TD) learning - a type of RL in which states that predict value acquire value through a time shifted reward prediction error signal (Sutton & Barto, 1990). Conditioned reinforcers have a strong influence on behavioral choice and are often viewed as goals themselves. If intrinsically valuable, consensus should be able to induce conditioned reinforcement in associated arbitrary stimuli.

Following Matz and Wood (2005), we employ a “mock jury” scenario to generate majority judgments with which a participant may agree or disagree. Here we attempt to show that contexts paired with high levels of consensus will be subsequently selected in a two alternative forced choice task, over those paired with dissent. Additionally, we used subjective ratings of likability to demonstrate an increase in appeal for contexts associated with consensus over dissent. In the following experiments, this was done by assessing how discrepancy between participants' own judgements and those of a unanimous jury influenced the likability of, and preference for distinctly colored courtrooms.

2.3 Experiment 1

A basic prediction of RL theory is that if consensus has intrinsic value, then this value should transfer to arbitrary stimuli associated with high levels of consensus. This prediction was tested by assessing how the congruence between participants' own judgments and those of a unanimous jury influenced the likability of, and preference for, distinctly colored courtrooms.

2.3.1 Methods

Participants: Twenty undergraduates at the University of California, Irvine (13 females, mean age = 19.6) participated in the study for course credit. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine, approved the study.

Task and Procedure*:* At the start of the experiment, participants were told that they would be acting as juror in a series of cases in various courtrooms. They were further told that to prepare for making decisions on the cases themselves, they would first have the opportunity to study previously adjudicated cases (henceforth referred to as the learning phase). All cases were potential violations of the California Vehicle Code, with a maximum penalty of 6 months of incarceration. The cases were constructed such that all defendants had violated the California Vehicle code but would not necessarily be found liable for the infraction (e.g. driving five miles per hour above the legal speed limit).

Before starting the learning phase, participants were asked to rate the likability of four differently colored courtrooms, in random order, on a scale from 0 (not at all likable) to 10 (extremely likable), with 5 indicating neutral affect. On each trial in the learning phase (shown in Figure 1), participants were first presented with a short synopsis describing the particular case while one of the four courtrooms were displayed in the background indicating that the case was heard in that room. They were then asked to press the left or right arrow key to indicate whether they believe the individual to be guilty or not. A grey avatar representing the participant would move beneath the corresponding "guilty" or "not guilty", label based on the participant's response. They would then be prompted to press the spacebar to see the jury's verdict, which was represented by five darker grey icons appearing beneath the relevant label. In two of the courtrooms (consensus rooms), the verdict of the jury was the same as that of the participant ~90% of the time and in the other two courtrooms (dissention rooms) the verdict of the jury will be the opposite of the participant's ~90% of the time. The colors of consensus and dissention rooms were counterbalanced across participants.

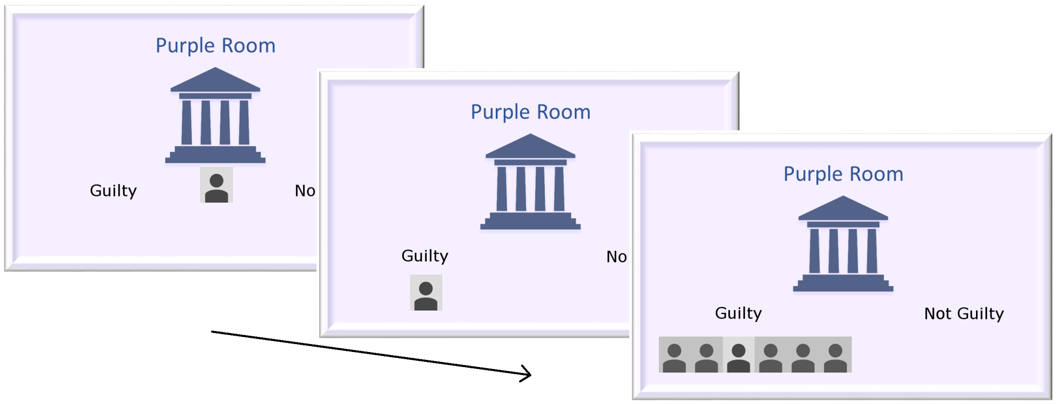

Figure 2.1. Trial illustration of the learning phase showing the initial choice screen, the participant's culpability choice, and the verdict of a unanimous jury together with the choice of the participant. Case summaries and response prompts are not shown in the figure.

Following the learning phase, participants were again asked to rate the likability of the four courtrooms before being moving on to a second phase, in which they served as jury members themselves.

In the second phase, participants were instructed that none of the previously observed jurors would serve on any juries of which the participant might be a member. Participants first selected between two courtrooms: one consensus room, in which the previous juries had frequently agreed, and one dissention room, in which the previous juries had frequently disagreed. Once entering the chosen courtroom, they were presented with a case and asked to indicate whether they believed the defendant to be guilty or not guilty. To assess explicit memory of consensus and dissention courtrooms, at the end of the experiment participants were asked to rate the degree to which the jury agreed with them in each of the differently colored courtrooms during the initial learning phase, on a scale of 0 (never) to 10 (always).

2.3.2 Results

Likability ratings: We predicted that participants would rate courtrooms where participants frequently agreed with the jury (consensus rooms) as more likable than courtrooms in which participants frequently disagreed with the jury (dissent rooms). A planned comparison revealed that this was indeed the case: subtracting the baseline (pre-learning) ratings for each courtroom, the mean rated likeability of rooms associated with consensus was significantly greater than that of rooms associated with dissent, $t(19)=2.96$, $p=0.008$, $d=0.139$ $BF_{10} = 6.13$, error percentage = 0.001. Notably, this difference was due to a decreased likeability of the dissention rooms rather than an increase in likability of the consensus rooms. There was also a significant difference in mean pre- and post-learning likability ratings for dissention rooms (-0.75 ± 1.51), $t(19)=2.22$, $p=0.04$, $d=1.509$ $BF_{10} = 1.71$, error percentage = 0.003, but not for consensus rooms (0.18 ± 1.45), $t(19)=0.54$, $p=0.60$, $d=1.45$ $BF_{01} = 3.78$, error percentage = 0.021. Mean (post-pre) likability ratings are shown on the left side of Figure 2.

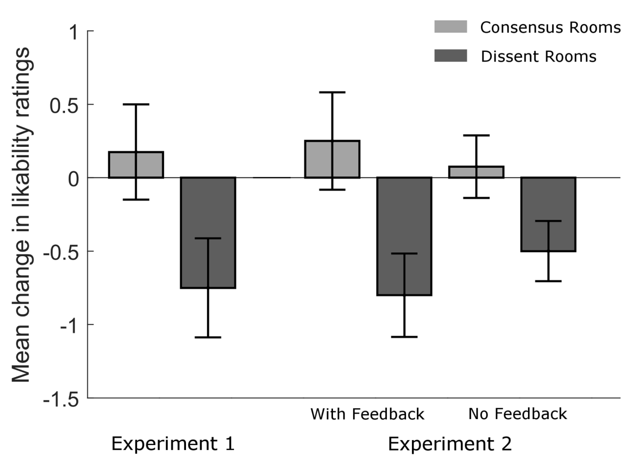

Figure 2.2. Likability ratings (post-pre learning) for consensus and dissent courtrooms from Experiment 1 (left) and from the two groups in Experiment 2 (right).

Choice Preference: We further hypothesized that there would be a significant preference for deliberating in consensus rooms over dissention rooms, despite instructions that emphasized that none of the jurors that had been present during the learning phase would serve with the participant during this phase of the experiment. Consistent with this prediction, we found that, when asked to select a room in which to serve on a jury, the mean proportion of consensus room choices was 65%, which was significantly greater than chance, $t(19)=2.48$, $p=0.02$, $d=0.19$ $BF_{10} = 2.62$, error percentage = 0.003.

Explicit recall of consensus and dissent: Finally, we confirmed that participants were able to accurately distinguish between consensus (5.73 ± 1.45) and dissention (4.60 ± 1.52) rooms, $t(19)=2.68$ $p=.01$, $d=1.88$ $BF_{10} = 3.70$, error percentage = 0.002. However, critically, the degree to which participants discriminated between consensus and dissention rooms was not correlated with the degree to which likability ratings differed across rooms (i.e., differences in consensus ratings across rooms did not predict differences in likability ratings across rooms), Pearson's $r=-0.08$, $p=0.73$ $BF_{01} = 3.42$.

2.4 Experiment 2

We interpret the effects of Experiment 1 in terms of an *intrinsic aversive property* of dissent that is transferred to contextual stimuli through reinforcement learning mechanisms. However, an alternative possibility is that participants were using the unanimous jury as a substitute for information regarding accuracy, and that the transferred negative valence was elicited by the perception of being wrong rather than by dissent. In Experiment 2, we addressed this possibility by including feedback about the "true" culpability of the defendants in our hypothetical court cases.

2.4.1 Methods

Participants: Forty undergraduates at the University of California, Irvine (21 females, mean age = 19.75) participated in the study for course credit. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine, approved the study.

Task and Procedure: There were two groups in this experiment. For the first (No Feedback) group, the task and procedures were identical to those in Experiment 1a. For the second (Feedback) group, the task and procedures were identical to those of the "No Feedback" group, with the following exceptions: First, at the end of each trial in the initial learning phase, participants were asked to press the space bar to view the *actual* culpability of the defendant in that particular case. On the culpability feedback screen, the icons representing the jury were removed, and a selection square appeared around the label indicating the actual culpability of the defendant. The culpability feedback was such that the participant's judgment was correct 50% of the time, in both consensus and dissention rooms. Thus, both types of rooms were equally associated with being wrong. Second, at the end of the study, participants in both groups were also asked to complete a written explanation regarding the reasons for their likability ratings.

2.4.2 Results

The results in both groups closely replicated those of Experiment 1: A 2-by-2 mixed analysis of variance (ANOVA) performed on the post-pre likability ratings, with feedback and consensus as between- and within-subjects factors respectively, revealed a significant main effect of consensus, $F(1,38)=9.73$, $p<.005$ $BF_{10} = 29.84$, error percentage = 1.03), but no effect of feedback (i.e., group) $BF_{01} = 3.63$, error percentage = 0.82, and no interaction, F's<0.84. Comparing the strength of the Bayes factor for a model that includes the interaction term against the null model including both feedback and consensus in all models (nuisance variables) yielded $BF_{01} = 2.21$ (error percentage = 2.45, weak evidence for the null hypothesis). As can be seen in Figure 2, the post-learning difference in likability between consensus and dissention rooms was again due to a decreased liking of dissention rooms, in both groups. Likewise, planned comparisons again revealed a preference for deliberating in consensus over dissention rooms that was significantly greater than chance, in both the "feedback" (62%, $p=0.04$, $BF_{10} = 1.64$ error percentage = 0.003) and "no feedback" (68%, $p=0.002$, $BF_{10} = 17.19$ error percentage = 0.0003) group.

While mean ratings of how often their judgment had agreed with that of the juries in a particular room during the learning phase were again greater for consensus (5.49 ± 1.33) than for dissention rooms (4.94 ± 1.19), this difference did not reach significance in either group, $p_{feedback}=0.52$, $p_{nofeedback}= 0.13$, $BF_{01\ feedback} = 3.54$ error percentage = 0.021, $BF_{01\ no\text{-}feedback} = 1.46$ error percentage = 0.004, nor did the degree of discrimination between consensus and dissention rooms significantly predict changes in likability ratings, in either group, $p_{feedback}=0.32$, $p_{nofeedback}= 0.15$, $BF_{01\ feedback} = 2.27$, $BF_{01\ no\text{-}feedback} = 1.40$. When asked, at the end of the study, about the basis for their likability ratings, only 15% of participants, 3 in each group, cited their consensus with the jury; importantly, in spite of the reduction in power, differences in likability ratings as well as choice preferences remained significant, in each group, when those participants were excluded (*p's*<0.05). The majority of participants, 53%, attributed their ratings to the (counterbalanced) colors of the rooms, while the remaining participants cited various reasons, including the specific cases presented in a particular room (13%), or simply a general "feeling" about the room (10%).

2.5 Discussion

In two experiments, we investigated the affective properties of agreeing or disagreeing with a unanimous majority, by measuring the transfer of valence to concomitant stimuli. We found that contexts repeatedly paired with consensus decisions were rated as more likable and selected more frequently in a two-alternative forced choice test than contexts repeatedly paired with dissent. In the second study, these evaluative differences emerged even when participants received explicit feedback regarding the "correct" answer. This suggests that the valence associated with agreement or dissent was not solely due to perceived accuracy. Notably, across studies, evaluative changes were driven by a decreased likeability of contexts paired with dissent rather than an increased likability of contexts paired with consensus. It is possible that this pattern of results reflects a general, exposure-based, decrease in the likability of all stimuli from which an association with consensus offered protection. However, given that we did not observe a neutral condition where participants completed the experiment without group feedback, we tentatively interpret our findings as evidence for an intrinsic cost of dissent.

There are several possible sources of negative affect associated with dissent. From a reinforcement learning perspective, the act of dissenting from a group majority may have acquired negative valence through a history of being paired with aversive outcomes (e.g., social rejection, losing access to group resources, inferior perceptual and economic decisions). Alternatively, the negative valence may not be directly related to dissent, but instead accompany more general processes. For example, a lack of consensus has been proposed to illicit cognitive dissonance—a feeling of discomfort induced by interpersonal or intrapersonal discrepancy (Festinger, 1957; Matz and Wood, 2005; Klucharev et al., 2009, 2011; Shestakova et al., 2013). Matz and Wood (2005) found that individuals who belonged to groups that were able to reach a consensus experienced reduced discomfort compared to individuals whose groups were not able to reach a consensus, whether consensus was reached through persuading other group members, yielding to other group members, or switching into a group that shared the individual's opinion. Additionally, at the neural level, the ACC and the posterior medial frontal cortex (pMFC), which closely borders and partially encompasses the ACC, have been implicated in both cognitive dissonance and consensus seeking (van Veen et al., 2009; Izuma et al., 2010; Falk et al., 2012; Kitayama et al., 2013).

Similarly, several studies have shown that conforming to a consensus reduces perceived uncertainty about decision outcomes which could indicate that the negative affect associated with dissenting is related to uncertainty aversion (McGarty et al., 1993; Smith et al., 2007; Petrocelli et al. 2007; Sherif & Harvey,1952). Uncertainty aversion is a well-documented phenomenon according to which, in choosing between a more certain moderately rewarding outcome and a less certain, highly rewarding outcome, individuals on average prefer the more certain outcome even when expected value is equal between the gambles (Kahneman and Tversky, 1979). Indeed, Sherif and Harvey (1952) showed an increase in certainty ratings regarding perceptual judgments when participant's decisions converged to the group mean estimates. Likewise, when manipulating group agreement with an individual's judgment about ambiguous stimuli, Mcgarty (1992) found that certainty regarding one's judgment increased when the judgment was consistent with that of the group and decreased when the judgment was inconsistent with that of the group. Moreover, the ACC has been implicated in encoding uncertainty and consensus information. Finally, negative emotions accompanying dissent may reflect inferences about the inaccuracy of one's own judgments in the face of an opposing view. Although this basis for changes in valence was largely ruled out in Experiment 2, in which explicit feedback regarding accuracy was provided on each trial, it seems plausible that, under some circumstances, perceived accuracy may modulate affective responses to dissent.

Of course, whether an aversive quality of dissent is induced by dissonance, uncertainty or a desire to be right, RL mechanisms may still be responsible for transferring that valence to actions and stimuli associated with dissent, as suggested by the current results. An important aspect of our effects, particularly from an RL perspective, is that they are apparently implicit in nature. When asked, most participants attributed the likability of contexts to their colors, and there was no correlation between memory of which courtroom had been paired with dissent and changes in courtroom likability (see Nisbett & Wilson, 1977 for a discussion on when participants may identify experimental manipulations as influential stimuli). This lack of correspondence between explicit recall of which contexts were paired with dissent and decreases in the likability of those contexts suggests that evaluative changes occurred on each trial, as the unanimous majority opinion was revealed, rather than through a retrieval of consensus information at the time that contexts were rated. Such an incremental, trial-by-trial, adjustment in value is consistent with a model-free RL algorithm, in which changes in value coincide with, and are proportional to, the discrepancy between expected and experienced reward. It is also consistent with a previously demonstrated overlap between neural regions implicated in model-free RL and those involved in detecting dissent from a majority opinion (e.g., Klucharev et al., 2009; Zaki et al., 2011; Nook & Zaki, 2015).

In summary, we found that contexts repeatedly paired with dissent from a unanimous majority were less likable, and less preferred in a two-alternative forced choice task, than contexts paired with consensus. Notably, these evaluative changes were driven by a decrease in likability of contexts paired with dissent rather than an increase in likability of contexts paired with consensus. Additionally, the evaluative changes were not predicted by explicit recall of which contexts had been paired with dissent and emerged in spite of explicit feedback regarding the accuracy of judgments. Our results suggest that an intrinsic cost of dissent may motivate and reinforce social conformity.

Chapter 3. Tradeoff Between Motivational Rewards and Consensus

3.1 Abstract

Having established that there is some valence attributed to consensus, or to dissent, our goal was to create a paradigm that would allow us to pit that value against conventional reward, such as money. Here, using a simple gambling task in which the decisions of ostensible previous gamblers were indicated on each trial, we examined the affective properties of agreeing with a group majority by assessing the trade-off between social and non-social currencies, and the transfer of social valence to concomitant stimuli. In spite of demonstrating near-perfect knowledge of objective reward probabilities, participant's choices and evaluative judgments reflected a reliable preference for conformity, consistent with the hypothesized value of social alignment. (Published in *Scientific Reports* at https://doi.org/10.1038/s41598-019-38560-4)

3.2 Introduction

In the previous chapter, we discussed two experiments that indicate that there is some valence ascribed to consensus. This suggests that the intrinsic value of consensus may drive and reinforce consensus seeking behavior. In neuroeconomic literature, the value of a specific reinforcer is frequently quantified by measuring how much individuals are willing to pay to obtain it (Hare et al., 2008; Chib et al., 2009; Peters & Büchel, 2010; Khaw et al., 2015). Specifically, the more money individuals are willing to pay to obtain a reinforcer, the more value that specific reinforcer has. While the phenomenon of apparently inconsequential conformity has been demonstrated across social psychology and neuroscience literatures, and while the notion that individuals conform to social norms at considerable personal cost is well rooted in evolutionary and social sciences, no previous study has, to our knowledge, provided direct experimental and formal evidence for a willingness to pay a price in order to conform (Sherif, 1935; Klucharev et al., 2009; Nook & Zaki, 2015; Gavrilets & Richerson, 2017). In Experiment 3, we directly pit the value of conforming against an alternative incentive, exploring a trade-off between social and non-social currencies. In Experiment 4, using a *conditioned reinforcement* procedure, we further probe the affective valence of majority alignment by assessing the transfer of such valence to concomitant stimuli.

3.3 Experiment 3

As discussed earlier, individuals often conform in the absence of any apparent social or economic gain. It is unclear, however, whether conformity also occurs in the face of conspicuous loss. Critically, in the neuroeconomic literature, the price that a participant is willing to pay for a commodity is a common measure of its value. In Experiment 3, the decision to conform often came at a price. Specifically, participants chose between gambling options that differed in terms of the probability of a fictitious monetary reward (henceforth the "pay-off"), given an ostensible majority endorsement by previous gamblers of the option associated with either a smaller or larger pay-off.

3.3.1 Methods

Participants: Thirty undergraduates at the University of California, Irvine (19 females, mean age = 20.70 ± 2.56) participated in the study for course credit. The sample size was determined through a post hoc power analysis of data from a pilot study, indicating that 26 subjects were required for a power of 90% given a 0.05 threshold for statistical significance ($d = 0.67$). All participants gave informed consent and the Institutional Review Board of the University of California, Irvine, approved the study. All aspects of the study conformed to the guidelines of the 2013 WMA Declaration of Helsinki.

Task and Procedure: At the start of the experiment, participants were instructed that would be playing a game in which, on each trial, they would be asked to select between two slot options, out of six available, on a gambling board. Each of the six numbered slots yielded a fictitious reward of $1 with some probability. Although informed that the monetary rewards were fictitious, participants were asked to treat the rewards as real and attempt to make as much money as possible. Participants were further told that, while they would receive initial training on the probability of reward for each slot, they would not be shown any monetary outcomes when actually gambling. They would, however, be shown which option “previous gamblers” had selected when given the same slot options, before choosing themselves on each gambling trial. The group of previous gamblers was stated to have been drawn from a cohort of students participating in the study during the previous academic quarter. Thus, participants made their gambling decisions with both knowledge of the expected pay-off associated with each available option and knowledge of their peers’ decisions when choosing between the same slot options.

Learning Phase: First, participants were trained to criterion on the probabilities with which each slot yielded the hypothetical $1 reward. To ensure equal sampling of, each slot was highlighted on 10 consecutive trials, indicating the availability of that slot only. When the participant pressed a key to select the highlighted slot, an image of a one-dollar bill or a red cross was displayed on the surface of the slot to indicate, respectively, whether or not the $1 reward was delivered. Following 10 trials with a given slot, participants were asked to rate the probability of reward for that slot. If they did not report the probability within 0.2 of the programmed probability, they had to repeat another 10 trials on that slot. After being trained on, and rating, each individual slot, participants were asked to rate the probability of reward for all slots. If they did not rate the probability of each slot within 0.2 of its programmed probability, they were required to repeat the entire pre-training phase. At the end of the experiment, to assess retention, participants again rated the probability of reward for each slot.

Gambling Phase: Before starting, participants were again instructed that their goal was to make as much money as possible. On each trial, participants were asked to select between two available slot options (See Figure 3). Available slot options were highlighted, and their corresponding slot numbers were printed at the top on the left and right side of the screen. The decisions of ostensible previous gamblers were indicated by grey icons. Specifically, each icon would line up under the slot number that the particular ostensible gambler had chosen when given the same slot options. Previous gambler icons were split across available options such that a randomly determined 5-6 out of 6 majority endorsed one of the two available slots. At the beginning of each trial, an icon representing the participant was presented at the top center of the screen. Once the participant pressed the left or right arrow key to select a slot option, the icon would move beneath the number representing that particular slot option, aligning itself with any previous gamblers already displayed beneath that number. To avoid additional learning during the gambling phase, participants were instructed that total earnings would not be revealed until the end of the experiment, but that they should assume that all outcomes were consistent with the reward probabilities established in the initial training phase.

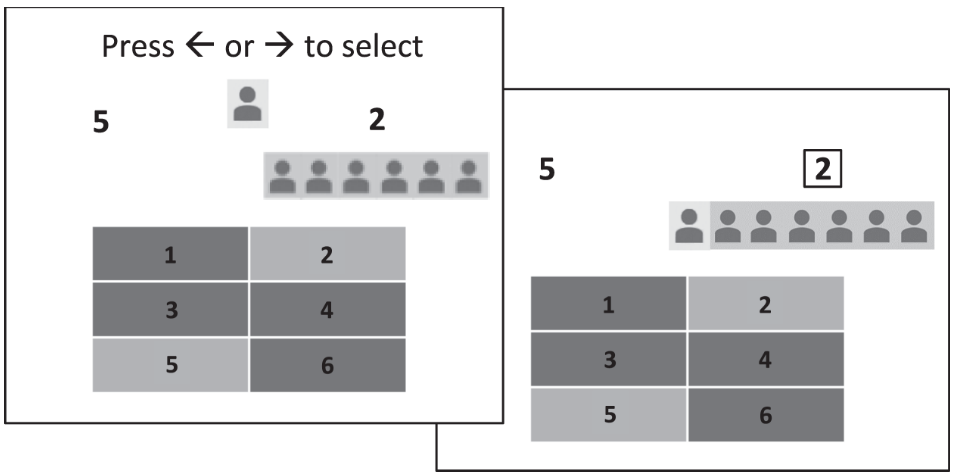

Figure 3.1. Choice and feedback screens on a trial in the gambling phase of Experiment 3. Participants pressed the left or right arrow key to indicate their choice of one of two available slot options, highlighted on the game board and displayed, respectively, to the right and left of the participant's avatar. Following the selection, the participant's avatar moved below the selected option, flanking a set of additional avatars that indicated the decisions of a group of previous players given the same options.

Two of the six game board slots had a 0.2 probability of reward, two had a 0.5 probability of reward, and two had a 0.8 probability of reward. Thus, on any given trial, there was either no difference, a small ($0.30) difference, or a large ($0.60) difference in expected pay-offs between the two options available on that trial. Moreover, when there was a difference, the majority of previous gamblers were shown to have selected the option with a lesser pay-off on half of the trials. Specifically, there were 24 trials with a small difference in expected pay-offs (on 12 of these the majority of previous gamblers chose the option with a lesser pay-off and on the remaining 12 they chose the option with a greater pay-off) and 24 trials with a large difference in expected pay-offs (again, the majority chose the option with a lesser pay-off on half of those trials and that with a greater pay-off on the other half). Finally, 12 trials were included in which the expected monetary pay-off was the same for both options, yielding a total of 60 trials. The order of trials was random, with the constraints that neither a majority-endorsed option nor a greater expected value option could appear on the same side of the screen on more than 3 consecutive trials, and that a particular trial type could not occur on more than 3 consecutive trials.

Debriefing: After completing the experiment, participants were informed that the decisions made by "previous gamblers" had been generated by a computer algorithm and were given the option to withdraw their data from the study in light of this information. Participants were also instructed that they could contact the experimenter if they had any questions or concerns regarding the experiment or their participation. All participants gave written consent to having their data included in the study after learning of the deception.

Computational Model: Gambling decisions were formalized using a model-based reinforcement learner (Doya, 2002) that maintains separate representations of transition probabilities and rewards. Transition probabilities and rewards are dynamically combined to yield decision values, DV, according to the following equation:

$$DV(s,a) = \sum_{s'} T(s.a.s')R(s'), \tag{3}$$

where a is the selection of a particular slot option, $T(s,a,s')$ is the probability of transitioning into a particular outcome state, s', from the current state, s, given a, and $R(s')$ is the reward associated with s'. On each trial, each available slot option was associated with four possible outcome states: receiving \$1 and agreeing with the decision of previous gamblers, receiving \$1 and disagreeing with the decision of previous gamblers, receiving \$0 and agreeing with the decision of previous gamblers, and receiving \$0 and disagreeing with the decision of previous gamblers. Since probabilities of monetary reward were trained to criterion prior to the gambling phase, and the probability of group affiliation could be deduced from the choice screen (see Figure 3), $T(s,a,s')$ was initialized to, and maintained at, the true transition function for the events being modeled.

We implemented two versions of the model: a non-social and a social version. In the non-social model the reward associated with a particular response contingent state, $R(s')$, was simply defined as the expected monetary pay-off of that state. To model the intrinsic value of conformity, an alternative social model was specified, in which the consensus associated with a state served as a surrogate reward, such that

$$R(s') = m(s') + wc(s'), \tag{4}$$

where $m(s)$ is the monetary amount associated with a trial outcome, $c(s)$ is the proportion of previous gamblers associated with an option, and w is a free parameter reflecting individual differences in the monetary value of conformity. For example, if a participant shows no preference between an option for which the probability of \$1 is 0.2 and $c(s)$ is 1.0, versus an option for which the probability of \$1 is 0.8 and $c(s)$ is 0.0, then for that particular participant, w should equal around \$0.6.

Both models assumed that participants select actions stochastically using probabilities generated by a softmax distribution, in which a free "inverse temperature" parameter, τ controls the degree to which choices are biased toward the highest valued action. More generally, the softmax rule is used to introduce noise and is one way of modeling exploration where exploratory decisions and selection of low value choices are determined probabilistically based on actions' relative expected values (Sutton & Barto, 1998). Thus, on each trial in the gambling phase, expected values were computed for the two available gambling options, using Equations 3 and 4, and the softmax rule was used to transform those values into choice probabilities, plotted, for each conformity and pay-off condition, in the left and middle panel of Figure 4. Free parameters were fit to behavioral data by minimizing the negative log-likelihood of obtained choices for each individual using MATLAB's fminsearchbnd function (MathWorks, 2017b), with upper-lower bounds of 0.01–1.01 for w (since the largest possible pay-off on a given trial was \$1.00) and 0.01–100.00 for τ.

The corrected Akaike information criterion (AICc) was used to select between models. We specifically chose the AICc over other information criterion because of our small sample size. The AICc is defined according to the following equation:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad \textbf{(5)}$$

The AICc provides a stronger penalty for more complex models for small sample sizes than the AIC. The number of parameters that can be reliably estimated from finite data changes with sample size (Burnham & Anderson, 2002). Although the AIC accounts for this, the AIC often overfits when the sample size divided by the number of parameters is less than about forty

(Burnham & Anderson, 2004). The BIC target model does not depend on sample size and BIC-selected models can be biased at small sample sizes as an estimator of their target model (Burnham & Anderson, 2004).

3.3.2 Results

We confirmed that participants retained accurate representations of reward probabilities, with 97% of estimates falling within 0.1 of, and 92% of estimates being identical to programmed probabilities. Additionally, we confirmed that participants were incentivized by the hypothetical monetary payoffs: Collapsing across social and pay-off conditions, whenever payoffs differed across available slot options, participants chose the option with a greater payoff 75% of the time, significantly more often than chance, p=1.13 e-7 BF_{10} = 134212.31 error percentage =3.75 e-8. Revealing a clear modulation of this preference, a two-by-two ANOVA performed on the proportion of choices favoring the slot option with a *lower* expected pay-off, with the social decision associated with that slot option (conforming or dissenting) and the size of the difference in pay-offs between options (large or small) as factors, yielded a main effect of social decision ($F(29) = 7.40$, $p < 0.05$, $\eta p^2 = 0.20$, $BF_{10} = 96.86$ error percentage = 8.48 e-8) and a main effect of the difference in pay-offs ($F(29) = 17.88$, $p < 0.001$, $\eta p^2 = 0.38$, $BF_{10} = 0.50$ error percentage = 0.006), as well as a marginally significant interaction ($p = 0.06$). Comparing the strength of the Bayes factor for a model that includes the interaction term against the null model including both pay-off difference and social decision in all models (nuisance variables) yielded $BF_{01} = 0.44$ (error percentage = 23.75, weak evidence for the null hypothesis). Furthermore, as can be seen in Figure 4, a comparison of model-derived choice probabilities with participants' actual choices suggests that the model that treats social alignment as a surrogate reward dramatically outperformed the non-social model; mean AICc scores were significantly smaller, indicating a better fit, for the social model than for the non-social model ($t(29) = 4.07$, $p = 0.0003$, $d = 0.60$, $BF_{10} = 88.43$ error percentage = 5.12 e-5). The means and standard deviations of the best-fitting parameters, and of the associated negative log likelihoods, are listed in Table 1.

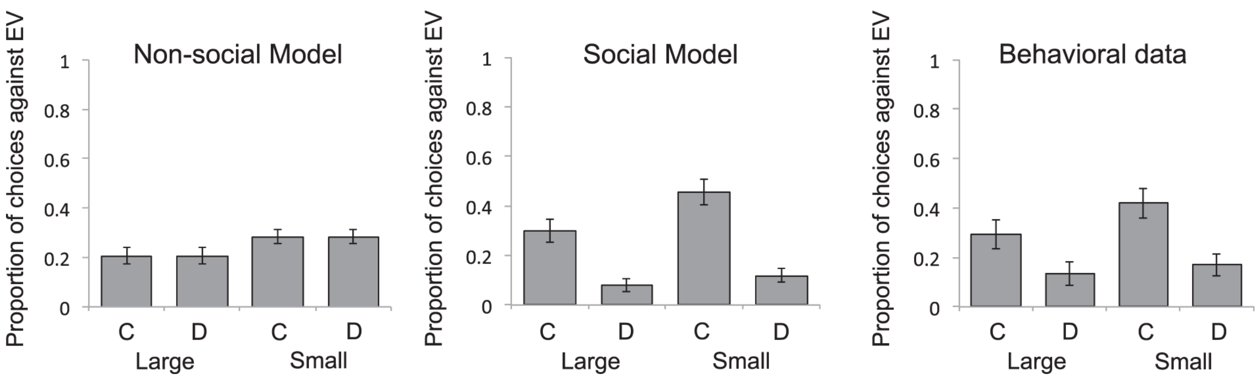

Figure 3.2. Model predictions and behavioral results from Experiment 3. Bars show mean proportions of two-alternative forced choices favoring the option with the *lower* expected monetary value (EV) in each of four conditions, defined by the magnitude of the difference in EV across available options (Large or Small) and by whether the lower EV option was associated with conformity (C) or dissent (D). In each condition, a proportion of 0.5 indicates chance performance (i.e., no preference based on EV). Subtracting the depicted proportion from 1.0 gives the proportion of choices favoring the option with a *greater* EV. The left and middle graphs respectively show mean choice probabilities generated by a non-social model of expected value and an alternative, social, model that uses majority alignment as a surrogate reward. The right graph shows participant's actual choices. Error bars = SEM.

Specifically, as predicted only by the social model, whenever the probability of reward differed across available options, participants were significantly more likely to choose the option associated with a *lower* pay-off if that option was endorsed by a majority of ostensible previous gamblers, whether the difference in pay-offs across options was large ($t(29) = 2.06$, $p = 0.048$, $d = 0.53$, $BF_{10} = 1.23$ error percentage = 4.80 e-6) or small ($t(29) = 3.13$, $p = 0.004$, $d = 0.85$, $BF_{10} = 9.84$ error percentage = 2.20 e-6). In other words, participants appeared willing to relinquish an alternative incentive in order to conform to the group norm.

	w	τ	**NL Likelihood**
Social	0.39 ± 0.38	7.33 ± 13.10	22.73 ± 9.99
Non-social	-	6.85 ± 17.92	30.20 ± 11.21

Table 3.1. Best fitting parameter values in Experiment 3 for the value of conforming (w), the softmax noise (τ), and the negative log (NL) likelihoods for the social and non-social model of decision value.

An important consideration when interpreting these results is the fact that monetary rewards were hypothetical: It is possible, therefore, that the apparent willingness to pay a price for majority affiliation reflected a lack of awareness of, or failure to be incentivized by, monetary pay-offs. We consider this unlikely for a couple of reasons. First, because a large number of behavioral and neuroimaging studies have found similar effects of fictitious and real rewards. Second, and more importantly, the results clearly demonstrated an overall preference for gambling options with *greater* pay-offs that increased significantly with the magnitude of the difference in pay-offs.

When the probabilities of reward were the same for both available options, participants on average chose the option endorsed by a majority of previous gamblers 66% of the time, significantly greater than chance; $t(29) = 2.66$, $p < 0.05$, $d = 0.49$, $BF_{10} = 3.68$ error percentage = 3.06 e-6. Importantly, these choice preferences did not depend on the degree to which a participant had learned or retained the pay-off probabilities: the accuracy of rated reward probabilities for all gambling slots obtained at the end of the study did not predict the degree to which participants favored options associated with conformity over dissent, $p = 0.64$ $BF_{01} = 3.98$, nor did it predict the difference in AICc scores between social and non-social algorithms, $p = 0.94$ $BF_{01} = 4.40$. Likewise, the free parameter *w*, which reflects individual differences in the value of conformity, was predicted neither by the accuracy of recalled reward probabilities at the end of the study, $p = 0.97$ $BF_{01} = 4.40$, nor by the number of training rounds required to learn those probabilities to criterion at the beginning of the study $p = 0.54$ $BF_{01} = 3.67$. By pitting conformity against an alternative incentive, Experiment 3 provided evidence for the hypothesized value of majority alignment. In the previous chapter we started to examine the valence attributed to consensus through conditioned reinforcement. In Experiment 4 we, again, use conditioned reinforcement to see how the value attributed to consensus modulates a transfer of valence to stimuli associated with different monetary values.

3.4 Experiment 4

The ability of hedonic stimuli to transfer valence to neutral stimuli with which they are paired, termed *conditioned reinforcement*, has been studied extensively using a wide range of stimuli, species and procedures (Arroyo et al., 1998; Goldberg 1973; Katz, 1979; Williams, 1994). Once established, previously neutral conditioned reinforcers can pass on their motivational significance to other neutral stimuli. In Experiment 4, we assess the degree to which the valence of conformity and dissent decisions are transferred to concomitant stimuli. In particular, we explore how the rewarding properties of social conformity may modulate a previously demonstrated increase in the rated likability of visual stimuli paired with monetary reward, as well as the preference for such stimuli when placed in a novel choice context (Cox et al., 2005).

Additionally, in this experiment we arbitrate between potential cognitive mechanisms that drive consensus seeking behavior. In the previous chapter, we discussed two experiments that showed decrease in likeability of contexts paired with dissent (specifically where the participant disagreed with the group). We tentatively interpreted the results as a negative property of dissent. A prominent theory of consensus seeking is that it reflects an attempt to escape from the cognitive dissonance elicited by a mismatch between one's own judgment and that of other individuals (Festinger 1957; Matz & Wood, 2005; Klucharev et al., 2009). This account implies that if measured against a neutral baseline, valence associated with conformity and dissent should be asymmetric, since it is the negative affect associated with dissent that motivates conformity. Additionally, a direct experience of conformity or dissention, such that it is *one's own* decisions that conflict with those of others, should be required for affect to emerge.

To arbitrate between cognitive mechanisms driving consensus we assessed the degree to which motivational significance is attributed to ostensible other individuals engaging in conforming and dissenting decisions. If consensus does possess intrinsic value, states defined by high levels of agreement or dissent should be symmetrically associated with gains and losses respectively, acquiring positive and negative valence accordingly.

3.4.1 Methods

Participants: Thirty undergraduates at the University of California, Irvine (23 females, mean age = 20.26 ± 2.01) participated in the study for course credit. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine, approved the study. All aspects of the study conformed to the guidelines of the 2013 WMA Declaration of Helsinki.

Task and Procedure: At the start of the experiment, participants were instructed that they would be playing a game where they would be required to select between pairs of slots on a game board that yield a fictitious monetary reward ($1) with various probabilities. They were further told that before playing the game themselves, they would have an opportunity to learn about the game and choices of prior participants, henceforth referred to as "players." In addition, they were instructed that during this learning phase, to ensure that they are paying attention, they would be asked to predict the choices of the presented player and must accurately predict player choices to proceed to the next phase of the experiment.

Training Phase and Pre-Likability Ratings: Before starting, participants were required to train on the probability of reward for each of the slots available in the learning phase and were asked to rate the likability of the six target players. During the training, participants were asked to select each slot 10 times and were shown whether or not they received the $1 reward each time they made a selection. Subsequently, they were asked to report the probability of receiving a reward on each of the slots in random order on a scale of 0 to 1 where 1 meant there was a 100% chance of receiving a reward on that slot. If participants did not rate the probability of reward within 0.2 of the true probability for each slot, they were asked to repeat the training. Likability was, again, measured on a scale from 0 (not at all likable) through 10 (extremely likable), with 5 indicating neutral affect. Players were represented by differently colored rectangular icons with each player's first initial and last name at the bottom. Player colors were randomized across participants.

Prediction Phase: During the prediction phase, on each trial, participants were shown the following: first game board, two available slots, choices made by six other players, and finally one of the six target players whose choice the participant will be asked to predict (see figure 5). The board was represented on a 3 by 2 grid, and the slots were all given a number label. The two available slots were highlighted in a light grey color. Additionally, the available slot numbers were presented at the top of the screen where player choices were represented based on which slot number the icons were beneath. At the beginning of each trial, there were six grey icons representing anonymous players beneath the two numbers—divided such that there were between four and six individuals in the majority. Participants were asked to predict the choice of the target player. After the participant made a prediction, the true choice of the player was revealed. If participants made incorrect predictions for more than 10 trials, the entire prediction phase was repeated.

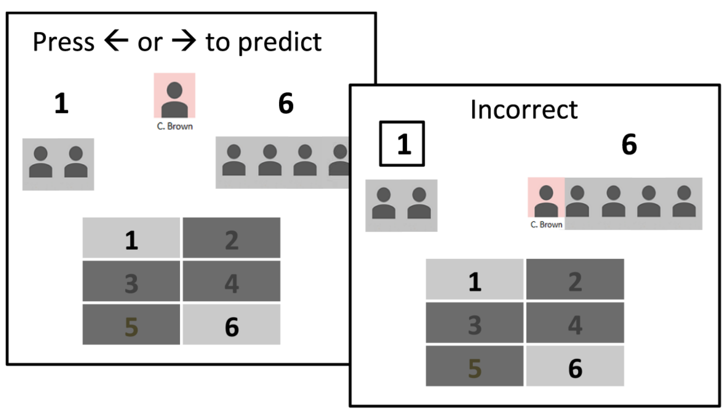

Figure 3.3. Illustration of trials in Experiment 4. Choice and feedback screens on a trial in the social learning phase. Participants pressed the left or right arrow key to predict which option the target gambler would choose. On the subsequent screen, the participant's selection was indicated by a square around the chosen option, the target gambler's avatar moved below the option ostensibly selected by the target gambler, flanking a set of additional avatars that indicated the decisions of several non-descript players given the same options, and accuracy feedback was displayed center-screen.

Player choices in the prediction phase were distributed such that 3 levels of estimated reward emerged ($3.5, $4.7, $5.9) across the players. Specifically, one player from the consensus group and one player from the dissent group were paired with each level of reward, by varying the frequency with which each player selected differently rewarded slots on the game board. See Table 2 for details.

	0.8-0.8	**0.8-0.5**	**0.8-0.2**	**0.5-0.8**	**0.5-0.5**	**0.5-0.2**	**0.2-0.8**	**0.2-0.5**	**0.2-0.2**
$5.9	1	4	1	1	0	0	1	1	1
$4.7	0	1	1	2	1	2	1	1	1
$3.5	0	0	1	0	1	2	1	4	1

Table 3.2. Possible pairings of slot reward probabilities into two-alternative forced choice scenarios during the social learning phase of Experiment 4, with the number of instances each scenario was presented to target gambler associated with high ($5.9), medium ($4.7) and low ($3.5) cumulative gain. The first number in a pair (e.g., 0.8 in 0.8-0.5) is the reward probability of the slot selected by the target gambler and the second number that of the alternative slot available on the trial.

Gambling Phase and Post-Likability Ratings: After the prediction phase, participants were once again asked to rate the likability of the six players using the same scale. In the final phase of the experiment, participants were instructed that they would now have an opportunity to play the game shown in the prediction phase of the experiment on a new game board. In addition, they were told that before making a selection, they would be shown which slot two of the target players had selected when given the same choice pair, with one player endorsing each slot. Finally, they were instructed that they would not be shown any rewards until the end of the experiment. Consensus and monetary reward were pitted against each other in the selection phase, such that participants were asked to choose between a player associated with consensus but a lower estimated obtained reward, and a player associated with dissent and a higher estimated obtained reward. We also included trials where consensus was balanced across expected value and trials where expected value was balanced across consensus. In addition, there were trials where the reward discrepancy between players associated with consensus and dissent was small and trials where the discrepancy was large. There were 15 possible combinations of target gamblers, with each combination being repeated 8 times for a total of 120 block randomized trials.

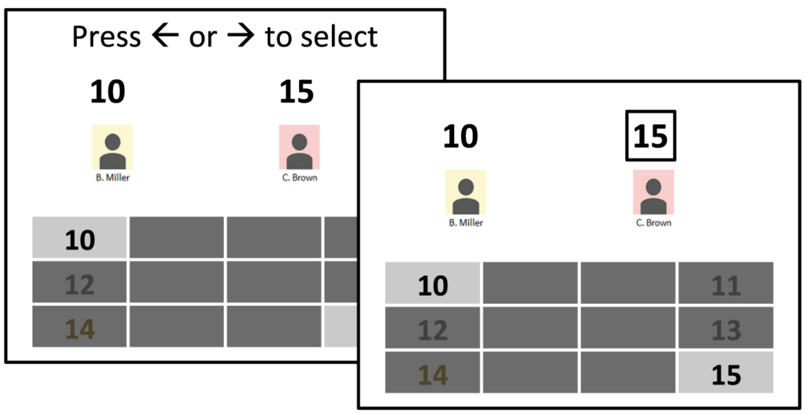

Figure 3.4. Choice and feedback screens on a trial in the gambling phase. Participants pressed the left or right arrow key to select one of two slot options drawn from a set of novel options with unknown reward probabilities. On critical trials, a conforming and dissenting target gambler from the previous phase respectively endorsed each slot option. On the feedback screen, the participant's selection was indicated by a square around the chosen option.

Majority Rating: At the end, participants were asked to complete a final evaluation of the players where they rated the degree to which each player agrees with the majority on a scale from 0 to 10, where 0 meant the player never agreed with the majority and 10 meant the player always agreed with the majority.

Debriefing: After completing the experiment participants were informed that decisions made by "previous gamblers" were in fact generated by a computer and had an opportunity to withdraw their data from the experiment in light of the information. All participants gave written consent to having their data included.

Computational Model: Changes in the valence of target gamblers was formalized by model-free reinforcement learner that incrementally updated the value of each gambler based on the difference in experienced and expected value on each trial (Rescorla & Wagner 1972; Sutton & Barto, 1998). Specifically, on a given trial, the change in value of a target gambler was defined as:

$$\Delta V(s) = \alpha[R(s')-V(s)], \qquad \textbf{(6)}$$

where alpha is the learning rate, s represents the color of the target gambler, and $R(s')$ is defined either by a conventional monetary reward or additionally treats conformity as a surrogate reward. We assumed that each slot had acquired value proportional to the probability of reward on that slot in the pre-learning phase, and that in the training phase, all target gamblers grew associated with the expected reward outcomes for the trials in which they were present. Thus, on each trial in the "social learning" phase, the value of the target gambler present on that trial was updated with an amount proportional to the difference between the initial value of that target gambler and the reward probability (and conformity level) of the slot outcome on that trial. In the gambling phase, expected decisions were determined by the sum over the products of the probabilities and rewards of outcome states:

$$V(a) = \sum_{s'} P(a, s')R(s), \qquad \textbf{(7)}$$

where a is the selection of a particular slot option, $P(a,s')$ is the probability of a particular outcome state, s', given a, and $R(s')$ is the reward of s'. Here, s' is defined as the target gambler associated with a particular slot option.

A softmax rule was used to generate choice probabilities, and free parameters were fit to data by minimizing the negative log-likelihood of choices made in the final gambling phase, as participants selected between slot options endorsed by different target gamblers. Thus, the greater the value of w, the greater the value acquired by a conforming target gambler in the social learning phase, and the greater the probability of choosing a slot option endorsed by that gambler in the subsequent gambling phase. Fits were performed using MATLAB's fminsearchbnd function (MathWorks, 2017b), with upper-lower bounds of 0.01–0.99 for α, 0.01–1.01 for w, 0.01–100.00 for τ. The corrected Akaike information criterion (AICc) was used to select between models.

3.4.2 Results

A two-by-three ANOVA performed on the likability ratings, with social decision and cumulative gain as factors, revealed a significant main effect of social decision ($F(29) = 19.50$, $p < 0.001$, $\eta p^2 = 0.40$, $BF_{10} = 4.34$ e11 error percentage = 2.60 e-14) but no effect of cumulative gain ($p = 0.23$ $BF_{01} = 12.29$ error percentage = 0.008) and no interaction ($p = 0.42$). Comparing the strength of the Bayes factor for a model that includes the interaction term against the null model including both cumulative gain and social decision in all models (nuisance variables) yielded $BF_{01} = 6.89$ (error percentage = 8.49, positive evidence for the null hypothesis). Planned comparisons revealed that mean ratings deviated significantly from the neutral point (5) on the rating scale for both conforming (t(29) = 5.04, p=2.27 e-5, $d = 0.92$, $BF_{10} = 997.35$ error percentage = 2.26 e-6) and dissenting gamblers (t(29) = 2.92, p =0.007, $d = 0.54$, $BF_{10} = 6.43$ error percentage = 2.51 e-14); that is, social valence changed in both a positive and negative direction. In the final gambling phase, when target gamblers associated with the same amount of cumulative reward endorsed different slot options, participants' preference for the option endorsed by a gambler associated with conformity over one associated with dissent was significantly greater than chance; 69%, t(29) = 2.76, $p < 0.01$, $d = 0.50$, $BF_{10} = 2.72$ e8 error percentage = 3.53 e-16.

As noted, we attribute both likability ratings and choice preferences to changes in the affective valence of target gamblers respectively associated with conformity and dissent. Consistent with this interpretation, choice preferences were significantly predicted by differences between conforming and dissenting target gamblers in likability ratings, $p = 0.0001$ $BF_{10} = 4.34$ e11 error percentage = 2.6 e-14. Importantly, the above results did not reflect a failure to learn or retain the reward probabilities associated with different slot options: participants' ratings of reward probabilities at the end of the study were highly accurate, with 97% of estimates falling within 0.1 of, and 91% of estimates being identical to, programmed probabilities. Moreover, to the extent that they existed, deviations of estimations from programmed reward probabilities did not predict individual differences in either preference for ($p = 0.20$ $BF_{01} = 2.04$) or likability of ($p = 0.57$ $BF_{01} = 3.79$) conforming gamblers. Participants were also clearly able to discriminate between conforming and dissenting target gamblers, t(29) = 8.14, p =5.63e-9, $d = 2.76$, $BF_{10} =$ 2.23 e6 error percentage =3.55 e-10; however, their ability to do so predicted neither differences in the likability of conformity and dissent gamblers, $p = 0.53$ $BF_{01} = 3.65$, nor their preference for options endorsed by conformity gamblers, $p = 0.38$ $BF_{01} = 3.04$.

	High EV	Mid EV	Low EV
Rated Likability			
Conforming	6.77 ± 2.13	6.90 ± 1.88	6.17 ± 2.18
Dissenting Player	3.87 ± 2.37	3.80 ± 2.31	3.73 ± 2.59
Social Model			
Conforming	0.47 ± 0.51	0.38 ± 0.39	0.37 ± 0.40
Dissenting Player	0.27 ± 0.26	0.22 ± 0.22	0.17 ± 0.18
Non-Social Model			
Conforming	0.43 ± 0.31	0.30 ± 0.22	0.25 ± 0.20
Dissenting Player	0.38 ± 0.28	0.33 ± 0.23	0.25 ± 0.21

Table 3.3. Mean rated likability ratings of conforming and dissenting target gamblers in Experiment 4, at each level of cumulative gain (i.e., expected value; EV), together with corresponding mean state values, *V(s)*, derived by social and non-social computational models.

A two-by-two ANOVA performed on the proportion of choices favoring options endorsed by target gamblers associated with lower cumulative reward, with social decision (conforming or dissenting) and the difference between target gamblers in cumulative gain (small or large) as factors, revealed a main effect of social decision ($F(29) = 9.11$, $p < 0.01$, $\eta p^2 = 0.24$, $BF_{10} =$ 47921.33 error percentage = 6.82 e-10) but no main effect of the difference in cumulative gain ($p = 0.90$, $BF_{01} = 5.14$ error percentage = 0.013), and no interaction ($p = 0.53$). Comparing the strength of the Bayes factor for a model that includes the interaction term against the null model including both cumulative gain and social decision in all models (nuisance variables) yielded $BF_{01} = 3.78$ (error percentage = 3.74, positive evidence for the null hypothesis). As predicted only by the computational model that treated conformity as a surrogate reward, and illustrated in Figure 7, when the history of conformity as well as of cumulative payoffs differed across target gamblers endorsing different options in the final gambling phase, participants chose the option endorsed by the target gambler associated with a lesser payoff significantly more often if that gambler had a history of making conforming decisions, whether the difference in cumulative payoffs was small, t(29) = 2.56, p = 0.016, d = 0.91, BF_{10} = 16.29 error percentage = 1.91 e-6 or large, t(29) = 3.35, p = 0.002, d = 1.07, BF_{10} = 3.06 error percentage = 3.29 e-6. The superior performance of the social model was confirmed by planned comparisons of AICc scores, which were significantly smaller, indicating a better fit, for the social than for the non-social model (t(29) = 5.41, p = 8.04 e-6, d = 1.05, BF_{10} = 2579.17 error percentage = 1.05 e-6). The means and standard deviations of the best-fitting parameters, and of the associated negative log likelihoods, are listed in Table 3.

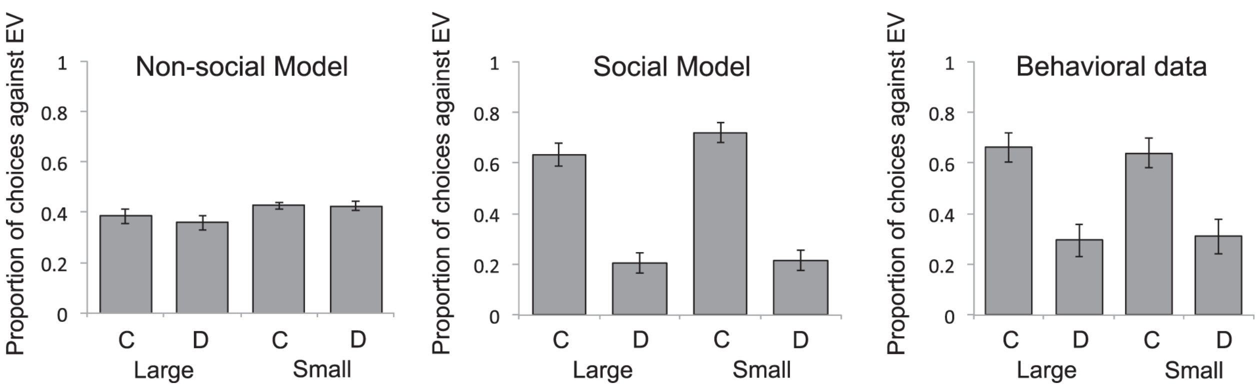

Figure 3.5. Model predictions and behavioral results from Experiment 4. Bars show mean proportions of two-alternative forced choices favoring the option endorsed by a target gambler with a *lower* cumulative monetary gain (expected value, EV, on y-axis), in each of four conditions, defined by the magnitude of the difference in cumulative gain across target gamblers (Large or Small) and by whether the gambler associated with a lower cumulative gain was associated with conformity (**C**) or dissent (**D**). In each condition, a proportion of 0.5 indicates chance performance (i.e., no preference based on the endorsing target gambler's cumulative gain). Subtracting the depicted proportion from 1.0 gives the proportion of choices favoring the option endorsed by a target gambler with a *greater* cumulative gain. The left and middle graphs respectively show mean choice probabilities generated by a non-social model of expected value and an alternative, social, model that uses majority alignment as a surrogate reward. The right graph shows participant's actual choices. Error bars = SEM.

	α	w	τ	**NL Likelihood**
Social	0.23 ± 0.35	0.49 ± 0.44	20.93 ± 28.76	42.70 ± 17.42
Non-social	0.40 ± 0.40	-	8.98 ± 19.74	57.55 ± 6.42

Table 3.4. Best fitting parameter values for the learning rate (α), the value of conforming (w) and softmax noise (τ), together with negative log (NL) likelihoods, for the social and non-social model of decision value.

As with differences in likability ratings and overall preferences, differences in AICc scores between the non-social and social model were not predicted by differences in the acquisition (p = 0.15, BF_{01} = 1.68) or retention of slot reward probabilities, $p = 0.15$, BF_{01} = 1.60. Moreover, differences in learning rates derived from the social model predicted neither likability ratings, $p = 0.72$, BF_{01} = 4.14, nor choice preferences, $p = 0.63$, BF_{01} = 3.93. Finally, the value of the w parameter was predicted neither by the accuracy of recalled reward probabilities at the end of the study, $p = 0.56$ BF_{01} = 3.74, nor by the number of training rounds required to learn those probabilities to criterion at the beginning of the study, $p = 0.83$, BF_{01} = 4.31. It might be argued, however, that in spite of the relatively large differences in slot reward probabilities, differences in the cumulative gain associated with conforming and dissenting target gamblers, based on their ostensible decisions in the social learning phase, were too subtle to be discernable, or deemed relevant, to participants in the subsequent gambling phase. To address this possibility, the gambling phase included trials on which both target gamblers were equally associated with either conformity or dissent, differing solely in terms of the monetary rewards accumulated in the social learning phase. On such trials, participants showed a significant preference for the option endorsed by a gambler associated with greater cumulative monetary gain; mean preference = 0.59 ± 0.21, $t(29) = 2.29$, $p = 0.029$, $d = 0.42$, BF_{10} = 1.84 error percentage = 4.06 e-6.

3.5 Discussion

In two experiments we investigated the affective properties of agreeing or disagreeing with an ostensible group majority by pitting conformity against monetary gain and assessing the transfer of social valence to concomitant stimuli. Furthermore, using a reinforcement learning framework to formalize the role of consensus in value-based decision making, we found that models that treat agreement with a group majority as a surrogate reward provide a better account of choice preferences than conventional algorithms. In Experiment 3, when the probability of a fictitious monetary reward differed across available options, participants chose the option associated with a *lesser pay-off* significantly more often if that option was also selected by a majority of ostensible previous gamblers. In Experiment 4, participants reported a greater likability of gamblers that had a history of agreeing with the majority over gamblers that had a history of dissent. Additionally, participants demonstrated a preference for options endorsed by gamblers that had a history of agreeing with the majority, even when dissent gamblers were associated with higher cumulative reward. Critically, these effects were not predicted by participants' ability to accurately recall the objective reward probabilities.

An important consideration is how ostensible previous gamblers were perceived by participants given their suboptimal decisions. Specifically, in Experiment 3 the majority chose the option with lower expected value on 40% of trials and in Experiment 4, even the gambler associated with the highest level of reward chose the less valuable option on 30% of the trials. In contrast, participants chose the option with lower expected value on only 20% of trials (averaging across consensus and dissent decisions) in Experiment 3. It is possible that participants attributed the performance of the ostensible gamblers to their level of expertise. Specifically, participants may infer that gamblers who chose inferior options had a lesser understanding of the reward structure. Further work is needed to explore how optimality of others' decisions informs perceived expertise.

Factors related to informational conformity such as majority size and expertise may also influence the hedonic aspects of social alignment (Deutsch & Gerard 1955; Rafaat et al., 2009; Toelch et al., 2013; Campbell-Meiklejohn et al., 2016). For example, in Experiment 4, we assume that the relationship between majority size and value is linear. However, previous work assessing the relationship between majority size and conformity is mixed with some showing a curvilinear relationship, others reporting linear relationships, and additional others showing diminishing functions where each additional majority members produces a smaller increment of conformity (Rosenberg, 1961; Gerard et al., 1968; Latané & Wolf, 1981). In our experiments, the majority size was not varied enough to assess linearity. Still, it is important to consider how size of majority opinion and expertise may shape reinforcement signals associated with conformity. For example, agreement with a small minority of experts may have a greater history of reward than agreement with a large majority of lay-people. Formally, this may be addressed by replacing c(s) – our implementation of consensus as a surrogate reward – with a power function. The exponent of the power function would represent the relationship between expertise, majority size and value.

Another possible source of negative affect associated with dissent is cognitive dissonance – a feeling of discomfort induced by interpersonal or intrapersonal discrepancy (Matz & Wood, 2005; Klucharev et al., 2009; Klucharev et al., 2011). Recall that in chapter 1, we found in two experiments that the change in valence of contexts associated with consensus was driven by a decrease in likability of courtrooms associated with dissent. From a reinforcement learning perspective, just as dissent may acquire negative valence from being paired with uncertainty aversion and cognitive dissonance, conformity should acquire positive valence from being associated with greater access to group resources, better outcomes, and better judgement accuracy. Notably, the likability ratings obtained in Experiment 4 that affective changes are also driven by a positive valence associated with social alignment. Since the likability ratings here were based on the behavior of ostensible other individuals, it is possible that the negative valence seen in Experiments 1 and 2 were partially due to cognitive dissonance. Further work is needed to determine the symmetry of valences associated with conformity and dissent.

Consistent with the idea that majority affiliation serves as a positive reinforcement signal, several neuroimaging studies have found greater activity in the ventral striatum (VS), a region known for coding reward prediction error, when individuals find a group norm to be in agreement with their own judgement (McClure et al., 2003; O' Doherty et al., 2003; Hare et al., 2008; Klucharev et al., 2009; Nook & Zaki, 2015). However, without baseline measures, it is difficult to figure out the directionality of VS responses to social feedback. Indeed, Klucharev et al., interpreted the effect as a deactivation of the VS in response to aversiveness of diverging from the group norm. Although this interpretation is supported by some studies showing decreased VS activity as a response to aversive stimuli, other studies have shown VS activity to be bivalent (increasing in response to both appetitive and aversive stimuli, or even nonvalent (increasing to neutral but surprising stimuli) (Maeda & Mogenson, 1982; Besson & Louilot, 1995; Ungless et al., 2004; Jensen et al., 2003; O'Doherty et al., 2006; Seymour et al., 2007; Levita et al., 2009; Horvitz, 2000; Zink et al., 2003). Additional work is necessary to assess how the results here and the formal framework of reinforcement learning relate to the neural basis of conformity.

In conclusion, we have used conventional measures of subjective value to explore the affective properties of conforming and dissenting decisions. Our results suggest a common value-scale for social and non-social currencies, and an ability of conforming decisions to transfer value to concomitant stimuli.

Chapter 4. Correlation Between Consensus Seeking and Exploitation

4.1 Abstract

There is converging neural and behavior evidence that reward learning may play a role in consensus seeking behavior. Individual differences in behavioral conformity may be a product of individual differences in the value of consensus or may reflect general differences in reward processing and the ability to overcome more habitual reward mechanism. Notably, in the exploration-exploitation literature, exploratory behavior has been attributed to overcoming habitual decision systems involved in exploitation. In this experiment, we use a frequently used conformity task that involves subjective judgements and normative feedback to create individual conformity scores. We then use a 30-armed bandit problem previously used in a social learning experiment looking at exploration and exploitation to derive a metric for each individuals' exploratory behavior and compare it to their conformity scores. As predicted, we found a significant negative correlation between conformity and exploratory behavior. Importantly, conformity was not correlated with performance on the 30-armed bandit problem.

4.2 Introduction

In the previous two chapters, we put forward several experiments that provide evidence for an intrinsic affective property of consensus. Moreover, in Experiments 3 and 4 participants continued to conform or select options associated with consensus when those decisions came at a cost. This suboptimal behavior might reflect the "incentive salience" of conformity: incentive salience is a property of a rewarding stimuli that elicits compulsive approach (Berridge, 2007). In other words, participants, in particular those that tend to generally fail to override reward impulses, may feel compulsively drawn to the reward of consensus options. In experiment 5, we assessed whether the tendency to conform is associated with a failure to explore the reward structure of the environment, indicative of compulsive reward approach. Specifically, we investigated whether individual differences in exploration predicted conformity.

Exploration is frequently defined as choosing an option that does not have the highest experienced expected value of the options available (e.g., Farrias & Megiddo, 2005; Daw et al., 2006; Hills et al., 2015; Beharelle et al., 2015). Often, some exploration is necessary to maximize reward. There is a large literature examining the strategic balancing of exploration and exploitation, where human decisions are compared to optimal solutions in specific contexts (Horowitz, 1973, Meyers & Shi, 1995; Banks, Olson & Porter, 1997; Anderson, 2001). Striking the correct balance between exploration and exploitation is a computationally complex task requiring careful regulation. Indeed, suboptimal decision strategies in this context have been linked to addiction and other compulsion related disorders (Harlé et al., 2015; Morris et al., 2015; Reiter et al., 2017; Addicott et al., 2017). For example, Morris et al. (2015) found that individuals with alcohol use disorder showed reduced exploratory behavior when compared to nonclinical volunteers which led to lower-yielding exploratory choices. Excessive attribution of incentive salience to stimuli associated with reward is a feature of such compulsive disorders (Albertella et al., 2019; Berridge, 2007).

Neural and genetic evidence suggests that exploratory decisions are made by overriding an exploitive tendency (Daw et al., 2006). Specifically, studies show that exploitation is linked to activity in reward substrates while exploratory decisions have been linked to activity in regions linked to attentional control (Laureiro-Martinez et al., 2015; Seymour et al., 2012; Daw et al., 2006). Moreover, genes controlling striatal dopamine function are associated with exploitative actions, while genes controlling prefrontal dopamine function are associated with exploratory decisions (Frank et al., 2009). Finally, reducing COMP enzymatic activity – which affects prefrontal dopamine function - results in an increase in exploratory actions (Kayser et al., 2015).

Similarly, lower COMP enzymatic activity is associated with lower levels of conformity (Campbell-Meiklehojn, 2010). It is possible that those with a tendency to conform have greater difficulty overriding the incentive salience attributed to consensus (Daw et al., 2006; Klucharev et al., 2009; McClure et al., 2003). Incentive salience is assigned to stimuli that have previously been paired with reward (Albertella et al., 2019). Through assignment of incentive salience such stimuli are subsequently transformed such that they are wanted, independent of reward (Berridge, 2007).

As discussed in previous chapters, there is both behavioral and neural evidence that the act of reaching consensus is intrinsically rewarding. Notably, studies have demonstrated that those with a tendency to conform show a greater change in neural activation in response to consensus information. Klucharev (2009) found that individuals who had a tendency to adjust their own subjective ratings towards an ostensible group mean rating (conform) showed stronger conflict-related deactivation in the ventral striatum. Furthermore, those that had a tendency to conform showed a smaller difference in deactivation in the ventral striatum when comparing trials in which participants conformed and did not conform. Similarly, Nook and Zaki (2015) found that ventral striatal activity, upon receiving feedback of an ostensible group rating, predicted individuals' tendency to conform.

It is possible that these differences in striatal response are driven by individual differences in the value of consensus. However, an alternate explanation is that more general differences in reward processing may contribute individual differences in conformity. Specifically, individual differences in attributing incentive salience to cues that predict reward may mediate individual differences in conformity, where those who have a tendency to conform may have a more compulsive approach to reward (Flagel et al., 2009). To assess this possibility, we compared individual differences in conformity to individual differences in exploratory behavior. If individuals who conform have greater difficulty overriding reward impulses, they should engage in less exploration, choosing instead to exploit options associated with higher experienced reward.

4.3 Experiment 5

4.3.1 Methods

Participants: Sixty undergraduates at the University of California, Irvine (42 females, mean age = 21.35± 3.43) participated in the study for course credit. We rounded up the sample size in the study done by Toyokawa et al. (2014) since we use their paradigm here. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine, approved the study.

Task and Procedure: This experiment consisted of a conformity task and an exploration-exploitation task. We counterbalanced the order of the tasks across participants.

Conformity Task: This paradigm was adapted by Nook and Zaki (2015) from similar experiments that have demonstrated conformity to a group norm in other contexts. (Zaki, Schirmer & Mitchell, 2011; Klucharev et al., 2009; Izuma & Adolphs, 2013; Campbell-Meiklejohn et al., 2010; Mason et al., 2009).

Participants were told that they were part of a large-scale study on food preferences of the University of California, Irvine undergraduate community and that hundreds of other students have already provided their preference ratings. Participants first completed 150 trials in which they rated how much they wanted to eat particular food items on a scale from 1 (dislike) through 8 (like). On each trial, after indicating their preference, a blue square appeared around the participant's rating. Subsequently a red square appeared for two seconds around a number that indicated the ostensible average rating of the 200 previous participants. If the participant rating and the group rating were identical, then the word "Agree" appeared above the rating. Otherwise a signed integer indicating the difference between the participant's score and the group score appeared above the group rating. There were three group norm conditions: on approximately one third of trials the group rated food items as 1, 2, or 3 points above the participant ratings, approximately one third of the trials had group ratings identical to the participant ratings, and approximately one third of the trials had group ratings 1, 2, or 3 points below the participant rating. Approximately one minute after the feedback task, participants completed 150 follow up trials in which they again rated their preferences for the 150 food items. During the follow up trials, participants did not receive group feedback. The order in which the food items appeared were randomized. Additionally, the number of points by with the group rating differed from the participant rating was randomly outputted on each trial with the constraint that the group mean had to remain on the number scale of 1 through 8.

Exploration-Exploitation Task: This paradigm used in this experiment was adapted from an experiment examining social learning in a dual exploration-exploitation task (Toyokawa et al., 2014). Specifically, to assess individual differences in exploration we had participants complete 100 rounds on a 30-armed bandit problem.

Generally, a bandit task is any economic choice task in which participants choose between two or more stochastic initially unknown, and sometimes changing, reward distributions. Multi-armed bandit problems are commonly used to study the exploration-exploitation dilemma because they allow for comparison between human and optimal decision strategies and the formal examination of cognitive mechanisms underlying exploratory behavior. For example, Steyvers, Lee and Wagenmakers (2009) used four-armed bandit problems to examine individual differences in how people balance exploration and exploitation using a Bayesian model of optimal decision-making where differences in decision making can be explained in terms of differences in assumptions individuals make about reward rate distributions. Similar tasks have been used to understand strategies and neural substrates involved in the exploration-exploitation dilemma (Laureiro-Martinez et al., 2015; Seymour et al., 2012; Daw et al., 2006). Notably, the bandit task is also used throughout the decision sciences to assess choice preferences for a range of theoretical topics including uncertainty aversion, ambiguity aversion, relative weighting of reward and punishment, social decision making, habitual behavior, incentive salience and novelty seeking (Wittman et al., 2008; Anderson, 2012; Seymour et al., 2012; Addicott et al., 2013; Tokoyawa et al., 2014; Harlé et al., 2015; Morris et al., 2016). The specific paradigm used in this experiment was selected because of its previous use in studying social learning and because it resulted in high exploration and more variance in the number of exploratory choices compared to other similar tasks (Toyokawa et al., 2014).

At the beginning of the task, participants were instructed that they would be shown a gameboard with 30 slots, and on each round of the experiment they would be asked to select a slot on the gameboard. They were further told that each slot had a unique probability distribution with which it outputted points and that the number of points outputted on a round would be randomly determined based on the chosen slot's unique probability distribution. Participants were asked to try to earn as many points as possible. Additionally, we tried to incentivize participants by informing them that at the end of the experiment, they would be able to learn how well they scored in comparison to previous participants. On each round the participant clicked on their preferred option and were next informed of their payoff. Payoffs were randomly generated from a stationary uniform probability distribution with an interval [min, min+150] where min was the minimum payoff. 11 of the 30 options had a min of 0, eight had a min of 15, five options had a min of 30, three options had a min of 45, two options had a min of 60, and one option had a min of 75. The distribution of minimum payoffs was meant to represent a foraging task where higher outcome options were more rare. Participants were not informed of the number of rounds that they would be required to complete but were shown their total number of points earned thus far at the top of the screen. The location of the various slot types on the gameboard was randomly determined.

Debriefing: At the end of the experiment, participants were informed that the group means they were shown were generated by a computer and not based on real participant ratings. Participants had an opportunity to withdraw their data from the experiment in light of the information and all participants gave written consent to having their data included.

4.3.2 Results

Conformity task: Consistent with previous work, we found that participants tended to shift their food ratings to be consistent with that of a group norm. Specifically, we conducted a mixed effects analysis of participant ratings on a trial level with follow up ratings as the response variable, group norm condition as a fixed effect, participant as a random effect, and participants' initial ratings as a fixed effect covariate and found a significant main effect of group norm condition ($F(59) = 18075.771$, $p < 0.001$). Furthermore, as in the original experiment, participant ratings were lowest for foods in the peers lower condition ($M = 3.33 \pm 1.05$), higher in the peers agree condition $M = 4.41 \pm 0.98$), and highest in the peers higher condition $M = 5.52 \pm 1.00$).

Also, as in the original experiment, participant level analysis of shift direction showed that participants were most likely to decrease their ratings for foods in the peers lower condition, not change their ratings in the peers agree condition, and increase their ratings for foods in the peers higher condition. The conformity effects here may have been stronger than the original paper due to the larger sample size and the lack of real reward.

Conditions	**Rating**	**Rating**	**Rating Increased**
Peers	31.11% ± 13.97	51.97% ± 16.96	16.92% ± 09.39
Peers Agree	22.57% ± 12.31	58.13% ± 15.08	19.30% ± 08.50
Peers	19.91% ± 09.62	53.83% ± 16.72	26.26% ± 13.68

Table 4.1. Mean proportion of trials in each condition for which participants decreased, did not change, or increased their ratings from Initial to Follow-up Ratings.

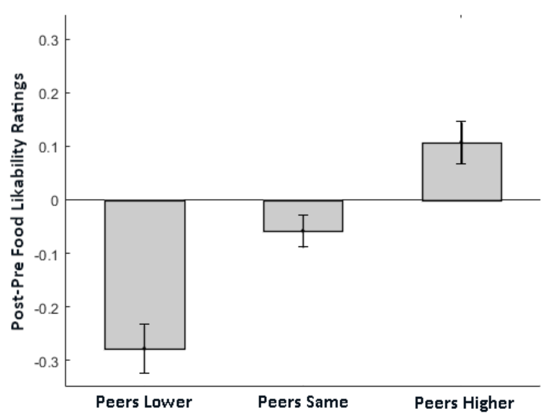

Figure 4.1. Difference in follow up rating and initial rating for various food items based on whether the group mean rating was lower than participants (peers lower), the same as participant (peers same), or higher than participants (peers higher). Adjusted ratings are significantly different across group norm conditions (*p*s<0.001).

Exploration-Exploitation Task: Here we defined exploration as choosing an option that had not, in the participant's experience, on average yielded the highest reward. Conversely, we defined exploiting as choosing the option that had yielded the highest average reward across trials selected. Similar to the previous experiment there was a large amount of exploration and participants explored significantly more often than choosing to exploit (*Mean explorations =* 92.75 ± 11.23 $t(59) = 29.4661$, $p < 0.0001$, 95% CI [89.85, 95.65] $BF_{10} = 3.85$ e33). Unlike the previous study, we did not find more exploratory behavior later in the experiment as opposed to earlier. Differences in exploratory behavior may have been due to the fictitious nature of rewards.

Correlations: To quantify individual differences in conformity, we computed behavioral conformity scores for each participant by calculating the Pearson's *r* correlation between 1) the difference between group ratings and the participant's original rating and 2) the difference between the participant's follow up rating and original rating. These correlations were computed across all trials for each participant and the raw scores were transformed to z scores to ensure normal distribution for subsequent analysis. We observed a significant negative correlation between conformity scores and exploratory behavior where exploratory behavior was defined as previously stated ($r = -0.4267$ $p=0.0007$, $BF_{10} = 45.42$). Moreover, there was a significant negative correlation between conformity scores and exploratory behavior when defining exploratory behavior as choosing an option different from the one selected on the previous trial ($r = -0.4240$ $p=0.0007$, $BF_{10} = 42.00$). To ensure that those with higher conformity scores, and lower levels of exploration were not simply paying more attention we also checked if conformity scores predicted performance in terms of average points earned across trials and found that it did not ($p = 0.3473$, $BF_{01} = 4.03$).

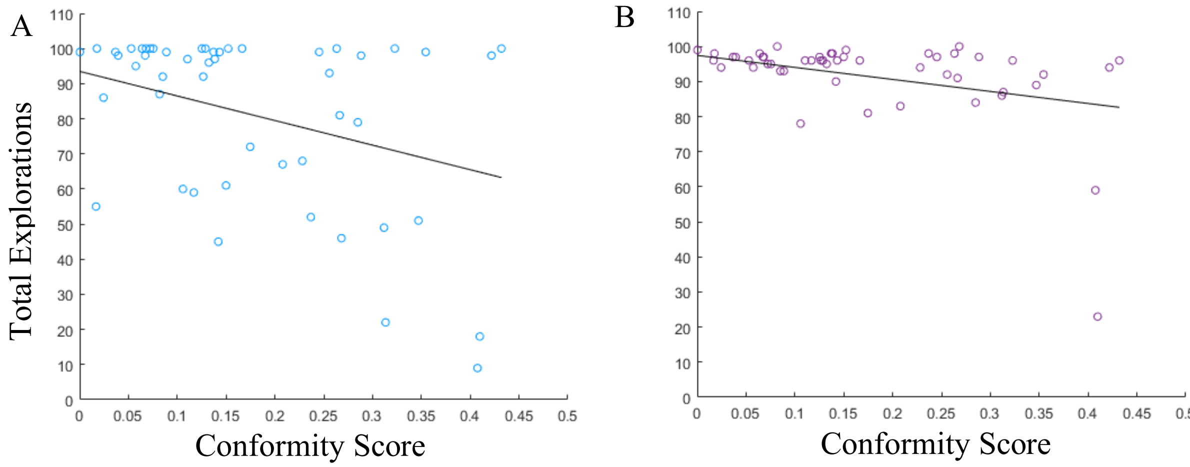

Figure 4.2. Correlations between exploratory behavior and behavior conformity scores (*p's<0.001*). The scatterplot on the right (A) defines exploratory behavior as choosing an option that is different as the one selected on the trial prior. The scatterplot on the left (B) defines exploratory behavior as choosing the option that does not yield the highest expected reward.

4.4 Discussion

Here we show that behavioral conformity scores are negatively correlated with exploratory behavior. Critically, conformity was not correlated with performance in the exploration-exploitation task. We tentatively interpret these results to as evidence that individual differences in consensus may be partially driven by general differences in attribution of incentive salience or an ability to overcome habitual striatal reward mechanisms.

An alternate interpretation of these results is that they actually reflect individual differences in uncertainty aversion. Other than expected reward, a factor frequently demonstrated to influence decision-making is the degree of perceived uncertainty about action outcomes. As discussed in previous chapters, consensus increases the perceived certainty of outcomes, and individuals are more likely to conform when uncertain. Similarly, by continuing to select a known rewarding option in a multi-armed bandit task, individuals are able to avoid uncertainty – especially in a 30-armed bandit task where all slots yield different rewards with unknown probabilities. The literature linking exploitive behavior to uncertainty aversion, however, is mixed. For example, Payzan-LeNestour and Bossaerts (2011) found that exploratory behavior was directed to less uncertain options. Likewise, Daw et. al. (2006) did not find that exploration was directed towards uncertainty. Moreover, Morris et. al. (2016) found that individuals engaged in more exploratory behavior in a gain framed exploration-exploitation task than a loss framed exploration exploitation task. However, there are also experiments showing that individuals strategically explore the least well-known options and evidence of uncertainty signaling in the rostrolateral prefrontal cortex for individuals who engaged in more exploratory behavior. (Frank et al., 2009; Badre et al., 2012). Future work could help arbitrate between uncertainty aversion and reward processing as mechanisms modulating conformity.

Unfortunately, we did not see the same pattern in exploratory behavior as the original experiment and found that exploratory choices were farther skewed towards the upper limit of the possible number of exploratory choices. A possible explanation is that all rewards were fictitious, and participants were less incentivized to reward maximize. Regardless, we were still able to find a significant negative correlation between exploratory behavior and behavioral conformity scores indicating that there may be some general mechanism modulating individual differences rather than the value different individuals attribute to consensus itself.

This negative correlation between exploratory behavior and behavioral conformity scores may also have implications for clinical populations. Specifically, if individual differences in conformity is partially driven by the compulsive reward approach, populations with compulsive disorders, such as addiction or obsessive-compulsive disorder, may exhibit higher levels of conformity. Additionally, high levels of conformity may serve as an early indicator of susceptibility to substance abuse. Further work is necessary to examine whether differences in attribution of incentive salience influence individual differences in conformity and, potentially, conformity in clinical populations.

Chapter 5. General Discussion and Future Directions

These studies contribute to a growing literature on social cognition and reward learning. Specifically, in the first chapter we found an intrinsic affective property of consensus using conditioned reinforcement and ruled out informational sources of conformity. Notably in those experiments, participants were either agreeing or disagreeing with a panel of ostensible jurors and we found an asymmetric change in valence such that our effects were driven by a decrease in likability of contexts paired with dissent. We tentatively attributed this to a negative affective property of dissent which could be related to more general processes such as cognitive dissent.

In chapter two we examined the tradeoff between consensus and expected monetary reward. Specifically, we measured the value of consensus in terms of a willingness to pay and implemented a formal application of reward-related algorithms to conformity. We found that participants selected options associated with lower monetary payoffs more often if that option was also associated with consensus. Additionally, we found that when pairing contexts with consensus among other individuals, we found a symmetric change in valence—an increase in likability of contexts paired with consensus *and* a decrease in likability of contexts associated with dissent—unlike the previous chapter. In both experiments we found that models that treated consensus as a surrogate reward outperformed conventional reinforcement learning models. Future work can examine how social brain areas interact with reward areas to mediate the integration of social and motivational rewards.

In the third chapter we looked at mechanisms that may drive individual differences in conformity. Specifically, we used an exploration-exploitation task to assess whether individual differences in conformity may be a result of individual differences in reward processing and ability to suppress habitual striatal mechanisms. We found a negative correlation between conformity scores and exploratory behavior. These effects may also be driven, however, by uncertainty aversion. Future work is necessary to arbitrate between the roles of uncertainty aversion, cognitive dissonance, and reward learning in consensus-seeking behavior.

While there are several possible sources of affect associated conformity and dissent decisions, including reinforcement history, cognitive dissonance and uncertainty aversion, it is important to note that informational inferences may also shape behavior independently of valence (Deutsch & Gerard 1955; Rafaat et al., 2009; Toelch et al., 2013; Campbell-Meiklejohn et al., 2016). We largely ruled this out in our experiments by including information about accuracy in Chapter 1 and finding in Chapter 2 that retention of reward structures did not predict subsequent conformity. In other circumstances, however, such inferences, potentially formalized by a Bayesian extension of expected value computations, could serve to reduce cognitive effort and reduce uncertainty. Future work can be done to differentiate between informational accounts of conformity and the hedonic aspect of conformity, potentially through the examination of how stress and physiological arousal modulate conformity. In addition, future work can address how observed accuracy, expertise, and majority size modulate effects of conformity.

An important caveat in interpreting the results of these experiments is that all reward outcomes were hypothetical. The use, however, of fictitious rewards is prevalent throughout decision sciences. Furthermore, studies have repeatedly shown equivalent behavioral and neural experimental results across the use of real and hypothetical monetary rewards (Smith & Walker, 1993; Kühberger et al., 2002; Johnson & Bickel, 2002; Madden et al., 2003; Kang et al., 2011). For example, Beattie and Loomes (1997) found that using real vs. fictitious monetary incentives did not significantly alter the common ratio effect—a classical violation of economic axioms—with analogous results reported for temporal discounting, preference reversals and framing effects (Grether, 1979; Kühberger et al., 2002; Madden et al., 2003). Additionally, in our experiments, when balancing for consensus, individuals selected the option associated with a higher monetary payoff more frequently. Thus, while we have not shown that people are willing to incur a real monetary loss to conform, we do provide evidence that they are willing to trade social alignment against a demonstrably rewarding alternative. Further work is needed to explore the influence of real monetary incentives on the effects reported here.

An understanding of how consensus can be used to constrain an action selection problem may provide new optimization criteria for reinforcement learning algorithms modeling the interaction between an agent and its environment and, therefore, have a significant impact on the development of artificial intelligence. Devaluation of consensus can also be an important signal for possible cognitive impairments or psychological disorders. For example, children with autism have been shown to engage in less social interaction and display less attention to social cues and information (Swettenham et al., 1998; Dawson et al., 2004; Moore et al., 2012). Specifically, conformity has been negatively correlated with children's autism quotient score amongst both clinical and nonclinical populations (Yafai et al., 2014). Thus, conformity behavior and valuation of consensus may be useful in developing clinical assessments and further research may be relevant to understanding learning strategies amongst children with autism. Separately, an understanding of how consensus affects evaluation of information may be relevant to our understanding of the current jury systems and the impact of targeted advertisements or articles.

To summarize, in five experiments we demonstrated an affective property of consensus and provided evidence for the role of reward learning in consensus-seeking behavior.

Bibliography

Addicott, M. A., Pearson, J. M., Wilson, J., Platt, M. L., & McClernon, F. J. (2013). Smoking and the bandit: A preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and clinical psychopharmacology*, *21*(1), 66.

Albertella, L., Le Pelley, M. E., Chamberlain, S. R., Westbrook, F., Fontenelle, L. F., Segrave, R., ... & Yücel, M. (2019). Reward-related attentional capture is associated with severity of addictive and obsessive–compulsive behaviors. *Psychology of Addictive Behaviors*, *33*(5), 495.

Anderson, C. M. (2001). *Behavioral models of strategies in multi-armed bandit problems* (Doctoral dissertation, California Institute of Technology).

Anderson, C. M. (2012). Ambiguity aversion in multi-armed bandit problems. *Theory and decision*, *72*(1), 15-33.

Arroyo, M., Markou, A., Robbins, T. W., & Everitt, B. J. (1998). Acquisition, maintenance and reinstatement of intravenous cocaine self-administration under a second-order schedule of reinforcement in rats: effects of conditioned cues and continuous access to cocaine. *Psychopharmacology*, *140*(3), 331-344.

Asaad, W. F., & Eskandar, E. N. (2011). Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *Journal of Neuroscience*, *31*(49), 17772-17787.

Asch, S. E., & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgments. *Documents of gestalt psychology*, 222-236.

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*(3), 595-607.

Banks, J., Olson, M., & Porter, D. (1997). An experimental analysis of the bandit problem. *Economic Theory*, *10*(1), 55-77.

Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, *14*(2), 155-168.

Beharelle, A. R., Polanía, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve exploration–exploitation trade-offs. *Journal of Neuroscience*, *35*(43), 14544-14556.

Berridge, K. C. (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*(3), 391-431.

Besson, C., & Louilot, A. (1995). Asymmetrical involvement of mesolimbic dopaminergic neurons in affective perception. *Neuroscience*, *68*(4), 963-968.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological bulletin*, *119*(1), 111.

Burnham, K., & Anderson, D. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, *20*(13), 1165-1170.

Campbell-Meiklejohn, D., Simonsen, A., Frith, C. D., & Daw, N. D. (2016). Independent neural computation of value from other people's confidence. *Journal of Neuroscience*, 4490-15.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2011). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral cortex*, *22*(11), 2575-2586.

Chen, E. E., Corriveau, K. H., & Harris, P. L. (2013). Children trust a consensus composed of outgroup members—but do not retain that trust. *Child Development, 84*(1), 269-282

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315-12320.

Cooper, J. C., & Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *Neuroimage, 39*(1), 538-547.

Cooper, J. C., Dunne, S., Furey, T., & O'Doherty, J. P. (2012). Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *Journal of cognitive neuroscience*, *24*(1), 106-118.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *20*(3), 372–377.

Cox, S. M., Andrade, A., & Johnsrude, I. S. (2005). Learning to like: a role for human orbitofrontal cortex in conditioned reward. *Journal of Neuroscience*, *25*(10), 2733-2740.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, *40*(2), 271.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*(3), 629.

Festinger, L. (1957). *A theory of cognitive dissonance*. *Scientific American* (Vol. 207).

Falk, E., Way, B., & Jasinska, A. (2012). An imaging genetics approach to understanding social influence. *Frontiers in human neuroscience*, *6*, 168.

Farias, D. D., & Megiddo, N. (2005). Exploration-exploitation tradeoffs for experts algorithms in reactive environments. In *Advances in neural information processing systems* (pp. 409-416).

Flagel, S. B., Akil, H., & Robinson, T. E. (2009). Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. *Neuropharmacology*, *56*, 139-148.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature neuroscience*, *12*(8), 1062.

Gavrilets, S., & Richerson, P. J. (2017). Collective action and the evolution of social norm internalization. *Proceedings of the National Academy of Sciences*, *114*(23), 6068-6073.

Gerard, H. B., Wilhelmy, R. A., & Conolley, R. S. (1968). Conformity and group size. *Journal of Personality and Social Psychology*, 8, 79-82.

Goldberg, S. R. (1973). Comparable behavior maintained under fixed-ratio and second-order schedules of food presentation, cocaine injection or d-amphetamine injection in the squirrel monkey. *Journal of Pharmacology and Experimental Therapeutics*, *186*(1), 18-30.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, *69*(4), 623-638.

Gürçay, B., Mellers, B. A., & Baron, J. (2015). The power of social influence on estimation accuracy. *Journal of Behavioral Decision Making*, *28*(3), 250-261.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, *28*(22), 5623–5630.

Harlé, K. M., Zhang, S., Schiff, M., Mackey, S., Paulus, M. P., & Yu, A. J. (2015). Altered Statistical Learning and Decision-Making in Methamphetamine Dependence: Evidence from a Two-Armed Bandit Task. *Frontiers in psychology*, *6*, 1910.

Harris, C. B., Barnier, A. J., & Sutton, J. (2012). Consensus collaboration enhances group and individual recall accuracy. *The Quarterly Journal of Experimental Psychology*, *65*(1), 179-194.

Hills, Thomas T., Peter M. Todd, David Lazer, A. David Redish, Iain D. Couzin, and Cognitive Search Research Group. "Exploration versus exploitation in space, mind, and society." *Trends in cognitive sciences* 19, no. 1 (2015): 46-54.

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, *96*(4), 651-656.

Horowitz, A. D. (1973). *Experimental study of the two-armed bandit problem* (Doctoral dissertation, University of North Carolina at Chapel Hill).

Hull, C. L. (1943). *Principles of behavior* (Vol. 422). New York: Appleton-century-crofts.

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron*, *78*(3), 563-573.

Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., & Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences*, *107*(51), 22014-22019.

Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., & Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron*, *40*(6), 1251-1257.

Jensen, J., Smith, A. J., Willeit, M., Crawley, A. P., Mikulis, D. J., Vitcu, I., and Kapur, S. (2007). Separate brain regions code for salience vs. valence during reward prediction in humans. *Human brain mapping*, *28*(4), 294-302.

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the experimental analysis of behavior*, *77*(2), 129-146.

Joiner, J., Piva, M., Turrin, C., & Chang, S. W. (2017). Social learning through prediction error in the brain. *NPJ science of learning*, *2*(1), 1-9.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 363-391.

Kang, M. J., Rangel, A., Camus, M., & Camerer, C. F. (2011). Hypothetical and real choice differentially activate common valuation areas. *Journal of neuroscience*, 31(2), 461-468.

Katz, J. L. (1979). A comparison of responding maintained under second-order schedules of intramuscular cocaine injection or food presentation in squirrel monkeys. *Journal of the experimental analysis of behavior*, *32*(3), 419-431.

Kayser, A. S., Mitchell, J. M., Weinstein, D., & Frank, M. J. (2015). Dopamine, locus of control, and the exploration-exploitation tradeoff. *Neuropsychopharmacology*, *40*(2), 454.

Khaw, M. W., Grab, D. A., Livermore, M. A., Vossler, C. A., & Glimcher, P. W. (2015). The measurement of subjective value and its relation to contingent valuation and environmental public goods. *PloS one*, *10*(7), e0132842.

Kim, B. R., Liss, A., Rao, M., Singer, Z., & Compton, R. J. (2012). Social deviance activates the brain's error-monitoring system. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(1), 65-73.

Kitayama, S., Chua, H. F., Tompson, S., & Han, S. (2013). Neural mechanisms of dissonance: An fMRI investigation of choice justification. *Neuroimage*, *69*, 206-212.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*(1), 140-151.

Klucharev, V., Munneke, M. A., Smidts, A., & Fernández, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *Journal of Neuroscience*, *31*(33), 11934-11940.

Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes*, 89(2), 1162-1175.

Latané, B., & Wolf, S. (1981). The social impact of majorities and minorities. *Psychological Review*, *88*(5), 438.

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*(3), 319-338.

Lee, M. D., Zhang, S., & Shi, J. (2011). The wisdom of the crowd playing The Price Is Right. *Memory & cognition*, *39*(5), 914-923.

Levita, L., Hare, T. A., Voss, H. U., Glover, G., Ballon, D. J., & Casey, B. J. (2009). The bivalent side of the nucleus accumbens. *Neuroimage*, *44*(3), 1178-1187.

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. Experimental and clinical psychopharmacology, 11(2), 139.

Maeda, H., & Mogenson, G. J. (1982). Effects of peripheral stimulation on the activity of neurons in the ventral tegmental area, substantia nigra and midbrain reticular formation of rats. *Brain research bulletin*, *8*(1), 7-14.

Mason, M. F., Dyer, R., & Norton, M. I. (2009). Neural mechanisms of social influence. *Organizational Behavior and Human Decision Processes*, *110*(2), 152-159.

Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: the consequences of disagreement. *Journal of Personality and Social Psychology*, *88*(1), 22–37.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339-346.

McGarty, C., Turner, J. C., Oakes, P. J., & Haslam, S. A. (1993). The creation of uncertainty in the influence process: The roles of stimulus information and disagreement with similar others. *European Journal of Social Psychology*, *23*(1), 17-38.

Meyer, R. J., & Shi, Y. (1995). Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem. *Management Science*, *41*(5), 817-834.

Moore, D. J., Heavey, L., & Reidy, J. (2012). Attentional processing of faces in ASD: A dot-probe study. *Journal of autism and developmental disorders*, *42*(10), 2038-2045.

Morris, L. S., Baek, K., Kundu, P., Harrison, N. A., Frank, M. J., & Voon, V. (2016). Biases in the explore–exploit tradeoff in addictions: the role of avoidance of uncertainty. *Neuropsychopharmacology*, *41*(4), 940.

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

Nook, E. C., & Zaki, J. (2015). Social norms shift behavioural and neural responses to foods. *Journal of Cognitive Neuroscience*, *7*(7), 1412–1426.

O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, *49*(1), 157-166.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329-337.

Oyama, K., Hernádi, I., Iijima, T., & Tsutsui, K. I. (2010). Reward prediction error coding in dorsal striatal neurons. *Journal of neuroscience*, *30*(34), 11447-11457.

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS computational biology*, *7*(1), e1001048.

Peters, J., & Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural brain research*, *213*(2), 135-141.

Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: attitude clarity and attitude correctness. *Journal of personality and social psychology*, *92*(1), 30.

Plassmann, H., O'Doherty, J., & Rangel, (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of neuroscience*, *27*(37), 9984-9988.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in cognitive sciences*, *13*(10), 420-428.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Reiter, A. M., Heinze, H. J., Schlagenhauf, F., & Deserno, L. (2017). Impaired flexible reward-based decision-making in binge eating disorder: evidence from computational modeling and functional neuroimaging. *Neuropsychopharmacology*, *42*(3), 628.

Reysen, S., & Branscombe, N. R. (2008). Belief in collective emotions as conforming to the group. *Social Influence*, *3*(3), 171-188.

Rosenberg, L. A. (1961). Group size, prior experience, and conformity. *Journal of Abnormal and Social Psychology*, 63, 436-437.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*, *27*(18), 4826-4831.

Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P., & Dolan, R. (2012). Serotonin selectively modulates reward value in human decision-making. *Journal of Neuroscience*, *32*(17), 5833-5842.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*, *187*, 60.

Sherif, M., & Harvey, O. J. (1952). A study in ego functioning: Elimination of stable anchorages in individual and group situations. *Sociometry*, *15*(3/4), 272-305.

Shestakova, A., Rieskamp, J., Tugin, S., Ossadtchi, A., Krutitskaya, J., & Klucharev, V. (2012). Electrophysiological precursors of social conformity. *Social cognitive and affective neuroscience*, *8*(7), 756-763.

Skinner, B. F. (1938). The behavior of organisms: an experimental analysis. Appleton-Century.

Smith, J. R., Hogg, M. A., Martin, R., & Terry, D. J. (2007). Uncertainty and the influence of group norms in the attitude–behaviour relationship. *British Journal of Social Psychology*, *46*(4), 769-792.

Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, *31*(2), 245-261.

Steyvers, M., Lee, M. D., & Wagenmakers, E. J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*(3), 168-179.

Sun, S., & Yu, R. (2016). Social conformity persists at least one day in 6-year-old children. *Scientific reports*, *6*, 39588.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological review*, *88*(2), 135.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of pavlovian reinforcement.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., ... & Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry*, *39*(05), 747-753.

Toelch, U., Bach, D. R., & Dolan, R. J. (2013). The neural underpinnings of an optimal exploitation of social information under uncertainty. *Social cognitive and affective neuroscience*, *9*(11), 1746-1753.

Tomlin D, Nedic A, Prentice DA, Holmes P, Cohen JD (2013) The Neural Substrates of Social Influence on Decision Making. *PLoS ONE 8*(1): e52630.

Toyokawa, W., Kim, H. R., & Kameda, T. (2014). Human collective intelligence under dual exploration-exploitation dilemmas. *PloS one*, *9*(4), e95789.

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science*, *303*(5666), 2040-2042.

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

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*(3-4), 279-292.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: collaborative and individual processes in remembering. *Journal of experimental psychology: Learning, memory, and cognition*, *23*(5), 1160.

Williams, B. A. (1994). Conditioned reinforcement: Neglected or outmoded explanatory construct?. *Psychonomic Bulletin & Review*, *1*(4), 457-475.

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, *58*(6), 967-973.

Yafai, A. F., Verrier, D., & Reidy, L. (2014). Social conformity and autism spectrum disorder: A child-friendly take on a classic study. *Autism*, *18*(8), 1007-1013.

Yu, R., & Sun, S. (2013). To conform or not to conform: spontaneous conformity diminishes the sensitivity to monetary outcomes. *PloS one*, *8*(5), e64530.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological science*, *22*(7), 894-900.

Zink, C.F., Pagnoni, G., Martin, M.E., Dhamala, M., & Berns, G.S. (2003). Human striatal response to salient nonrewarding stimuli. *Journal of Neuroscience, 23*(22), 8092-8097.

Zink, C. F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., & Berns, G. S. (2006). Human striatal activation reflects degree of stimulus saliency. *Neuroimage*, *29*(3), 977-983.

Zubarev, I., Klucharev, V., Ossadtchi, A., Moiseeva, V., & Shestakova, A. (2017). MEG signatures of a perceived match or mismatch between individual and group opinions. *Frontiers in neuroscience*, 11.