The Intrinsic Cost of Dissent

Prachi Mistry (prachim@uci.edu), Mimi Liljeholm (m.liljeholm@uci.edu)
Department of Cognitive Sciences, UC Irvine
Irvine, CA 92697

Abstract

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. 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.

Keywords: Conformity; Dissent; Reinforcement Learning; Decision Making; Conditioned Reinforcement

Introduction

Social animals must often reach a consensus with other members of their group in making collective decisions. By agreeing with a majority opinion, individuals are able to avoid social rejection and retain access to group resources (Bond & Smith, 1996; Reysen & Branscombe, 2008). Moreover, relying on a group majority often yields superior memory retrieval (Harris et al., 2012), improved perceptual judgment accuracy (Gürçay et al., 2014), and greater monetary pay-offs in gambling tasks (Toyokawa et al., 2014). While consensus seeking in the face of such conspicuous contingent gain is unsurprising, individuals also consistently conform to a group majority in the absence of any apparent social or monetary reward (e.g., Sherif, 1936; Klucharev et al., 2009; Corriveau et al., 2009; Chen et al., 2013; Nook & Zaki, 2015), suggesting that the act of reaching consensus might have *intrinsic value*. Whether due to evolutionary pressure or an individual's reinforcement history, a basic but untested prediction of reinforcement learning (RL) theory is that, once established, such intrinsic value should "rub off" on stimuli associated with high levels of consensus. 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). We report two experiments testing this prediction.

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). For example, in a recent neuroimaging study by Nook & Zaki (2015), participants rated the perceived likability of various foods and, after each rating, viewed the ostensible average rating of a large group of peers. After some delay, participants were given an opportunity to re-rate each food item. 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. Nonetheless, participants significantly shifted their follow-up food ratings – subjective judgments for which no "correct" answer existed – in the direction of the group norm. Moreover, neural activity in the ventral striatum, a brain region heavily implicated in processing reward (McClure et al, 2003; O'Doherty, 2004), increased with consensus relative to disagreement between individual and group ratings, and predicted subsequent adjustments towards the group norm. These, and similar results (Klucharev et al., 2009; Zaki et al., 2011; Campbell-Meiklejohn et al., 2010, 2012; Yu & Sun, 2013; Zubarev 2017), strongly suggest that conformity to a group majority may be intrinsically rewarding.

In spite of ample evidence of apparently inconsequential conformity, it is problematic to conclude that consensus-seeking decisions are rewarding simply because such decisions are made. 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; first, since those same neural regions also respond to valence-neutral but surprising, or otherwise salient, stimuli (Horvitz, 2000; Zink et al., 2003, 2006; Jensen et al., 2007; Levita et al., 2009) and second, because neural signals identified in social conformity studies often appear more consistent with error adjustment than with hedonic reinforcement (e.g., Zaki et al., 2011). 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 suggests 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. Following Matz and Wood (2005), we employ a "mock jury" scenario to generate majority judgments with which a participant may agree or disagree.

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. In Experiment 1, we tested this prediction 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.

Method

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: The task is illustrated in Figure 1. At the start of the experiment participants were told that they would be acting as a juror in a series of cases in various courtrooms. They were further told that, in preparation for making decisions on cases themselves, they would first have an opportunity to study some previously adjudicated cases. All cases were potential misdemeanors under the California Vehicle Code, punishable by incarceration for 6 months or less. Cases were constructed such that all defendants had violated the California Vehicle Code but without necessarily being held liable (e.g. driving five miles per hour above the legal speed limit). Prior to starting a learning phase, participants were asked to rate, in random order, the likability of four differently colored courtrooms, on a scale from 0 (not at all likeable) to 10 (extremely likeable), with 5 indicating neutral affect.

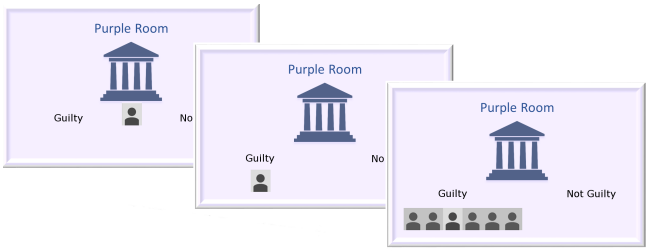

Figure 1. Trial illustration showing the initial choice screen, the participant's culpability choice, and the verdict of a unanimous jury together with the choice of the participant. Case summaries and response prompts (see text) are not shown in the figure.

On each trial in the learning phase, participants were first presented with a short synopsis describing the particular case while one of the four courtrooms was 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 believed the defendant was guilty or not. A grey avatar representing the participant would move beneath the relevant, "guilty" or "not guilty", label based on the participant's response. They were then prompted to press the spacebar to see the jury's verdict, which was represented by five darker shade grey icons appearing beneath the relevant label (screen 2 in Figure 1). 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 was the opposite of the participant's ~90% of the time. The colors of consensus and dissention rooms were counterbalanced across participants. There were 12 trials in each of the four distinctly colored courtrooms, for a total of 48, randomly ordered, trials. The 48 specific case descriptions were randomly distributed across trials.

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 would serve as jury members themselves. They were instructed that none of the previously observed jurors would serve on any juries of which the participant might be a member. On each of eight trials, participants first selected between two courtrooms: one was always a consensus room, in which previous juries had frequently agreed with them, and the other was a dissention room, in which the previous juries

had frequently disagreed. Once they entered the chosen courtroom, they were presented with a case and asked to provide a guilty or not guilty judgment. They were not shown the decisions made by other jurors, nor the final verdict, for any of these eight cases. To assess explicit memory of consensus and dissent courtrooms, at the end of the experiment participants were asked to rate, for each courtroom, the degree to which the jury had agreed with them in that room during the initial learning phase, on a scale of 0 (never) to 10 (always).

Results

Likability ratings: Mean (post-pre) likability ratings are shown on the left side of Figure 2. We predicted that likability for courtrooms in which participants' judgments had frequently agreed with those of the unanimous jury, would be greater than that for courtrooms in which they had frequently dissented. 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, $p<0.01$, $d=0.139$. Notably, this difference was not due to an increased likability of the consensus rooms, but to a decreased likeability of the dissention rooms. A significant difference between mean pre- and post-learning ratings was observed for dissention rooms (-0.75 ± 1.51), t(19)=2.22, $p<0.05$, $d=1.509$, but not for consensus rooms (0.18 ± 1.45), t(19)=0.54, $p=0.60$, $d=1.453$.

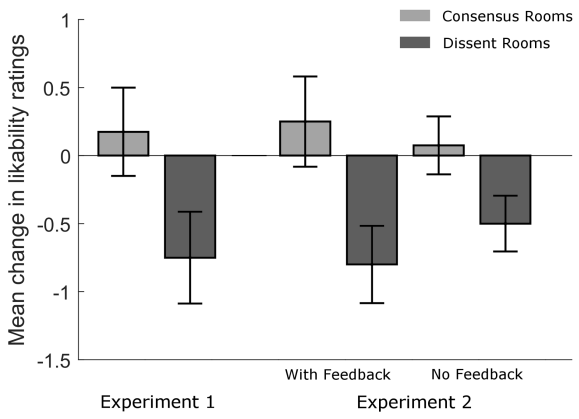

Figure 2. Likability ratings (post-pre learning) for consensus and dissent courtrooms from Experiment 1 (left) and from the two groups in Experiment 2 (right).

Choice preference: We further hypothesized that there would be a significant preference for deliberating in consensus rooms over dissention rooms, in spite of instructions emphasizing 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, $p<0.05$, $d=0.19$.

Explicit recall of consensus and dissent*:* Finally, using a scale on which participants indicated the degree to which the juries had agreed with them in a particular courtroom during the training phase, we confirmed that participants were able to accurately distinguish between consensus (5.73 ± 1.45) and dissention (4.60 ± 1.52) rooms, t(19)=2.22 *p<.05, d=1.877.* 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 $r=-0.08$, $p=0.73$.

Experiment 2

The results of Experiment 1 suggest that arbitrary stimuli repeatedly paired with dissent from a unanimous majority can acquire negative valence. We interpret these effects in terms of an *intrinsic aversive property* of dissent that is transferred to concomitant stimuli according to basic reinforcement learning mechanisms. However, an alternative possibility is that participants assumed that the unanimous jury was always correct, and that the transferred negative valence was elicited by the perception of being wrong, rather than by dissent per se. In other words, the results reflected an *informational* (Deutsch & Gerard, 1955) rather than normative basis of apparent social valence. In Experiment 2, we address this possibility by including feedback about the "true" culpability of the defendants in our hypothetical court cases.

Method

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 the experiment. For the first (No Feedback) group, the task and procedures were identical to those in Experiment 1. For the second (Feedback) group, the task and procedures were identical to those of the "No Feedback" group, with the following exception: 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 panel representing the jury was absent, the icon representing the participant remained under the selected "guilty" or "not guilty" label, and a selection square appeared around the label indicating the actual culpability of the defendant; in

other words, this screen was identical to the 2nd screen in Figure 1, but with the addition of a selection square around the "accurate" culpability label. 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. At the very end of the study, participants in both groups were asked to explain the reasons behind their likability ratings.

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, $F(1,38)=9.73$, $p<0.005$, but no effect of feedback (i.e., group) and no interaction, $F's<0.84$. 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%, $p<0.05$) and "no feedback" (68%, $p<0.005$) 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, $p's>0.12$, nor did the degree of discrimination between consensus and dissention rooms significantly predict changes in likability ratings, in either group, $p's>0.23$.

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%).

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 were contexts repeatedly paired with dissent. In the second study, these evaluative differences emerged even when explicit feedback provided the "correct" answer, suggesting that the valence associated with agreement or dissent was not solely due to perceived accuracy. Moreover, across studies, evaluative changes were driven, not by an increased likability of contexts repeatedly paired with consensus, but by a decreased likeability of contexts paired with dissent. Although 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, we tentatively interpret our findings as evidence for an intrinsic cost of dissent.

There are several possible sources of negative affect associated with dissent. First, as noted, by diverging from a majority opinion, individuals may be subject to social rejection, lose access to group resources, and make inferior perceptual and economic decisions. Thus, from a reinforcement learning (RL) perspective, the act of dissenting from a group majority may acquire negative valence through a history of being paired with aversive outcomes. Alternatively, the negative valence may not be directly related to dissent, but instead accompany more general processes. For example, lack of consensus has been proposed to elicit 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). Moreover, several studies have demonstrated that conforming to a consensus reduces perceived uncertainty about decision outcomes (McGarty et al., 1993; Smith et al., 2007; Petrocelli et al. 2007; Sherif & Harvey,1952), suggesting that negative affect associated with dissent might be related to uncertainty aversion (Kahneman and Tversky, 1979). 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 queried, most participants attributed the likability of contexts to their colors (see Nisbett & Wilson, 1977 for a discussion on when participants may identify experimental manipulations as influential stimuli), and there was no correlation between memory of which context had been paired with dissent and changes in context-likability. 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.

Our results are also generally consistent with demonstrations of a decreased neural signal in the ventral striatum (VS), an area frequently implicated in model-free reward learning (O'Doherty et al., 2003), when participants make decisions that dissent from a group norm (e.g., Klucharev et al., 2009; Zaki et al., 2011; Nook & Zaki, 2015). Interpretation of such VS signals are complicated, however, by the bi-directionality of dissent employed by the relevant studies: a group rating of a stimulus' subjective value (e.g., the attractiveness of a face or desirability of a food) may be either greater or lesser than a participant's rating. While both types of deviation have been shown to deactivate the VS as group norms are revealed, Zaki et al. (2011) found that, during subsequent re-exposure to rated stimuli, activity in the VS scaled with the *signed* difference between the participant's rating and the group norm. Such signed signals could reflect new stimulus values that had been error-adjusted towards the group reference, or a retrieval of the previously experienced divergence from the group norm. They are not consistent, however, with a hedonic reinforcement signal, which should simply increase the value of stimuli paired with the positive hedonics of consensus decisions, and decrease the value of stimuli associated with the aversiveness of dissent. Notably, conventional demonstrations of reinforcement signaling in the VS primarily entail increased activity in response to unexpected reward and, critically, the transfer of such responses to stimuli associated with reward – that is, an increased signal in response to a stimulus that is repeatedly paired with a rewarding outcome (e.g., O'Doherty et al., 2003) – consistent with the affective changes demonstrated here. Further work is needed to determine how the current results, and the formal framework of reinforcement-learning more broadly, relate to recently demonstrated neural correlates of social conformity

In summary, we found that contexts repeatedly paired with dissent from a unanimous majority were less likable and less preferred given forced choice than were contexts paired with consensus. These 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.

Acknowledgements

This work was supported by a start-up fund from the University of California, Irvine to Mimi Liljeholm.

References

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.

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. K., Simonsen, A., Jensen, M., Wohlert, V., Gjerløff, T., Scheel-Kruger, J., ... & Roepstorff, A. (2012). Modulation of social influence by methylphenidate. *Neuropsychopharmacology*, *37*(6), 1517-1525.

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.

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

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)

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.

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.

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

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

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

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.

Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: the consequences of disagreement. *Journal of personality and social psychology*, *88*(1), 22.

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.

Nisbett, r. E., & Wilson, T. D. (1997). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.

Nook, E. C., & Zaki, J. (2015). Social norms shift behavioral and neural responses to foods. *Journal of cognitive neuroscience*.

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.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current opinion in neurobiology*, *14*(6), 769-776.

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.

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

Sherif, M. (1936). The psychology of social norms.

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. (2013). Electrophysiological precursors of social conformity. *Social cognitive and affective neuroscience*, *8*(7), 756-763.

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.

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

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

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*.

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*.