NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

As a manuscript

Vasily Klucharev

NEUROCOGNITIVE MECHANISMS OF THE SOCIAL INFLUENCE
(MECHANISMS OF CONFORMITY)

Summary
for the purpose of obtaining academic degree of
Doctor of Science in Cognitive Sciences

Moscow, 2024

The dissertation was prepared at the Institute for Cognitive Neuroscience, HSE university.

Seven published articles were selected for the defense:

1. Klucharev V., Hytönen K., Rijpkema M., Smidts A., Fernández G. Reinforcement learning signal predicts social conformity. Neuron. 2009. 61(1). 140-151. doi.org/10.1016/j.neuron.2008.11.027
2. Klucharev V., Munneke M.A., Smidts A., Fernandez G. Downregulation of the posterior medial frontal cortex prevents social conformity. Journal of Neuroscience. 2011. 31. 11934-11940. doi.org/10.1523/jneurosci.1869-11.2011
3. Shestakova A., Rieskamp J., Tugin S., Ossadtchi A., Krutitskaya J., Klucharev V. Electrophysiological precursors of social conformity. Social Cognitive and Affective Neuroscience. 2013. 8(7). 756-763. doi.org/10.1093/scan/nss064
4. Klucharev V., Zubarev I., Shestakova A. Neurobiological mechanisms of social influence. Experimental Psychology (Russia) 2014. 7(4). 20–36. psyjournals.ru/en/journals/exppsy/archive/2014_n4/72898
5. Zinchenko O., Klucharev V. Commentary: The Emerging Neuroscience of Third-Party Punishment. Frontiers in Human Neuroscience. 2017. 11. 1-3 doi.org/10.3389/fnhum.2017.00512
6. Zubarev I., Klucharev V., Ossadtchi A., Moiseeva V., Shestakova A., MEG Signatures of a Perceived Match or Mismatch between Individual and Group Opinions. // Frontiers in Neuroscience. 2017. 10(11). 1-9. doi.org/10.3389/fnins.2017.00010
7. Gorin A., Klucharev V., Ossadtchic A., Zubarev I., Moiseeva V., Shestakova A. MEG signatures of long-term effects of agreement and disagreement with the majority. Scientific reports. 2021. 11(1). 1-10. doi.org/10.1038/s41598-021-82670-x

The results are also published in the following articles on this topic:

8. Klucharev V., Smidts A., Fernández G. Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. Social Cognitive and Affective Neuroscience. 2008. 3(4). 353-66. doi.org/10.1093/scan/nsn022

9. Stallen M., Smidts A., Rijpkema M., Klucharev V., Fernández G. Celebrities and shoes on the female brain: The neural correlates of product evaluation in the context of fame. Journal of Economic Psychology. 2010. 31(5). 802-811. doi.org/10.1016/j.joep.2010.03.006

10. Huber R., Klucharev V., Rieskamp J. The neural underpinnings of Informational Cascades — An fMRI study of probability updating in a social context. Soc Cogn Affect Neurosci. 2015. 10(4). 589-97. doi.org/10.1093/scan/nsu090

11. Zinchenko O., Belianin A., Klucharev V. The Role of the Temporoparietal and Prefrontal Cortices in a Third-Party Punishment: A tDCS Study. Psychology. Journal of Higher School of Economics. 2019. 16(3). 529-550. doi.org/10.17323/1813-8918-2019-3-529-550

12. Jaaskelainen I., Klucharev V., Panidi K., Shestakova A. Neural Processing of Narratives: From Individual Processing to Viral Propagation. Frontiers in Human Neuroscience. 2020. 14. article 253. doi.org/10.3389/fnhum.2020.00253

13. Zinchenko O., Savelo O., Klucharev V. Role of the prefrontal cortex in prosocial and self-maximization motivations: an rTMS study. Scientific reports. 11(1). 2021. 1-11. doi.org/10.1038/s41598-021-01588-6

14. Zinchenko O., Nikulin V., Klucharev V. Wired to Punish? Electroencephalographic Study of the Resting-state Neuronal Oscillations Underlying Third-party Punishment. Neuroscience. 471. 1-10. 2021. doi.org/10.1016/j.neuroscience.2021.07.012

15. Ntoumanis I., Panidi K., Grebenschikova Y, Shestakova A., Kosonogov V., Jääskeläinen I., Kadieva D., Baran S., Klucharev V. "Expert persuasion" can decrease Willingness to Pay for sugar-containing food. Frontiers in Nutrition. 2022. 9. article 926875. doi.org/10.3389/fnut.2022.926875

16. Ntoumanis I., Davydova A., Sheronova Y., Panidi K., Kosonogov V., Shestakova A., Jaaskelainen I., Klucharev V. Neural mechanisms of expert persuasion on willingness to pay for sugar // Frontiers in Behavioral Neuroscience. 2023. Vol. 17. No. 1147140. P. 1-13. doi.org/10.3389/fnbeh.2023.1147140

I. Introduction

General research problem

Historically, neuroscience has focused on the neuronal mechanisms of individual behavior. Relatively recently, social psychology, in collaboration with newly developed social neuroscience, made an attempt to clarify the social factors shaping individual behavior, decisions, and attitudes (Ajzen & Fishbein, 1980; Cacioppo & Decety, 2011; Cialdini & Goldstein, 2004). Some theoretical accounts indicate that *homo sapiens* display "superorganismic" properties and provocatively suggest that we have made a dramatic evolutionary transition from "primate colonies" to a "superorganismic" unit of individuals (Foster & Ratnieks, 2005; Kesebir, 2012; D. S. Wilson & Wilson, 2007). Indeed, people form very complex social groups that are self-organized by a system of communication and complex social interactions. Furthermore, social groups are vital for our survival and normal development. Thus, such a complex "superorganismic" unit requires cognitive mechanisms to integrate individuals, synchronize their actions, increase group coherence, avoid conflicts, reduce within-group behavioral variations, and increase between-group differentiation (Kesebir, 2012). To achieve this, evolution via multilevel selection may favor not only our tendency to cooperate but also our tendency to follow group norms that would require a neurocognitive mechanism of social conformity (E. Wilson & Wilson, 2008). Social psychology has documented the profound effect of social norms on human behavior (Cialdini & Goldstein, 2004), but the neurocognitive mechanism of conformity remains largely unknown. The present dissertation summarizes the studies of our research group on conformity and suggests and explores a general mechanism of social conformity.

Following David Marr's three levels of analysis, here, we suggest the following three explanatory levels of conformity: (1) Conformity is an evolutionarily advantageous form of social learning that increases group coherence; (2) The brain detects deviations from descriptive norms and generates a reward prediction error-like (learning) signal; (3) Large-scale brain networks integrate the learning signal with personal and social contexts and convert it to the behavioral adjustments. Importantly, the current dissertation focusses

primarily on the statement (2), while supporting the statement (1) and (3) based on the literature review.

Theoretical basis of the current work

The set of studies presented in the current dissertation adopts theoretical accounts of social psychology, social neuroscience, and learning theory. First, our studies focused on the concept of social norms—one of the key concepts in social psychology and social sciences (e.g., (Rivis & Sheeran, 2003; J. R. Smith & Louis, 2009)). Social psychology differentiates *injunctive* and *descriptive norms*. Injunctive norms or "behavioral standards" are prescriptions of the referent social group, but descriptive norms represent typical behavioral patterns of the referent social group (Cialdini et al., 1990; Kallgren et al., 2000). Injunctive and descriptive norms do not necessarily match each other. For example, an injunctive norm could prohibit corruption, whereas a descriptive norm could signal the prevalence of some forms of corrupt behavior. Importantly, descriptive norms are particularly effective predictors of social behavior, including healthy eating and drinking behavior, energy conservation, and recycling ((Cialdini et al., 1990), see (Manning, 2009), for a meta-analytic review). However, this raises an important theoretical question: *By which cognitive mechanism do we detect and learn descriptive norms?*

Second, our studies adopted assumptions of social neuroscience, suggesting that all social behavior is implemented (neuro)biologically (Cacioppo & Decety, 2011). Social neuroscience adopts theoretical assumptions of *constitutive reductionism,* a systematic approach to studying the parts of the system and the interplay across parts/levels of the system to better understand the whole system (Cacioppo et al., 2000). The social neuroscience approach raises another important theoretical question: *Which neurobiological mechanism underlies the effects of descriptive norms on human behavior and how is it implemented in neural architecture?*

Finally, we employed r*einforcement learning theory*, which models the effects of experience on value functions and subsequent choices (e.g., (Rescorla & Wagner, 1972)). The reinforcement learning theory assumes that organisms encode an *action value*

function as the sum of future outcomes of an action (*Q(s,a)*, *s* = the state of the environment; *a* = the particular action) and *state value function* (*V(s)*) as the sum of rewards that are expected in the state of the environment. Here, reinforcement learning refers to the idea that to reduce errors, value functions are continuously updated according to the received rewards. The theory hypothesizes a learning signal—*reward prediction error* (RPE)—that encodes a difference between the actual and the expected reward (Sutton & Barto, 1998; for a similar approach, see Bechtereva et al., 2005). *Model-free* reinforcement learning algorithms assume that outcome-based RPE is the key mechanism for updating values. However, *model-based* reinforcement learning algorithms suggest that the values could also be updated prior to outcomes, based on relevant information regarding the motivational state or environment. Reinforcement learning theory also admits that organisms can simulate possible outcomes and learn, for example, based on the *fictive reward prediction error—the* difference between hypothetical and predicted outcomes (Boorman et al., 2011; Lohrenz et al., 2007). Overall, reinforcement learning provides an effective framework for modeling adaptive decision-making processes and raises another important theoretical question: *Can conformity to descriptive norms be based on a reinforcement learning-like (and essentially not uniquely social) mechanism implemented in the dopaminergic system of the human brain?*

Summary of scientific novelty

1. The current work invented a novel research paradigm—a face judgment *conformity task* that induces well-controlled conflicts between the subject's own judgment and the group opinion in neuroimaging settings.

2. Our functional magnetic resonance imaging (fMRI) data, for the first time, showed that the activity of the posterior medial frontal cortex (pMFC) and ventral striatum (a) reflects social influence and (b) predicts subsequent adjustments of opinion, in line with that of the descriptive group norm.

3. Our transcranial magnetic stimulation (TMS) results, for the first time, demonstrated the causal role of pMFC in conformal adjustments of opinion, in line with that of the descriptive group norm.

4. Our magnetoencephalography (MEG) study of conformity, the first of its kind, showed that conflicts with the descriptive norms modulate activity of the posterior cingulate cortices (PCC, including precuneus), right temporoparietal junction (TPJ), ventromedial prefrontal cortex (vMPFC), bilateral anterior cingulate cortices (ACC), and right superior occipital gyrus.

5. Our electroencephalography (EEG) study of conformity, for the first time, demonstrated temporal dynamics of neurocognitive correlates of conformity as the cascade of neuronal responses to perceived conflicts with the group norms, from a frontal negativity reflecting a conflict with the group opinion to a later evoked response (peaking at 380 ms), reflecting a conformal behavioral adjustment in line with the descriptive group norm.

6. For the first time, we showed the MEG markers of the long-lasting effect of group pressure on the processing of visual information.

Theoretical significance

Our articles suggested for the dissertation elaborate on a theoretical neurocognitive mechanism of social influence. Our results indicate that decision values are adjusted in line with descriptive norms based on a fundamental rather than a uniquely social mechanism, similar to reinforcement learning. We suggest that a conflict with a group opinion might generate a “social” RPE-like signal. More precisely, a difference between an individual and the (in)group behavior (or judgments) could be perceived as an error. According to this framework, the ventral striatum and pMFC generate the RPE-like learning signal to perceived deviations from descriptive norms. Such dopamine-related activity triggers a learning mechanism via interaction with other large-scale brain networks that “put” the learning signal into a broader social and personal context. Importantly, our theoretical framework suggests that the neural error signal detecting deviations from descriptive norms shares the same neurocognitive mechanism as the standard RPE underlying reinforcement learning.

Applied significance

A better understanding of social influence and conformity is critical to a more precise control of many forms of maladaptive behavior. Daily, descriptive norms provide information about appropriate smoking (Schofffild et al., 2001), drinking (Johnston & White, 2003), healthy eating (Louis et al., 2007), and environment-relevant behaviors, such as littering (Reno et al., 1993), recycling (White & Hyde, 2012), and energy conservation (Goldstein et al., 2008; P. W. Schultz et al., 2007).

Often, peers initiate the use and abuse of drugs or alcohol by adolescents. For example, the greater the popularity of drinking and marijuana use among friends, the more likely adolescents are to drink or use marijuana (Burkett, 1977). Overall, many studies have suggested that individual sensitivity to group pressure, in combination with strong peer pressure, leads to delinquent behavior. Therefore, our studies of the neurocognitive mechanisms of conformity will stimulate the development of effective social interventions that control or prevent maladaptive behavior.

Further, the effectiveness of social norms in encouraging pro-environmental behaviors (Bodin, 2017; Byerly et al., 2018; Centola et al., 2018; Nyborg et al., 2016; Otto et al., 2020) suggest that our experimental and theoretical data can be used to design effective social norm interventions, thus providing a tool for policymakers to promote behavior that is beneficial to the environment (e.g., (Beretti et al., 2013)).

Statements for the defense

1. The activity of the pMFC and ventral striatum, measured with fMRI, reflects mismatches between individual judgments and group judgments (descriptive norms). This activity is predictive of subsequent conformal adjustments of opinion in line with group judgments (descriptive norms).

2. The pMFC plays a causal role in conformal adjustments to group judgments (descriptive norms) as probed with TMS.

3. Mismatches between individual judgments and group judgments (descriptive norms) trigger a cascade of neuronal responses, including earlier frontocentral response (peaking at 200 ms, similar to feedback-related negativity) that reflects a conflict with

descriptive norms to a later evoked response (peaking at 380 ms) that reflects a conformal behavioral adjustment.

4. Partially distinct neural circuits monitor matches and mismatches of individual judgments with group judgments (descriptive norms). Mismatches of individual judgments with group judgments evoke activity of the anterior and posterior medial prefrontal cortices, as well as activity of the TPJ and vMPFC. However, matches with group judgments evoke an increase in the amplitude of beta oscillations (13–30 Hz) in the anterior and vMPFC.

5. Conformity to group judgments (descriptive norms) is based on a neurocognitive learning mechanism that shows features of reinforcement learning and is implemented in the pMFC and ventral striatum in interaction with large-scale brain networks.

Data collection and author contribution statement

Five out of seven articles selected for the defence report psychophysiological fMRI, EEG, TMS and MEG studies. Overall, selected papers describe six laboratory studies of over 120 participants. The laboratory experiments were run at the Institute of Cognitive Neuroscience (HSE university, Moscow, Russia) – MEG and theoretical studies, Donders Institute for Brain, Cognition and Behaviour (The Netherlands) in collaboration with Erasmus Research Institute of Management (Erasmus University Rotterdam, The Netherlands) – fMRI and TMS studies, Faculty of Psychology (University of Basel, Switzerland) in collaboration with Saint Petersburg State University (Russia) – EEG study.

The author confirms contribution to the papers selected for the defense as follows:

- Conception, design, supervision of the studies (Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013; Zubarev et al., 2017; Gorin et al., 2021; Klucharev et al., 2014; Zinchenko, Klucharev, 2017)
- Data collection (Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013)

- Data analysis and interpretation (Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013; Zubarev et al., 2017; Gorin et al., 2021)
- Drafting and critical revisions the articles (Klucharev et al., 2009; Klucharev et al., 2011.; Shestakova et al., 2013; Zubarev et al., 2017; Gorin et al., 2021; Klucharev et al., 2014; Zinchenko, Klucharev, 2017)

Reliability of the results and conclusions, public presentations on the topic and grant support

The reliability of the results has been confirmed by the required number of observations and modern neuroimaging methods. The scientific results and conclusions of the dissertation are based on actual empirical data reported in a number of peer-reviewed publications. Statistical analysis and interpretation of the results were carried out using contemporary methods of neuroimaging data processing and statistical analysis. All the results are supported by statistically significant tests at 95% significance level.

The results of the dissertation were publicly presented in more than 20 talks and poster sessions at conferences in Russia and worldwide, including Congress on Brain, Behavior and Emotions (Brazil, 2022), Society for NeuroEconomics Conference (2005, 2007, 2008, 2009, 2012, 2015, 2017), Volga Neuroscience Meeting (Russia, 2018), XVI European Congress of Psychology (Russia, 2019), Annual Cognition, Computation, Communication and Perception Conference (Russia, 2015), Colloquium lecture Ecole Normale Supérieure, Département d'Études Cognitives (France, 2013), 3rd International Conference on Neuroeconomics and Neuromanagement (China, 2012), 1st Conference of the European Society for Cognitive and Affective Neuroscience (France, 2012), Colloquium lecture Brain & Cognition Seminar, Neuroscience Center, Université de Genève (Switzerland, 2012), Social Psychology Colloquium, University of Basel (Switzerland, 2011), 4th International Conference on Cognitive Science (Russia, 2010), 8th Dutch Endo-Neuro-Psycho Meeting (Netherlands, 2009), 3rd International Conference on Cognitive Science (Russia, 2008).

The studies were supported by FACI (Federal Agency for Science and Innovation, Russia), 2010-2012 and 2012-2013, Erasmus Institute of Management, Swiss National Science Foundation (SNSF), 2011-2013, Russian Academic Excellence Project “5–100”, by the International Laboratory for Social Neuroscience of the Institute for Cognitive Neuroscience HSE, RF Government Grant No. 075-15-2019-1930 and has been carried out using HSE university unique equipment (Reg. num 354937).

II. Concepts of social conformity and descriptive norms

Conformity is a type of social influence in which individuals change their attitudes, beliefs, and behaviors in line with the reference group without an explicit request. Conformity strongly affects various forms of human behavior, from criminal to pro-environmental behavior (Cialdini & Goldstein, 2004; O'Keefe, 2002). Importantly, people are usually unaware of this strong tendency to conform to group norms (Bryan & Test, 1967). Already two-year-olds (Haun et al., 2012, 2014) and preschoolers (Sun & Yu, 2016) show a strong tendency to conform to the majority. Conformity was observed in various species: fruit flies (Danchin et al., 2018), fish (Day et al., 2001; Pike & Laland, 2010), rats (Galef & Whiskin, 2008; Konopasky & Telegdy, 1977), monkeys (Dindo et al., 2009), and great apes (Whiten et al., 2005). Furthermore, a genome-wide association study suggested the role of specific genes in social conformity (Chen et al., 2018). Overall, previous behavioral and genetic findings indicate an evolutionary basis of conformity, suggesting that natural selection favors the behavioral tendency to conform to group norms. Thus, evolution may select an "automatic" neurocognitive learning mechanism that continuously adjusts our attitudes, beliefs, and behavior in line with group norms. Indeed, conformist transmission of beliefs or behaviors is adaptive because it allows for the integration of the outcomes of multiple individuals (Boyd et al., 2005).

Social psychology suggests that conformity could be induced by injunctive or descriptive norms (Cialdini & Goldstein, 2004). Moral injunctive norms signal what people have to do, whereas descriptive norms signal what the majority of people actually do, regardless of the injunctive norms. Unsurprisingly, people demonstrate a tendency to conform to injunctive norms, which are often reinforced by various types of social punishments and rewards. The strong impact of descriptive norms on human behavior is more surprising since they have a more informational nature: normally, they simply signal the most popular behavioral strategy.

The classic experiments of the pioneering Asch study (Asch, 1951, 1955) showed that individuals frequently conform to a clearly erroneous majority opinion when personally faced with erroneous answers. In a modified version of the Asch's paradigm

(Crutchfield, 1955), participants received erroneous feedback from the other group members indirectly while sitting in individual cubicles. A meta-analysis of 133 studies (Bond & Smith, 1996) showed that conformity was often even higher in the Crutchfield paradigm than in the original Asch's paradigm. Thus, conformity occurs even when participants do not personally face social groups, which is critical for standard lab settings in neuroimaging studies. Nevertheless, when we initiated our studies of conformity, neuroimaging studies in this field were scarce.

A pioneering neuroimaging study investigated the fMRI signatures of conformity to group decisions during a mental rotation task (Berns et al., 2005). Conformity was associated with activity in the striatum and occipital-parietal cortices. Various neurophysiological studies have suggested that the striatum encodes reward values and participates in learning (Carelli, 2002; Knutson & Wimmer, 2007). Therefore, the first study implicated performance monitoring mechanisms in the effects of group norms. Nevertheless, the neurocognitive mechanism underlying conformity remains unclear.

Traditionally, psychological studies have focused on the rewarding value of social affiliation with others (Cialdini & Goldstein, 2004), while behavioral economics have emphasized the role of punishment in supporting social norms (Fehr & Fischbacher, 2004). Interestingly, both frameworks somehow imply that conformity is underlined by a fundamental learning mechanism reinforcing normative behaviors (Klucharev et al., 2009; Montague & Lohrenz, 2007). Therefore, in our studies, we addressed the following key research questions: Is the neurocognitive system that supports conformity functionally and structurally (at the level of neural networks) similar to the neurocognitive system underlying nonsocial reinforcement learning? Does group pressure result in a true modification of beliefs or merely in public compliance?

III. Neuroimaging signatures of conformity

A neural substrate of conformity

Articles selected for the defense: Klucharev et al., 2009; Klucharev et al., 2011.

fMRI study. Progress in neuroimaging studies of social conformity critically depends on effective behavioral paradigms. Therefore, we modified the classic Asch's conformity task, and designed a task in which the participant's initial judgments matched or mismatched group opinion (Klucharev et al., 2009). During the first (neuroimaging) session of our conformity task (Session 1), participants rated facial attractiveness (Figure 1). At the end of each trial, the "group rating" (descriptive norm) was presented. Importantly, according to a preprogrammed (semi-random) algorithm, the group rated the face differently than the participant or assigned the same rating. Such a procedure enabled us to systematically manipulate conflicts between the individual and the group opinion. In order to identify conformal changes of the attractiveness ratings, we instructed the participants to rate the same faces again during the behavioral Session 2. The participants indeed changed the ratings of attractiveness in line with the (normative) group's ratings. After controlling for the regression to the mean (Schnuerch, Schnuerch, et al., 2015), this task provides an effective tool to investigate conformity using neuroimaging methods. Importantly, the participants were not instructed to conform to the group norm but they did so automatically.

In the first study selected for the defense (Klucharev et al., 2009), we hypothesized that the neurocognitive mechanism of conformity may share some features with the fundamental performance-monitoring mechanisms. Therefore, a perceived deviation of the individual opinion from the group's opinion should evoke a neural activity that is fundamentally similar to RPE in reinforcement learning, signaling that participants adjust their judgments in line with group norms.

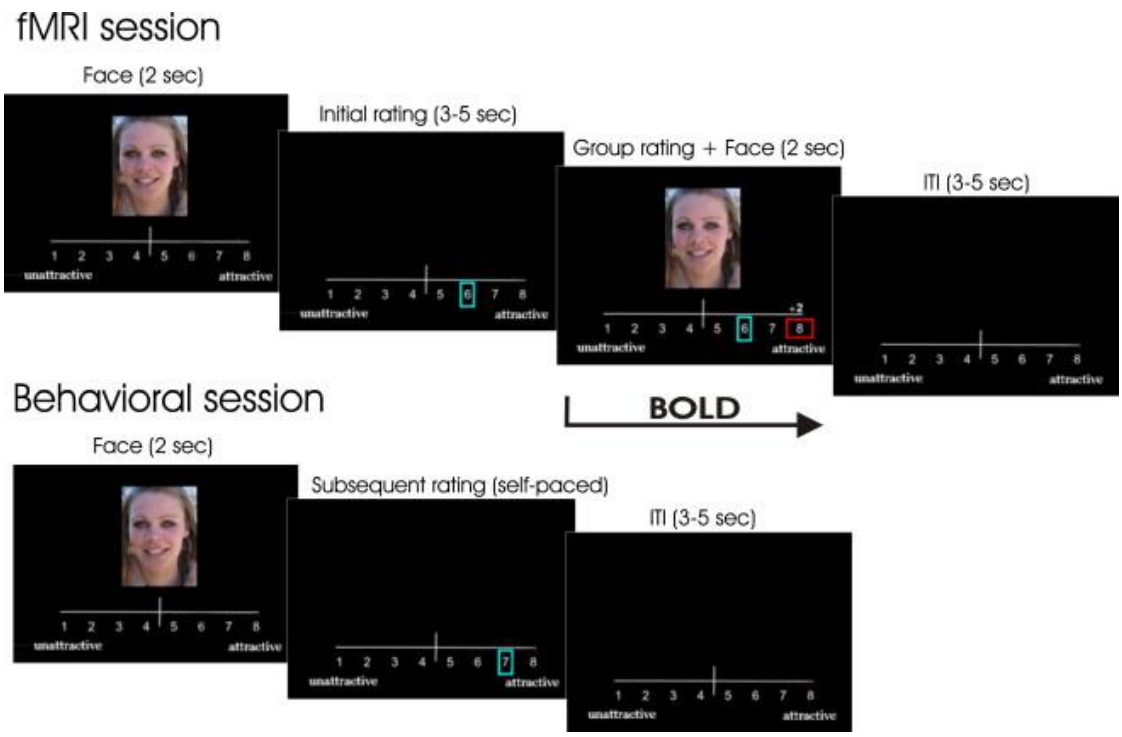


Figure 1. The conformity task that was used in the fMRi study. During the fMRI Session 1, the participants rated facial attractiveness. At the end of each trial, participants observed group ratings that matched or mismatched the participants' own ratings. During the behavioral session (Session 2), the participants again rated the same faces, and conformity to the group ratings was identified (Experiment N1, for details, see (Klucharev et al., 2009)).

Various learning models include an RPE that signals whether the outcome of an action is better or worse than expected (Rescorla & Wagner, 1972). The RPE encodes the learning signal that indicates the need to adjust one's behavior. The neuroimaging studies suggested that the pMFC encodes the RPE (Holroyd & Coles, 2002). Various studies have suggested that the pMFC is not alone in monitoring behavioral outcomes (Oldham et al., 2018). In particular, the ventral striatum plays a prominent role in reward prediction and performance monitoring. Therefore, a hypothetical reinforcement learning-like mechanism underlying conformity would require a learning signal at the pMFC and ventral striatum that should reflect deviations from peer opinions and trigger conformal adjustments.

To test our hypothesis, we used the aforementioned conformity task, in which the participants' initial judgments were influenced by group opinions (Figure 1). A total of 46 female students participated in the social (Experiment N1) and nonsocial control (Experiment N2) neuroimaging experiments (for details, see (Klucharev et al., 2009)).

Our fMRI study was performed using Sonata 1.5T (Siemens) scanner with ascending slice acquisition, using a $T2^*$-weighted echo-planar imaging sequence (33 axial slices; volume repetition time = 2.28 s; echo time = 35 ms; slice thickness, 3.5 mm with a gap of 0.5 mm; 90° flip angle; slice gap, 0.5 mm). For structural MRI, we used a T1-weighted MP-RAGE sequence (176 sagittal slices; field of view = 256 mm; volume repetition time = 2.25 s; echo time = 3.93 ms; 15° flip angle; slice matrix, 256 × 256; slice thickness, 1.0 mm with no gap).

First, to detect neural correlates of social influence, we compared brain responses in 'mismatch group' trials with brain responses in 'match group' trials (social *conflicts* contrast). Second, to detect neural correlates of conformity, we compared brain responses in 'mismatch group' trials followed by conformity and brain responses in 'mismatch group' trials not followed by conformity (*conformity* contrast). Lastly, to identify the similarity of neural correlates of social conflicts (social influence) and conformity, we used conjunctive analysis (Figure 2).

Conflicts with group opinion modulated the activity of the pMFC, insular cortex, middle frontal gyrus, and striatum. Our results suggest that the sources of neuronal correlates of conformity were quite similar to the RPE signal in reinforcement learning. Furthermore, striatal activity significantly correlated with individual differences in conformity and substantially differed between conformists and nonconformists (Figure 3).

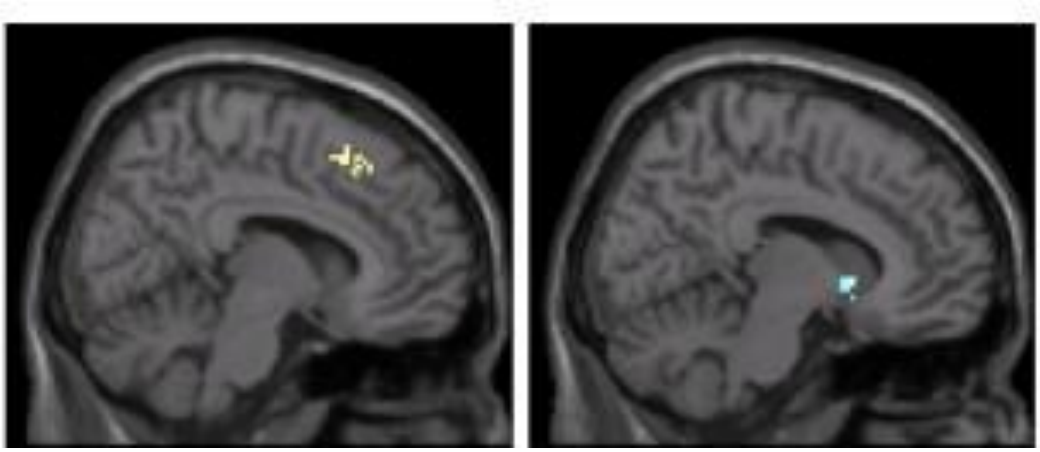


Figure 2. The results of the conjunction analysis show that activity in the posterior medial frontal cortex and ventral striatum reflects deviations from group opinion and subsequent conformity (Experiment N1).

Reinforcement learning has been robustly linked to the midbrain dopaminergic system (Wolfram Schultz, 2006). To account for this, we analyzed the activity of the ventral tegmental area and substantia nigra using a region of interest analysis. We showed that activity in these regions was modulated by perceived conflicts with the group's opinion (Klucharev et al., 2009). Overall, our results indicated that activity in the midbrain, pMFC, and ventral striatum monitored deviations from group opinion and reflected a degree of reward (or punishment) related to the level of affiliation (or disaffiliation) with the reference group. Interestingly, deviations from descriptive norms activated the pMFC but deactivated the ventral striatum, which may indicate various neurocognitive subroutines underlying conformity. Our first study suggested that conformity could indeed be an automatic neurocognitive process, in which dopaminergic mechanisms shift individual judgments to align with the group's judgment.

Importantly, to verify the social relevance of our findings, we designed and conducted a nonsocial version of the conformity task (Experiment N2, 22 females, aged 19–29 years; one participant was rejected from the future data analysis due to excessive head motion). In the new version of the task, we replaced group ratings with computer ratings (see (Spitzer et al., 2007; Zink et al., 2008), for a similar approach). Otherwise,

the design of the task and experimental setup were identical to the original conformity task. We compared the fMRI results of the social and nonsocial versions of the task. The statistical analysis showed an interaction of *conflicts* (within-group factor: mismatch versus match with the group opinion) and *social task* (between-group factor: social versus computer versions of the task) at the pMFC, ventral striatum, and midbrain regions (see Figure 4).

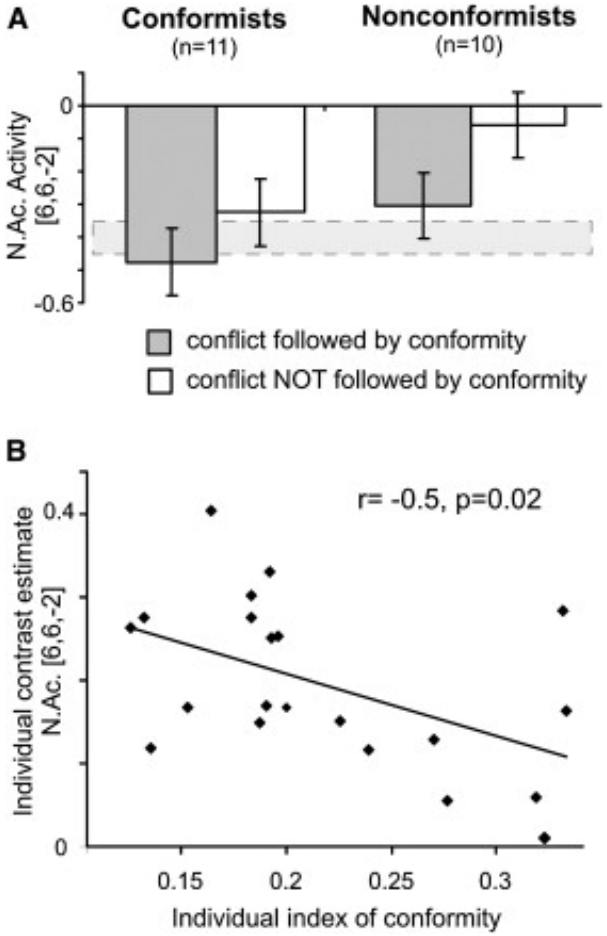


Figure 3. The activity of the ventral striatum reflects individual differences in conformal adjustments of opinion.

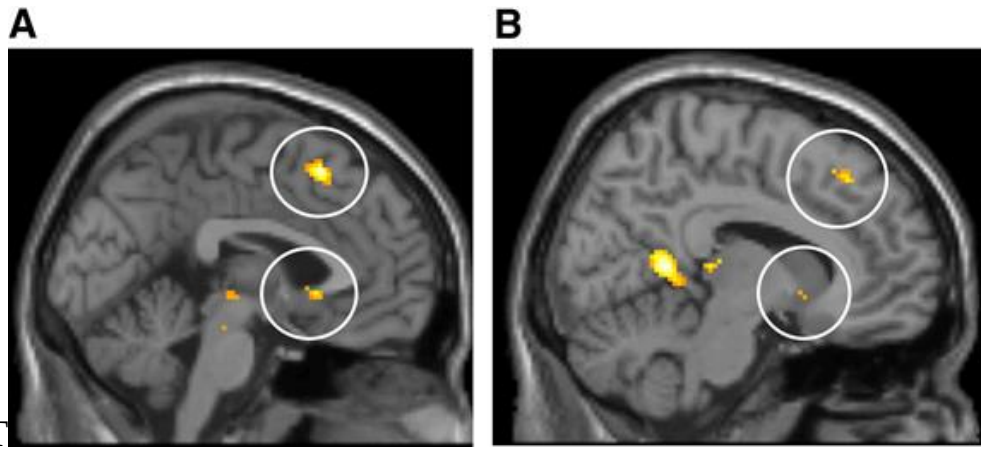
(A) Conformists (participants easily conforming to group (Klucharev et al., 2009) ratings in the conformity task) showed stronger deactivation of the ventral striatum to mismatches with group opinion. The gray rectangular area suggests a hypothetical threshold of conformity. (B) A significant correlation of fMRI correlates of *conformity* with the individual level of conformity. Conformists showed a smaller difference between brain responses to perceived mismatches, with group ratings that were followed by conformity and those that were not followed by conformity.

Importantly, the neural correlates of the conflicts (mismatches) with the normative group ratings at the medial frontal cortices and dopaminergic system were significantly stronger in the "social" version of the task than in the control study. Statistical analysis showed that the mismatch with the computer's ratings activated the precentral gyrus, right insula and precuneus in the control non-social study, similar to the social version of the study. Mismatches with the computer's ratings evoked quite weak activity of the pMFC and ventral striatum that could be observed only using a very low statistical threshold ($p < 0.006$). Overall, the deviations from group opinion evoked significantly weaker responses in the performance-monitoring brain regions in the control study than in the main study. The fMRI results of the two experiments indicated that the neural correlates of peer pressure were modulated by social factors.

We also analyzed the neural correlates of conformity (ratings changed in line with the computer rating vs. ratings not changed) in the control study. Data analysis showed an activation of the pMFC and ventral striatum predicting adjustments in line with computer ratings only with a relaxed statistical threshold ($p < 0.003$). Our results indicated that neural correlates of adjustments of opinion were similar in both experiments but they were strongly modulated by the social context.

The social and nonsocial versions of the conformity task also showed behavioral differences. Participants changed their opinions more after a mismatch with a social group than after a mismatch with a computer ($p = 0.004$). We also conducted a correlation analysis of the magnitude of the mismatch with a computer and conformity. Moreover, the correlation was significantly lower in the control computer condition than in the social version of the conformity task ($p = 0.001$). In the control computer condition, 12 of 21 participants did not show a significant correlation between the magnitude of the mismatch with a computer and the conformal adjustments of ratings. Overall, social descriptive norms modulated participants' opinions (and underlying neural activity) more strongly than did the control nonsocial stimuli.

The results of our fMRI study suggested that the neural mechanism underlying social conformity could be similar to a fundamental performance monitoring mechanism, for example, reinforcement learning. The fMRI data showed that the mismatches with descriptive norms modulated the activities of the pMFC, ventral striatum, insular cortex, precuneus, cerebellar tonsil, and other areas implicated in general error processing (Diedrichsen et al., 2005; Richard Ridderinkhof et al., 2003; Wolfram Schultz, 2006)). These results suggest that a perceived mismatch with descriptive norms evokes a neural response that is both functionally and neurophysiologically similar to the RPE in reinforcement learning. Hypothetically, a mismatch with group opinion may trigger an RPE-like response at the pMFC, ventral striatum, and midbrain: if such an "error"-related neural signal crosses a "learning" threshold, then conformal adjustment is initiated. A correlation of striatal activity with individual levels of conformity also supported a link of neural correlates of conformity with actual conformal adjustments of opinion.



T

Figure 4. Comparison of the fMRI correlates of conformity in the social (fMRI Experiment N1) and nonsocial (fMRI Experiment N2) versions of the conformity task (Klucharev et al., 2009).
(A) *Conflict* × *social task* interaction. (B) *Conformity* × *social task* interaction. White circles indicate the pMFC and ventral striatum.

TMS study. In the second study selected for the defense (Klucharev et al., 2011), we used repetitive transcranial magnetic stimulation (TMS) to confirm the causal role of

the pMFC in social influence. The stimulation site in the pMFC was selected based on our previous fMRI study (Klucharev et al., 2009). We used the precuneus as the control stimulation site. Participants attended one of three experimental groups: 1) TMS of the pMFC (*pMFC group*, 17 subjects); 2) TMS of the medial parietal cortex (*Control group*, 15 subjects); 3) a sham TMS of the pMFC (*Sham group*, 17 subjects).

For stimulation, we connected a 110 mm double cone coil (Magstim Company) to the Magstim Rapid magnetic stimulator. Importantly, the double cone coil consists of two angled windings that improves coupling to the head and allows stimulation of relatively deep cortical brain areas. For the first time within the subject of social neuroscience, we used theta-burst stimulation (600 pulses, main frequency = 50 Hz, inter-burst interval = 200 ms). Such a rapid TMS protocol decreases neural activity for approximately 60 minutes (Huang et al., 2005). Active motor thresholds were measured during TMS stimulation of the midline toe/leg area (primary motor cortex). We determined the minimum pulse intensity that produced a visible electromyography response in 50% of the trials during isometric contraction of the tibialis anterior muscle. In the current study, we used 80% intensity of this active motor threshold for the theta-burst TMS. Overall, participants received 40 s cTMS over either the pMFC (experimental condition) or parietal cortex (control stimulation), or sham stimulation (10% of the maximum output). Participants performed the conformity task for ∼3–5 min after the theta-burst TMS in the same laboratory.

The transient downregulation of the pMFC by the theta-burst TMS reduced the extent and probability of conformal behavioral adjustments relative to a sham and a control stimulation (Figure 5). Thus, we provided the first evidence of the causal role of the pMFC in social influence. Our TMS results showed that pMFC downregulation is capable of reducing conformity. Importantly, the pMFC is connected to the ventral striatum (Groenewegen et al., 1982; Hauber & Sommer, 2009; Parkinson et al., 2000). Thus, TMS of the pMFC can dysregulate the performance monitoring neural mechanism and inhibit behavioral adjustments to descriptive norms.

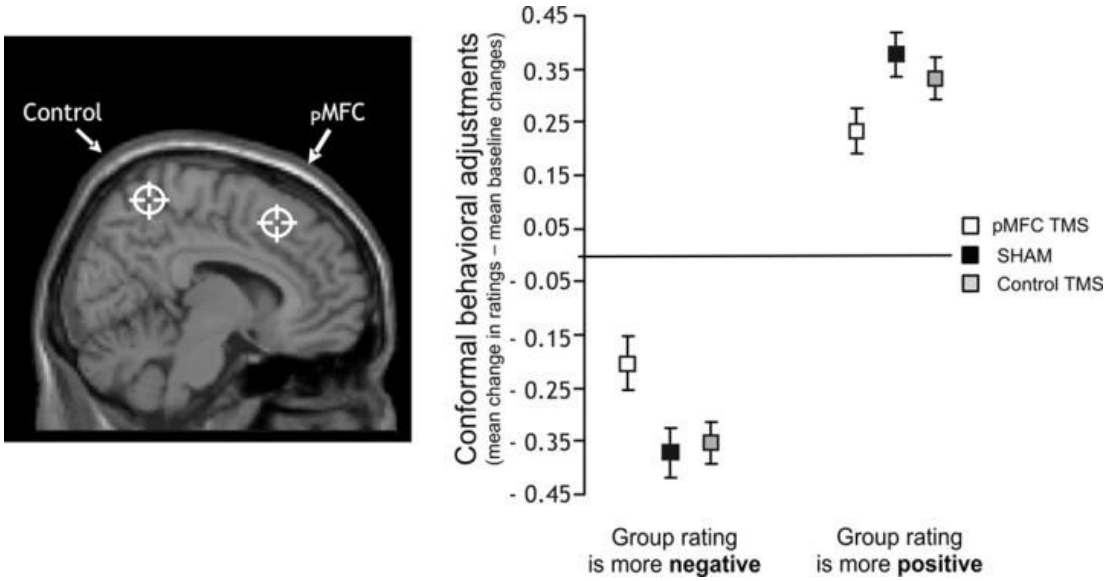


Figure 5. Effects of offline theta-burst transcranial magnetic stimulation on mean conforming adjustments of judgments. The left panel indicates the targets of theta-burst transcranial magnetic stimulation (Klucharev et al., 2011).

Figure 6 illustrates that TMS of the pMFC changed the probability of conformal adjustments relative to the Control and Sham TMS conditions. The probability of conformity decreased from 0.43 (Sham stimulation) and 0.42 (Control stimulation of the medial parietal cortex) to 0.38 after TMS of the pMFC: $F(2,46) = 3.55$, $p = 0.04$. Post hoc statistical analysis using the Tukey HSD test confirmed that the transient downregulation of the pMFC decreased the probability of conformity relative to the Control ($p = 0.033$) and Sham ($p = 0.025$) stimulation conditions.

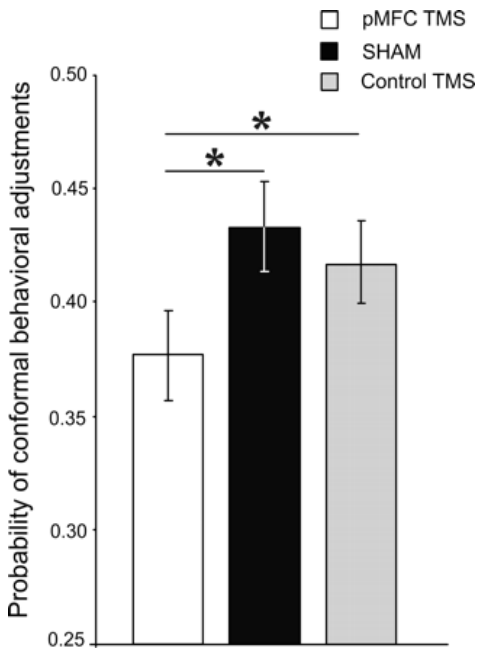


Figure 6. The effect of theta-burst transcranial magnetic stimulation on the probability of conformity in line with a group opinion (Klucharev et al., 2011).

Error bars – the standard error of the mean; * – Significant differences at the level of $p < 0.05$.

Notably, a number of later neuroimaging studies have confirmed that the pMFC is involved in different forms of conformity (e.g., (Berns et al., 2010b; Campbell-Meiklejohn et al., 2010; E. B. Falk et al., 2010; Ulf Toelch et al., 2018)).

IV. Temporal and spatial dynamics of neural signatures of conformity and group pressure

Articles selected for the defense: Shestakova et al., 2013, Zubarev et al., 2017.

EEG correlates of social conformity. In the third study selected for this defense, we measured the electrical activity of the brain to explore temporal and spatial dynamics of neural signatures of conformity (Shestakova et al., 2013). Electro- and magneto-encephalographic (EEG/MEG) are effective tools to study temporal changes in the neural correlates of conformity. A neural RPE signal can be measured using event-related potentials (ERPs). Importantly, an RPE signal is manifested as a component of evoked responses generated at the pMFC, which is often called *feedback-related negativity* (FRN) (see (Cohen & Ranganath, 2007; Miltner et al., 1997) or *reward positivity* (RewP) (Mulligan & Hajcak, 2018; Sambrook & Goslin, 2015). Our EEG study focused on FRN, which is a negative shift in the ERP that occurs within 200–400 ms after an individual receives negative performance feedback (Miltner et al., 1997).

Sixteen female students (aged 17–26 years) participated in the conformity task during two experimental sessions: an ERP session (Session 1) and a behavioral session (Session 2) that was separated by ∼15 min. According to the hypothetical reinforcement learning mechanism of conformity, mismatches with a group opinion should evoke FRN that has been associated with pMFC, performance monitoring, and subsequent adjustment of behavior.

EEG data were recorded (sampling frequency = 250 Hz, Mitsar Medical Diagnostic Equipment, EEG-201) using nineteen scalp electrodes (Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, T3, T4, P3, C3, Cz, C4, Pz, P4, O1, Oz, O2) and two ocular electrodes (one in the corner of the eye and another above the right eye). Electrode impedances were kept below 10 kΩ. We firstly band-pass filtered (0.1–70 Hz) EEG online and secondly filtered EEG offline at 0.5–20 Hz. During the recordings, EEG electrodes were on-line referenced to the average of all scalp electrodes. However, during prepossessing, EEG electrodes were offline referenced to the average of the two mastoids.

To identify the EEG signature of social influence, we compared the ERPs in all ‘mismatch group’ trials with ERPs in all ‘match group’ trials. To study subsequent conformity effects, we compared ‘mismatch group’ trials followed by conformity and ‘mismatch group’ trials not followed by conformity. We found that mismatches between individual and group opinions triggered a frontocentral negative deflection with a maximum of 200 ms, similar to FRN (Figure 7). Overall, the analysis showed that mismatches (conflicts) with group opinion evoked a sequence of neuronal responses: an earlier FRN-like response that indicated a mismatch with the group norm followed by a later ERP component that reflected a conformity. These results disentangled in time the neural signature of norm monitoring and the neural signature of conformal adjustments.

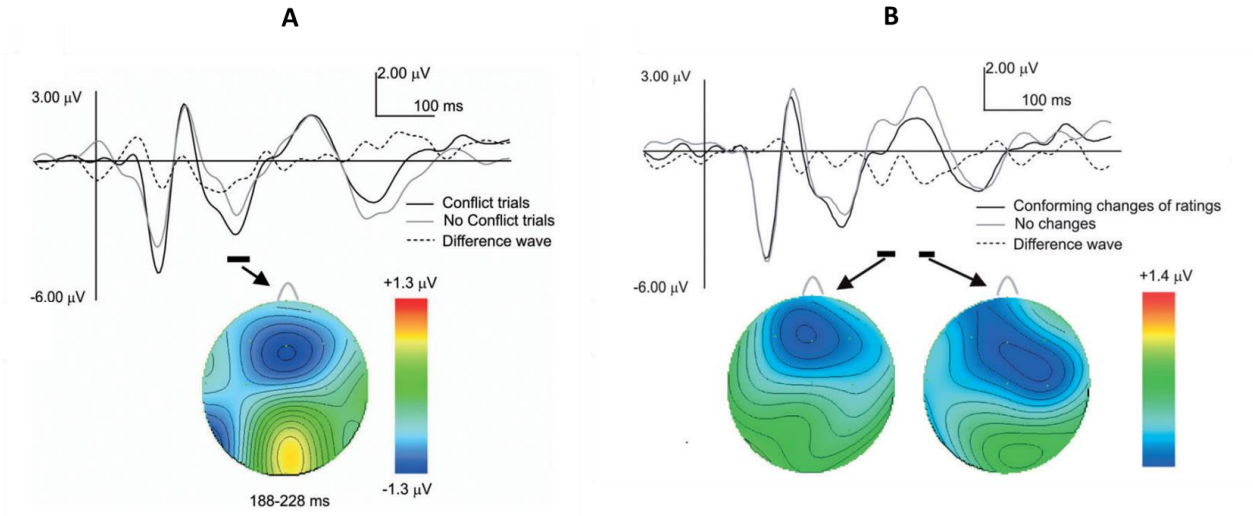


Figure 7. EEG “fingerprints” of group pressure and conformity. (A) Grand-averaged ERPs recorded in the trials where participants’ ratings mismatched the group opinion (black line) or matched the group opinion (gray line). The dotted line indicates the difference wave. (B) Grand-averaged ERPs recorded in the trials where participants changed their opinion in line with the group opinion (black line) or did not change it (gray line). Topographic maps show the voltage field distribution of the difference waves (Shestakova et al., 2013).

Several ERP studies have confirmed that a mismatch between individual opinion and group opinion triggers an FRN-like response (Chen et al., 2012; Kim et al., 2012). Importantly, FRN has been associated with pMFC activity, performance monitoring, and subsequent behavioral adjustment (e.g., (Cohen & Ranganath, 2007)). Neuroimaging studies have suggested that pMFC activity is modulated by a dopamine-related RPE signal encoding whether the outcome of an action is worse or better than expected (Holroyd & Coles, 2002; Matsumoto et al., 2007). The prefrontal cortex was shown to receive extensive dopaminergic projections (Fields et al. 2007), and Campbell-Meiklejohn's research group used a dopamine and noradrenaline agonist (methylphenidate) to modulate social conformity (Campbell-Meiklejohn et al., 2012). The results suggest that methylphenidate can enhance conformity and further support our hypothesis that the dopamine-related performance monitoring system is involved in social influence.

MEG correlates of social conformity. In the fourth study selected for the defense (Zubarev et al., 2017), we collected more detailed information on spatial and temporal dynamics of neural signatures of conformity using MEG. The vast majority of previous neuroimaging studies investigated conformity using fMRI, which could limit our understanding of the neurocognitive mechanisms of conformity. Twenty young females were invited to participate in our MEG study.

Our previous EEG study focused on the FRN component (Shestakova et al., 2013), which brain source remains debated. Various fMRI studies have reported the error-monitoring activity of the pMFC in response to negative outcomes. However, more recent MEG/EEG studies questioned the sole MFC origin of the error-monitoring activity, such as FRN (Doñamayor et al., 2011; Doñamayor, Schoenfeld, et al., 2012) and the closely related error-related negativity (Agam et al., 2011). Furthermore, these studies suggested the more posterior source of the activity in the posterior cingulate (PCC). For instance, in contrast to fMRI data a combined MEG–EEG analysis localized error-monitoring activity at the PCC of the same participants (Agam et al., 2011). Up to date, in studies of error-monitoring neural activity, the fMRI and EEG/MEG findings cannot be integrated straightforwardly. In the current study, we used a 306-channel Elekta Neuromag System

(102 magnetometers and 204 planar gradiometers, low-pass filter with a 333 Hz cut-off, sampling rate = 1000 Hz) to further investigate the temporal and spatial aspects of the detection of perceived mismatches or matches of individual and group opinions. .

We analyzed MEG brain activity that was collected during the conformity task. The statistical analysis showed that mismatches between individual and group opinions evoked activity of the posterior medial frontal regions (Figure 8): evoked responses at the precuneus and PCC (220–320 and 380–530 ms, respectively) increased frontal theta oscillations (4–8 Hz). Our results confirm an expectation bias toward conforming with group opinion, because the increase in theta oscillatory activity has been linked with unsigned prediction error signals (Cavanagh et al., 2012). We also conducted the source modeling of oscillatory activity: in the 'mismatch group' trials, frontal theta activity originated from the ventral MFC, PCC, and ACC, partly overlapping with the sources of the evoked responses. Thus, in our study, stronger theta activity in 'mismatch group' trials indicates an expectation bias toward a consistency with the group opinion.

We also found that mismatches between group opinions and individual opinions boosted beta oscillations (13–30 Hz) in the vMPFC. Thus, our MEG results suggest that an affiliation with descriptive norms is rewarding: matching the group opinion elevates beta band oscillatory activity in the ventral MFC, one of the key brain regions for processing reward information. Therefore, our MEG results also suggest that distinct neural circuits monitor deviations from norms and norm compliance.

fMRI studies of social influence consistently found elevated posterior MFC activity during mismatches of individual and group opinions (for a review, see (Izuma, 2013)). By contrast, our MEG results showed no such activity in the pMFC. Our MEG instead demonstrated more anterior and more posterior medial sources of such brain responses, similar to other EEG/MEG studies of the FRN (Doñamayor, Heilbronner, et al., 2012; Doñamayor, Schoenfeld, et al., 2012; Talmi et al., 2012). The multimodal EEG-MEG-fMRI neuroimaging study of error-related neural activity also reported more posterior MEG activity in contrast to the more anterior fMRI activation (Agam et al.,

2011). Overall, our MEG results indicate some discrepancy between fMRI and MEG responses to the perceived mismatches of individual and group opinions. The MEG results also highlighted the role of the posterior medial cortices in social influence. Further multimodal neuroimaging studies are needed to resolve the inconsistency between fMRI and MEG findings.

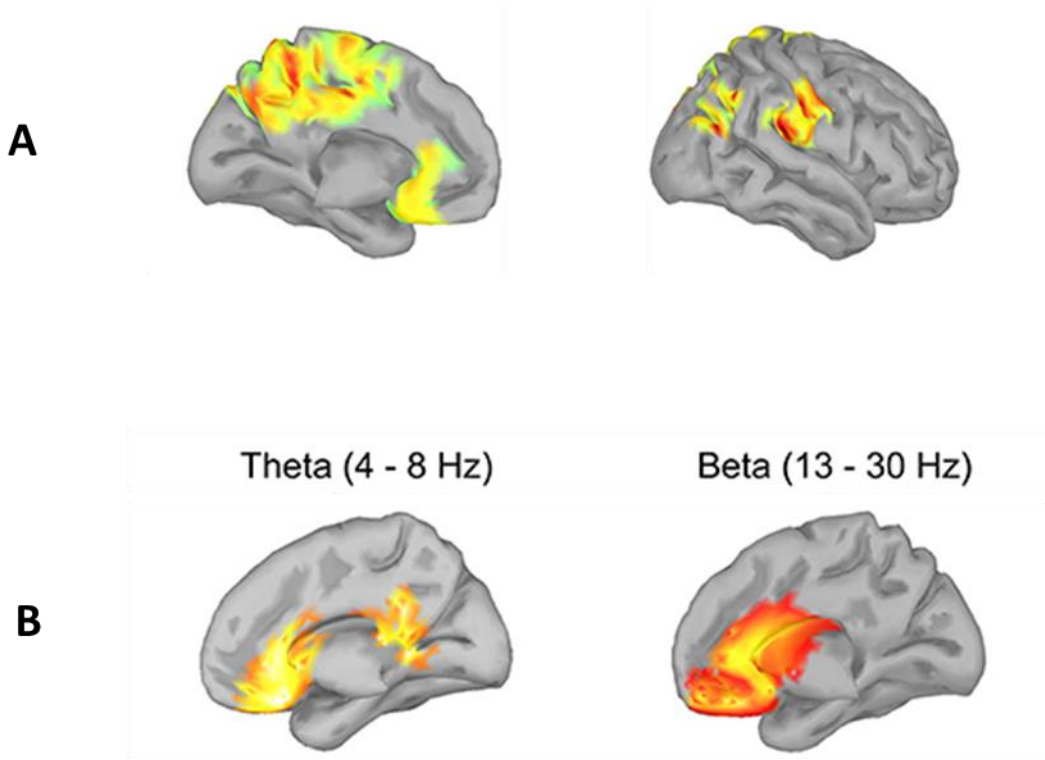


Figure 8. MEG correlates of the peer pressure. (A) Source reconstruction illustrates differences between evoked responses in 'mismatch group' trials and 'match group' trials. (B) Grand-averaged source localization for event-related (de)synchronization in theta (left) and beta (right) bands that differed significantly between 'mismatch group' and 'match group' trials (Zubarev et al., 2017).

Overall, our MEG results suggest that different learning subroutines may underlie social conformity. First, we observed an RPE-like signal in response to a perceived discrepancy with group opinion: both the evoked responses and frontal theta oscillations signaled deviations from group opinion. This mechanism engages error-processing circuitry in the anterior and posterior medial cortices. Second, the activity of the vMPFC

and ACC signals rewarding matches with opinion by bursts of beta oscillations. The role of the vMPFC in behavior integration is well documented (Knudsen & Wallis, 2022; Rolls et al., 2020). For example, Roy, Shohamy, and Wager suggested that the vMPFC plays a role as an integrative hub for sensory, memory, emotional, social, and self-related information processing (Roy et al., 2012). The vMPFC is highly interconnected with key functional regions of the brain that enable it to bind large-scale brain networks together during decision-making, emotional processing, memory formation, self-perception, and social cognition. Thus, the activity of the vMPFC may reinforce normative behavior and promote group coherence by making such behavior immediately rewarding.

V. Neural signatures of the internalization of group pressure

Articles selected for the defense: Gorin et al., 2021.

In the fifth study selected for the defense (Gorin et al., 2021), we used MEG to explore internalization of group pressure. Most neurocognitive studies have focused merely on neural responses to group norms. A limited number of studies have investigated the long-term effects of group pressure. Does group pressure result in a real modification of beliefs and opinions or just in public compliance (Berns et al., 2010a; Edelson et al., 2011; Schnuerch, Koppehele-Gossel, et al., 2015; Schnuerch, Schnuerch, et al., 2015; Zaki et al., 2011)? In a pioneering fMRI study, Jamil Zaki and colleagues used the conformity task to investigate the effect of social influence on intrinsic preferences (Zaki et al., 2011). The participants were scanned while they rated the stimuli a second time with no group ratings (Session 2). The data analysis searched for traces of social influence occurring 30 min prior to the session recording of the neural activity. This social influence echoed in the activity of the ventral striatum and vMPFC, indicating that norms modulate the neural representations of values assigned to stimuli. The results showed that social influence during the conformity task had relatively long-lasting effects on the valuation system of the brain. A follow-up study showed that after participants learned their peers' preferences, neural activity in the vMPFC started to depict the popularity of food items (Nook & Zaki, 2015).

In the current study, we used the magnetic source imaging to further study the after-effects of conformity. We used a modified conformity task in which, during Session 1, participants rated the facial trustworthiness and observed the group (trustworthiness) rating of each face. To detect the neural correlates of the internalization of group pressure, we analyzed MEG activity during Session 2 while participants rated the facial trustworthiness again, with no social normative information presented. Twenty female volunteers participated in the experiment (mean age 24.2 years; one participant was excluded from further analysis due to extensive artifacts).

The MEG data were recorded and processed according to good practice guidelines (Gross et al., 2013). We used a 306-channel Elekta Neuromag System comprising 102 magnetometers and 204 planar gradiometers, with a sampling rate of 1000 Hz and a low-pass filter with a 333 Hz cut-off frequency.

Our data analysis showed at the parietal cortices MEG-traces of past matches or mismatches with the group opinion. These early parietal correlates of mismatches with the majority (230 ms after face onset) were followed by vMPFC activity peaking around 320 ms after the face onset. The earliest traces of group pressure may indicate modulation of the social processing of faces or/and modulation of the memory for faces (Figure 9). The latest prefrontal markers of social influence could be related to modulation of the valuation process in the vMPFC. Thus, our MEG results have clarified spatiotemporal details of the long term effects of social influence. Our MEG results demonstrated the dynamics of the MEG "fingerprints" of the long-term effects of peer pressure: early parietal signatures of modified cognitive facial processing are followed by later prefrontal markers of long-term social influence on the neural valuation process.

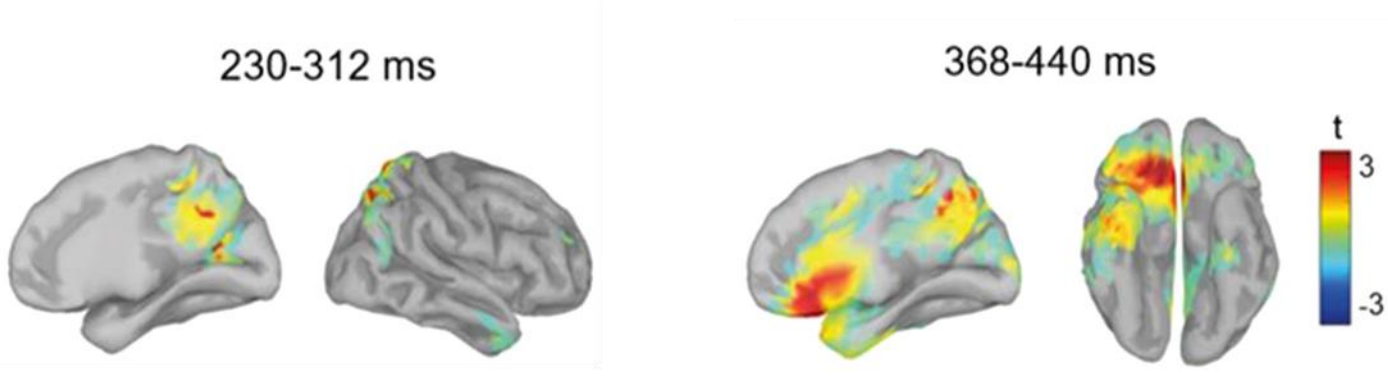


Figure 9. MEG signatures of the long-term effects of social influence (Session 2 of the conformity task). MEG markers of matches versus mismatches with the majority in the source space (t-statistics) (Gorin et al., 2021).

Interestingly, in the current study, the MEG correlates of the internalization of the peer pressure (Session 2 of the conformity task) shared some features with the neural

correlates of the mismatches with descriptive norms (Session 1 of the conformity task) reported in our previous MEG study (Zubarev et al., 2017). Mismatches with the (group) descriptive norms (regardless of whether they happened right now (Zubarev et al., 2017) or 30 min ago (Gorin et al., 2021)) evoked activity in the medial parietal and prefrontal cortices in both studies. Interestingly, our MEG data support other EEG studies of social influence (Chen et al., 2012; Kim et al., 2012; Shestakova et al., 2013) that demonstrated neural correlates of peer pressure also at the latency of around 200 ms. We can speculate that the long-term MEG signatures of mismatches with the descriptive norms fit the proposed (reinforcement learning-like) performance monitoring mechanism of social influence. Stimuli in Session 2 of the conformity task may trigger an error-related response associated with RPE circuitry. In line with this hypothesis, previous EEG studies have demonstrated that cues predicting future losses trigger stronger error-related negativity responses as compared to cues predicting future rewards (Baker & Holroyd, 2011; Dunning & Hajcak, 2008). Further, probabilistic information about future losses triggers stronger neural activity than probabilistic information about future rewards (Holroyd et al., 2009; Liao et al., 2011; Walsh & Anderson, 2012). Interestingly, evoked responses to cues predicting future losses often have a topography similar to the evoked responses to the losses themselves. Thus, in our MEG study, stimuli for which participants previously disagreed with the group opinion could become cues signaling mismatches with descriptive norms, leading to potential conflicts.

Our MEG results also highlighted the role of the precuneus in the internalization of group pressure. Previous fMRI studies showed stronger precuneus activity during mismatches with group opinion regarding healthy and unhealthy food items (Nook & Zaki, 2015), facial attractiveness (Klucharev et al., 2009), and facial trustworthiness (Zubarev et al., 2017). Since the precuneus has been implicated in general error processing (Nadig et al., 2010), it can encode long-term traces of social RPE-like signals. Furthermore, the precuneus has been implicated in trustworthiness information updates when current moral information mismatches with past moral information (Mende-Siedlecki et al., 2013). Alternatively, the parietal MEG signatures of social influence in

our study may indicate long-term changes in the processing of faces evoked by mismatches with descriptive norms.

Our MEG results also show signatures of the mismatches with group opinion in other brain regions: the superior parietal lobule and the intraparietal sulcus. Studies of top-down attention implicated the superior parietal lobule in memory retrieval (Cabeza et al., 2008; Corbetta & Shulman, 2002). Thus, in the present study, the activity of the superior parietal lobule may signal stronger memory strength for faces that have been associated with mismatches with group opinion (Hutchinson et al., 2014). Alternatively, mismatches with group's judgments could make such stimuli more salient, strengthening their memory engrams to facilitate future social interactions. Interestingly, we did not observe a neural echo of conformal changes of opinion that took place during Session 1, but we observed echoes of mismatches with group's judgments. Further studies are clearly needed to further explore the MEG signatures of the long-term effects of conformity. New paradigms that are more sensitive to minor changes in attitudes may discover fragile MEG signatures of the internalization of social influence.

Intraparietal sulcus activity observed in our MEG study confirmed the previous fMRI finding that showed the effect of social influence on this parietal region during a mental rotation task (Berns et al., 2005). We found that the activity of the right intraparietal sulcus was stronger if individual opinions were previously mismatched with group opinions. Interestingly, an influential line of research suggests that the activity of the intraparietal sulcus encodes a "mental line" (Nieder & Dehaene, 2009). When people think about numbers or evaluate facial attractiveness/trustworthiness using a Likert scale, they use a mental number line that maps numbers in a linear fashion. Neuroimaging and neuropsychological studies have indicated that numbers are represented along a continuous, left-to-right oriented mental line in the intraparietal sulcus. For example, damage to the intraparietal sulcus disrupts number processing (Ganor-Stern et al., 2020). Thus, in the present MEG study, the intraparietal sulcus activity may indicate a recalibration of the facial ratings induced by group's judgments.

In the present MEG study, we also observed that the activity of the vMPFC and ACC echoed previous mismatches with descriptive norms. Interestingly, ventromedial correlates of disagreement with the majority were observed within the 388–412 ms time interval at the vMPFC. It supports and updates previous fMRI findings that demonstrated signatures of social influence internalization in vMPFC activity (Zaki et al., 2011). Importantly, the role of the vMPFC in reward processing, value-based learning, decision making, and social cognition is well documented (e.g., (Elliott et al., 2010; O'Doherty, 2004; Padoa-Schioppa & Assad, 2006; Rushworth et al., 2007)). Thus, our MEG findings may further support our hypothesis that social influence manipulates an RPE-like learning mechanism and modulates our neural value representations of stimuli.

Similar to our EEG study (Shestakova et al., 2013), we found no significant effect of mismatches with the group judgments in Session 1 on the face-specific M170 component in Session 2. The M170 component has been associated with the early stages of facial perception (Halgren et al., 2000). Interestingly, an EEG study (Schnuerch, Koppehele-Gossel, et al., 2015), showed a larger posterior P2 component (155-175 ms) to faces for which participants previously matched with the group judgment, as compared to those on which they mismatched with the group. In our MEG study (Gorin et al., 2021), in Session 2, within a 158–312 ms time window, we observed a MEG echo of the earlier mismatches with the group judgments in Session 1. Possibly, this MEG signature of mismatches with the descriptive norm could be equivalent to the P2 component in EEG studies. This finding may suggest that mismatches with descriptive norms enhance attention to the relevant stimulus. However, we have to take into account that it is difficult to correctly compare EEG and MEG studies without a simultaneous EEG/MEG recording, since EEG and MEG are sensitive to different cortical sources of brain activity (Hämäläinen & Ilmoniemi, 1994). Thus, future studies combining EEG-MEG recordings are necessary to clarify whether the neural fingerprints of the long-term effects of social influence indicate enhanced attention or private acceptance.

In contrast to previous fMRI studies, the long-term signatures of group pressure in our MEG study were unspecific and insensitive to the sign and the level of mismatches with group opinions. Importantly, some discrepancies between the results of fMRI and MEG studies of social influence can be explained by methodological differences. First, MEG studies often require a larger number of trials than fMRI studies. A limited number of trials in our MEG study studies could decrease the statistical power of our data analysis. Second, behavioral paradigms often differ between studies of social influence. For example, in the present MEG study, participants rated facial trustworthiness, but in a similar fMRI study, participants rated facial attractiveness (Zaki et al., 2011). Future multimodal MEG-EEG-fMRI experiments are clearly needed to fully reconcile the results of different studies.

VI. A general neurobiological mechanism of social conformity

Articles selected for the defense: Klucharev et al., 2014, Zinchenko, Klucharev 2017.

A neurobiological mechanism of conformity

In the sixth and seventh studies selected for the defense (Klucharev et al., 2014; Zinchenko & Klucharev, 2017), we discussed and suggested a general mechanism of conformity. Three main neural mechanisms have been proposed for the social influence of descriptive group norms. According to the first proposed mechanism, perceived deviations from descriptive norms trigger a negative affect that calls for conformal adjustments (Berns et al., 2005). The second mechanism highlights the key role of cognitive inconsistency that is encoded in the pMFC (Izuma & Adolphs, 2013). The third perceived mechanism is similar to reinforcement learning (Klucharev et al., 2009; Montague & Lohrenz, 2007) that is also implemented in the pMFC and ventral striatum, which occurs when individuals adjust their intrinsic decision values in line with the group opinion based on an RPE-like neural signal. Below, we propose a hypothetical general mechanism of social conformity that attempts to integrate our current knowledge of the neural signature of group pressure.

A hypothetical neurocognitive mechanism of conformity should account for a number of key factors. The behavior of a social group normally signals relevant information to the individuals. Since our behaviors are tested by natural selection, the majority would adopt only the same behavioral patterns if they were advantageous in the current environment. Thus, from the evolutionarily perspective, the group's normative behavior is an "evolutionarily stable strategy" that cannot be bettered by an alternative strategy (J. M. Smith & Price, 1973). It follows that the best individual strategy is one that matches the strategy of the majority (Dawkins, 1989), since natural selection penalizes deviations from descriptive norms. Therefore, the monitoring of deviations from the descriptive norms is vital for survival, and could be integrated into the general neuronal performance-monitoring mechanisms.

On the whole, conformity to descriptive norms requires a learning signal that monitors deviations from norms and calls for adjustments in line with norms. In a constantly changing environment, efficient behavioral adjustments are vital for survival. Through millions of years of evolution, organisms have developed a learning mechanism that assigns expectations to all available options that are continuously updated via a reinforcement learning mechanism (Schultz et al., 1997). Many experts in the field have concluded that social influence may share the same mechanisms as nonsocial learning (e.g., Behrens et al., 2009; Heyes, 2012). For instance, Behrens and colleagues demonstrated that two neighboring subregions of the pMFC were involved in learning about social- and rewards-based information, further suggesting that social influence is underlined by the basic learning mechanism, that is, the dopamine-related activity of the pMFC (Behrens et al., 2009).

Thus, a hypothetical error signal, indicating deviations from descriptive norms, could initiate a learning mechanism similar to the standard RPE in reinforcement learning. A single exposure to a social influence in many conformity tasks makes it virtually impossible to apply conventional (model-free) reinforcement learning models to explain conformity. Nevertheless, a perceived conflict with a group opinion might generate a "social" RPE signal that reflects a difference between an individual and normative behavior. Such an error signal is then used to adjust individual beliefs, depending on how much weight is assigned to it. Previous neuroimaging studies largely support our hypothesis and indicate that the ventral striatum and pMFC continuously monitor and update subjective values based on a comparison of our own judgments/actions and normative ones.

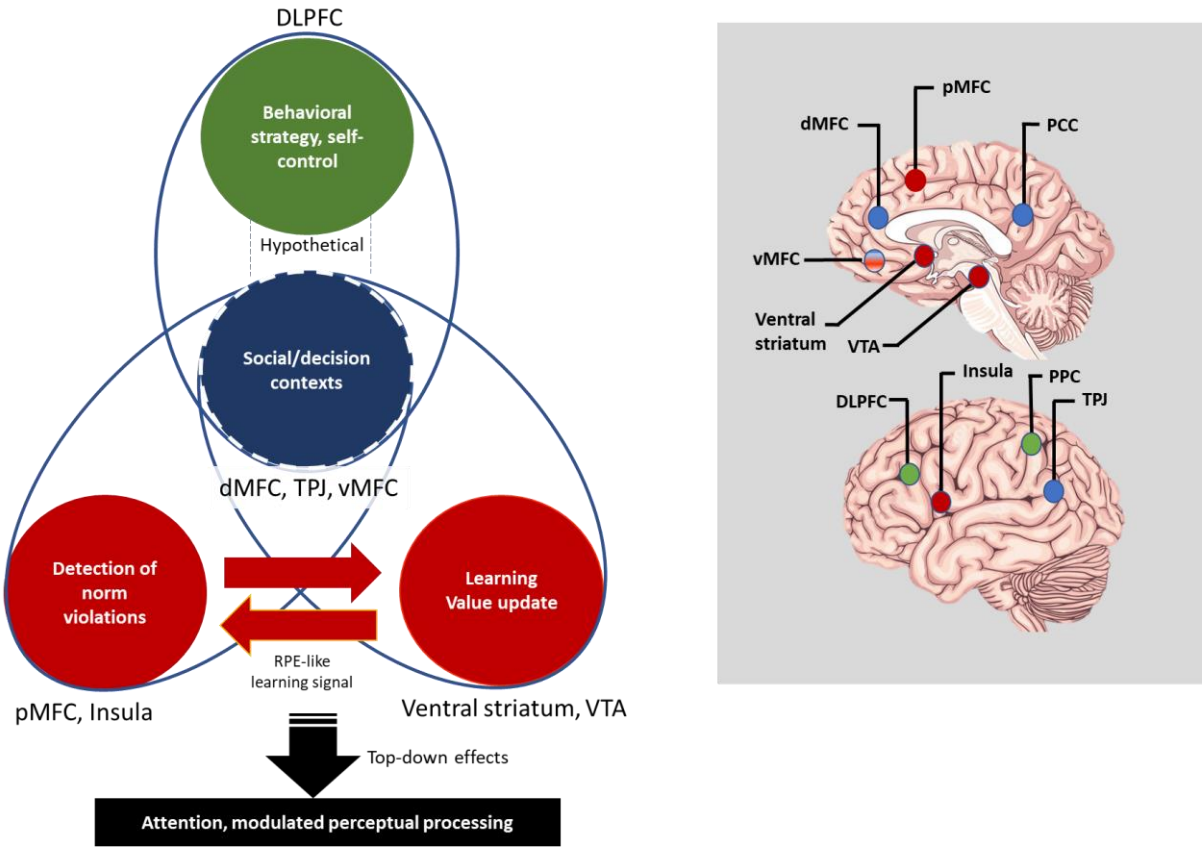


Figure 10. A hypothetical neuropsychological framework (left) and large-scale networks (right) of social conformity. Here, social conformity is initiated by the detection of descriptive norm violations. The *salience network* (red) detects (dMFC) violations of norms, and together with the dopaminergic system (ventral striatum), generates an RPE-like learning signal. Based on other lines of research (Krueger & Hoffman, 2016), we hypothesize that the aforementioned learning signal should be controlled for the social contexts (e.g., violations of in- versus out-group's norms should have different behavioral effects) via the *default mode network* (blue), including mentalizing regions (PCC, TPJ). Finally, the *central executive network*, which includes the dorsolateral prefrontal cortex (DLPFC), translates the context-dependent learning signal into decisions and attitude changes.

Abbreviations and notations. *Salience network*, in red: pMFC, posterior medial frontal cortex, VTA, ventral tegmental area. *Default-mode network*, in blue: mPFC, medial prefrontal cortex; dMFC, dorsomedial frontal cortex; PCC, posterior cingulate cortex; TPJ, temporoparietal junction. *Central executive network*, in green: DLPFC, dorsolateral prefrontal cortex; PPC, posterior parietal cortex.

A number of neuroimaging studies have implicated the pMFC, insula, and ventral striatum in neural mechanisms of social influence (e.g., Behrens et al., 2009; Berns et al., 2010a; Falk et al., 2010; Klucharev et al., 2009). An extensive meta-analysis demonstrated that the dorsal pMFC, ventral striatum, and anterior insula are consistently involved in normative decision-making (Wu et al., 2016): mismatches of individual opinion with group opinion reliably deactivate the ventral striatum and activate the dorsal pMFC and insula. Importantly, activity in the pMFC consistently predicts conforming adjustments in neuroimaging studies. The extent to which social conformity and reinforcement learning share a common neural mechanism remains unclear (Levorsen et al., 2021).

To suggest a neuropsychological framework of social influence, we adopt the approach of Krueger and Hoffman, who applied it to neural mechanisms of social punishment of norm violations (Krueger & Hoffman, 2016). Importantly, modern cognitive neuroscience focuses on large-scale networks at the level of multiple areas. Neuroimaging suggests that large-scale networks, rather than individual brain regions, build the biological basis of human cognitive architecture. A large corpus of studies identified several large-scale brain networks that operate during judgment and decision-making tasks, such as the *central executive network*, *default mode network*, and *salience network.* Overall, to suggest a neuropsychological framework of social influence, we must acknowledge that cognitive functions are distributed within the brain and implemented by large-scale brain networks.

Our studies suggest that the *salience network* plays a key role in detecting deviations from descriptive norms (Figure 10). The *salience network* includes the anterior insula and pMFC and includes subcortical structures such as the amygdala, the ventral striatum, and the substantia nigra/ventral tegmental area (Menon, 2015). We suggest that this network recognizes descriptive norm violations (pMFC) and generates an RPE-like learning signal (ventral striatum), which is a critical predictor of conformal adjustments. The role of the *default mode network* has been largely ignored in previous studies of conformity. Here, we hypothesize that during social influence, the *salience network*

interacts with the engagement of the *default mode network*, understood generally to be associated with mentalizing, memory, and self-monitoring (Bressler & Menon, 2010). Social and personal contexts modulate the effects of social influence. For example, we have a tendency to conform to the in-group majority but can largely ignore the out-group majority. Thus, the *default mode network*, including the medial prefrontal cortex, posterior cingulate cortex, and TPJ, should be involved in integrating emotional processing of the personal and social contexts with the learning signal, indicating violations of descriptive norms. Further, conformal adjustments of beliefs and behavior should involve the *central executive network*, including the dorsolateral prefrontal cortex, which is involved in higher-order context-dependent valuation, self-control, and decision making (Bressler & Menon, 2010). Such a framework assumes that the *central executive network* should translate the context-dependent learning signal (encoding the norm violation) into an actual behavioral adjustment. Overall, this process is likely to consist of calculating the personal relevance of norm violation (*default mode network*) and then selecting a way to adjust or ignore the group pressure (*central executive network*). Lastly, such a cascade of neurocognitive computations results in the interiorization of social norms and top-down effects on attention and sensory processing. Previous studies have largely focused on the role of the *salience network* in conformity. Further studies are needed to test our hypothesis about the role of the *default mode network* and the *central executive network* in the effects of social influence on beliefs and behavior. Most probably, the proposed neurocognitive mechanism of conformity can be extended to other forms of social influence.

The broader context of social conformity

It is essential to support the proposed mechanism of conformity by using other behavioral paradigms and social contexts. For example, fMRI data showed the involvement of the pMFC in the processing of group behavior during an ultimatum game, in which participants were exposed to decisions of other players (Wei et al., 2013). A mismatch with group opinion activated the *salience network*, including the pMFC and the insular cortex. Notably, the stronger pMFC and insula activity led to stronger adjustments

of the participants' initial choices, in line with the group descriptive norms. Another incisive fMRI study that focused on the learning of descriptive norms (Apps & Ramnani, 2017) demonstrated that participants learned the group's normative preferences during a delay-discounting task and then performed a similar task. Specifically, pMFC activities encode the value of rewards during normative choices.

It is also vital to verify whether the proposed neural mechanism of conformity is affected by key factors that modulate social influence. Toelch and colleagues created a novel paradigm that manipulates perceptual information, descriptive group norms, and financial bonuses for following group norms (Toelch et al., 2018). Similar to our study, mismatches with others, compared to matches with others, activated the pMFC. The special pattern of this activation differed in trials in which participants received a bonus for the agreements with the group as compared with trials in which they received a bonus for the disagreements with the group. Another fMRI study manipulated the source of group pressure (Izuma & Adolphs, 2013). The participants observed group's judgments of the liked in-group or the disliked out-group. Notably, the pMFC activity was shown to track the mismatches between individual and group judgments, also indicating whether an individual judgment differed from that of a liked or disliked group. A cross-cultural fMRI study of American and Chinese participants confirmed the differential effects of the in- and out-groups on conformity-related activity of the pMFC, insular cortex, and ventral striatum (Lin et al., 2018). In line with the robust findings of social psychology, participants demonstrated stronger conformity to the in-group than the out-group and a similar differential modulation of the activity of the pMFC, insular cortex, and ventral striatum. Thus, a broad spectrum of neuroimaging studies has confirmed that the *salience network* (pMFC, ventral striatum, and insular cortex) is involved in the detection of norm violations and conformal adjustments. The activity of this network is affected by relevant social and reward contexts that daily modulate the behavioral effects of social influence.

Our MEG results (Gorin et al., 2021) suggested that social influence may lead to long-term effects, internalization of the group pressure, or modulate perceptual processing of the socially relevant information. Thus, some cognitive and neuroimaging findings may

suggest that descriptive norms modulate attention to (or/and processing of) the relevant information. Nonsocial rewards can enhance attention toward relevant stimulus features that form involuntary biases (Hickey et al., 2010). For example, our EEG study demonstrated that reward-based learning enhances the attention-related neural response to reward-prediction cues (Krugliakova et al., 2019). Interestingly, descriptive norms modulate the activity of the occipital and parietal cortices, suggesting that peer pressure may modify sensory processing (Berns et al., 2005). Few EEG studies have also indicated the effect of social influence on visual processing and perceptual attention (Schnuerch et al., 2016; Schnuerch, Koppehele-Gossel, et al., 2015). Social influence has been shown to modulate early P1 components related to early visual processing (Trautmann-Lengsfeld & Herrmann, 2013) and this effect is even stronger in participants with low levels of autonomy (Trautmann-Lengsfeld & Herrmann, 2014). Using an incisive behavioral paradigm, Germar and colleagues demonstrated that social influence biases choices by altering the uptake of available sensory evidence (Germar et al., 2014, 2016). Thus, under social influence, people may analyze information more carefully, which is reflected in a diffusion model analysis as a smaller *drift rate*. We have also replicated these interesting findings in our lab using different visual stimuli (unpublished data). Overall, both neuroimaging and behavioral data indicate the effect of descriptive norms on the processing of visual stimuli.

Notably, it is hard to disentangle the value- and attention-based effects of conformity. Emotionally salient stimuli are often processed pre-attentively or get priority access to selective attention (e.g., Compton et al., 2003). Therefore, future studies should further elucidate the mechanisms of social influence—particularly in the context of the long-lasting effects of conformity that can be either value- or attention-based.

It can be somewhat challenging to integrate all neuroimaging studies of social influence due to substantial differences in methodology. For example, the EEG method differs from fMRI because it has a much better temporal resolution but lower spatial resolution. Since social influence is very much driven by subcortical dopaminergic activity, EEG can be particularly insensitive to the value-related effects of conformity.

However, fMRI is not sufficiently sensitive to disentangle the bottom-up and top-down effects of social influence on sensory processing. There are also clear differences in behavioral paradigms used in investigating social influence: in some studies (Berns et al., 2005; Germar et al., 2016; Huber et al., 2013; Trautmann-Lengsfeld & Herrmann, 2013; Welborn et al., 2016), the group judgments were presented prior to the participant's individual judgments, while other studies presented group judgments after the individual judgments (e.g., Izuma & Adolphs, 2013; Klucharev et al., 2009). Importantly, the former approach may highlight the informational nature of conformity, while the latter approach may boost the effects of conformity.

Here, we focused on social conformity. Unfortunately, until now, only a limited number of neuroimaging studies have investigated the neurocognitive mechanisms of other forms of social influence (e.g., Edelson et al., 2011; Falk et al., 2012, 2013; Klucharev et al., 2008; Stallen et al., 2010). A pioneering fMRI study explored brain activity evoked by effective persuasive smoking-cessation messages (Chua et al., 2009). The findings showed that pMFC and vMPFC activities were enhanced by more effective antismoking arguments. In a unique line of research, Emily Falk and Christin Scholz focused on the neurocognitive mechanisms of social influence from the perspectives of both communicators and receivers (Falk & Scholz, 2018). The researchers demonstrated that the neural mechanisms of effective persuasive communication and receptive conformity may substantially differ. For example, social approval could be particularly motivating for communicators of social influence. Neuroimaging studies have shown that both decisions to share information and the high success of the communicator are associated with activity in the brain valuation system (Baek et al., 2017; Falk et al., 2013; Scholz et al., 2017; Tamir et al., 2015). This suggests a positive value for sharing information with others. Information promoted by communicators more enthusiastically increased activity of the pMFC, vMPFC, and ventral striatum of communicators (Falk et al., 2012). Such brain activity of effective persuaders, particularly in the ventral striatum and vMPFC, suggests that sharing persuasive information with others is rewarding. However, some cognitive skills can be particularly important for effective

communicators; for example, mentalizing ability and theory of mind are essential for persuasive communication. Indeed, more effective salespeople tend to be stronger self-reported mentalizers and show stronger activity in the brain regions associated with mentalizing, including the TPJ and MFC (Dietvorst et al., 2009).

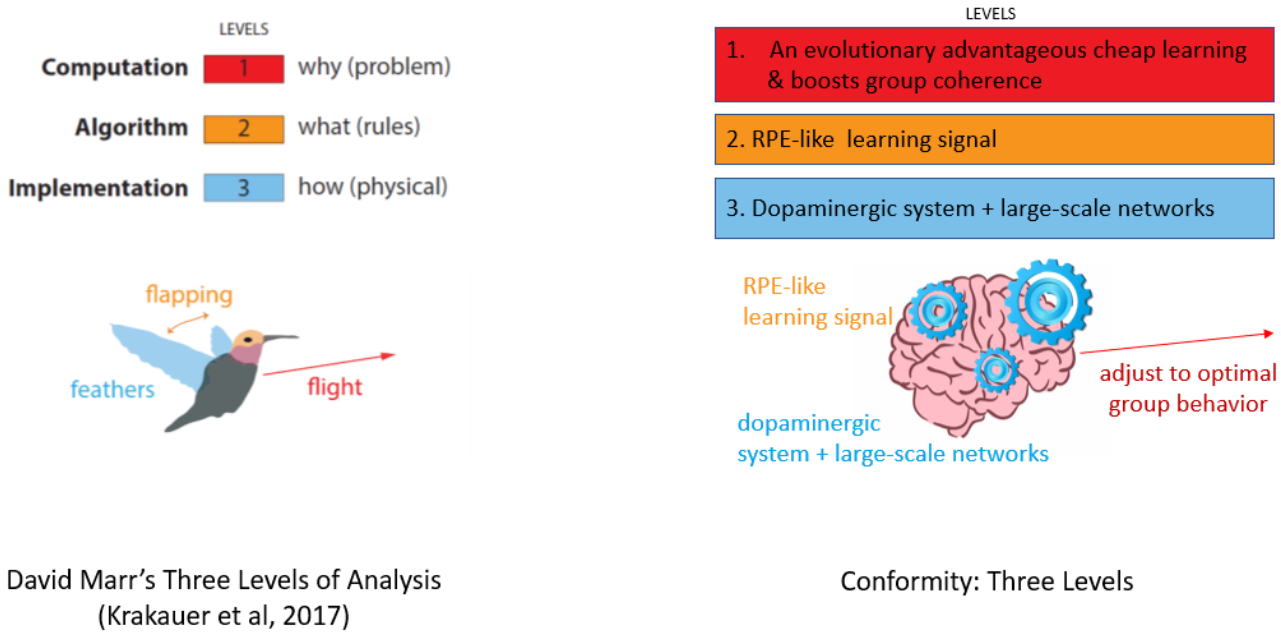


Figure 11. Conformity in a nutshell. Following David Marr's three levels of analysis of any information processing system (left panel), we suggest the following three explanatory levels of conformity to descriptive norms (right panel): (1) Conformity is an evolutionarily advantageous form of cheap social learning that boosts group coherence. (2) The brain detects deviations from descriptive norms and generates an RPE-like learning signal. (3) Large-scale brain networks, such as the *salience network*, *default mode network*, and *central executive network*, integrate the learning signal with personal and social contexts and convert it to the final behavioral conformal adjustments.

Figure 11 summarizes our current view of the neural mechanism of social conformity. We suggest that in order to implement an evolutionary advantageous social learning of group norms, the *salience network* (pMEC, ventral striatum, insular cortex) detects deviations from descriptive norms and generates an RPE-like learning signal that interacts with large-scale brain networks, such as the *default mode network* (PCC, TPJ:

mentalizing, memory, and self-monitoring) and the *central executive network* (DLPFC: higher-order context-dependent valuation, self-control, and decision making). The *default mode network* integrates emotional processing of personal and social contexts with the learning signal, while the *central executive network* converts this context-dependent learning signal into behavioral conformal adjustment that could play a role in resistance to social influence. Here, the vMPFC participates in linking large-scale networks. Future studies will verify this neurocognitive mechanism of conformity and may extend it to other forms of social influence.

Overall, much research remains to be done to clarify the detailed neural mechanism of social conformity using other computational models (e.g. Friedman et al., 2017), but recent neuroimaging studies have provided a tool to explore one of the most dramatic and elusive “mistakes” in our lives—that of being different from others.

VII. Conclusions

Day by day, our behavior is monitored and controlled by the social environment. Cognitive and social psychology demonstrated the profound impact of social descriptive norms on judgments, beliefs, and behavior. Nevertheless, the neurocognitive mechanisms of conformity to descriptive norms remain largely unknown. Our studies investigated the neural underpinnings of conformity. The studies selected for this defense were among the first to examine the neurocognitive mechanism of conformal adjustments of individual judgments to descriptive norms.

First, in the fMRI study, we demonstrated the involvement of the pMFC, ventral striatum, and insular cortex in social conformity to descriptive norms. The fMRI data showed that the activity of the pMFC and the ventral striatum not only reflects social influence but also predicts conformal adjustments of opinion in line with descriptive norms. Second, through the TMS study, we verified the causal role of the pMFC in conformal adjustments of opinion, in line with that of the group. The downregulation of the pMFC by theta-burst TMS reduced the extent and probability of conformity to descriptive norms relative to a sham and a control stimulation. Overall, we provided the first evidence of the causal role of the pMEFC in social influence.

Third, our EEG study demonstrated that mismatches between individual judgments and descriptive norms triggered a frontocentral FRN-like evoked response with the maximum at 200 ms, followed by a later ERP component (peaking at 380 ms), reflecting conformal adjustment to descriptive norms. Therefore, we disentangled the EEG signature of norm monitoring and the EEG signature of conformal adjustments.

Fourth, in the MEG study, we showed that mismatches between individual and descriptive norms modulated the activity of the PCC, TPJ, vMPFC, ACC, and occipital cortex. Our MEG results suggest that two generic learning sub-mechanisms may underlie social influence: (1) an RPE-like signal at dorsal medial frontal cortices that monitors the mismatches between individual judgments and descriptive norms, and (2) activity of the vMPFC and ACC, as indicated by an increase in power of beta oscillations that positively reinforces conformity.

Fifth, MEG source imaging (MEG study) showed signatures of the prolonged effect of descriptive norms. During the conformity task, when participants were exposed to the stimuli for the second time, previous mismatches between individual judgments and descriptive norms were still featured in (a) early signatures (230 ms) of modified stimuli processing at the parietal cortices and (b) later markers of social pressure on the valuation at the vMPFC.

Finally, based on two theoretical papers offering neuroimaging evidence, we concluded that individuals adjust their intrinsic decision values so that they are in line with group opinion based on a fundamental learning mechanism that is implemented in the *salience network*. Our results suggest that this network monitors descriptive norm violations and generates an RPE-like learning signal at the pMFC and ventral striatum. We also hypothesized that during social influence, the learning signal emitted by the *salience network* is modulated by the *default mode network* and *central executive network* to encode broader social and personal contexts and form an actual behavioral adjustment.

References

Agam, Y., Hämäläinen, M. S., Lee, A. K. C., Dyckman, K. A., Friedman, J. S., Isom, M., Makris, N., & Manoach, D. S. (2011). Multimodal neuroimaging dissociates hemodynamic and electrophysiological correlates of error processing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), 17556–17561. https://doi.org/10.1073/PNAS.1103475108/-/DCSUPPLEMENTAL

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall. https://searchworks.stanford.edu/view/808504

Apps, M. A. J., & Ramnani, N. (2017). Contributions of the Medial Prefrontal Cortex to Social Influence in Economic Decision-Making. *Cerebral Cortex*, *27*(9), 4635–4648. https://doi.org/10.1093/cercor/bhx183

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

Asch, S. (1955). Opinions and Social Pressure. *Scientific American*, *193*(5), 31–35. https://doi.org/10.1038/scientificamerican1155-31

Baek, E. C., Scholz, C., O'Donnell, M. B., & Falk, E. B. (2017). The Value of Sharing Information: A Neural Account of Information Transmission. *Psychological Science*, *28*(7), 851–861. https://doi.org/10.1177/0956797617695073

Baker, T. E., & Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological Psychology*, *87*(1), 25–34. https://doi.org/10.1016/J.BIOPSYCHO.2011.01.010

Bechtereva, N.P., Shemyakina, N.V., Starchenko, M.G., Danko, S.G., Medvedev, S.V. (2005). Error detection mechanisms of the brain: background and prospects. International Journal Psychophysiology, 58(2-3), 227-34. https://doi.org/10.1016/j.ijpsycho.2005.06.005.

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The Computation of Social Behavior. *Science*, *324*(5931), 1160–1164. https://doi.org/10.1126/science.1169694

Beretti, A., Figuières, C., & Grolleau, G. (2013). Using Money to Motivate Both 'Saints' and 'Sinners': a Field Experiment on Motivational Crowding-Out. *Kyklos*, *66*(1), 63–77. https://doi.org/10.1111/KYKL.12011

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010a). Neural Mechanisms of the Influence of Popularity on Adolescent Ratings of Music. *NeuroImage*, *49*(3), 2687. https://doi.org/10.1016/J.NEUROIMAGE.2009.10.070

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010b). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, *49*(3), 2687–2696. https://doi.org/10.1016/j.neuroimage.2009.10.070

Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., & Richards, J. (2005). Neurobiological Correlates of Social Conformity and Independence During Mental Rotation. *Biological Psychiatry*, *58*(3), 245–253. https://doi.org/10.1016/j.biopsych.2005.04.012

Bodin, Ö. (2017). Collaborative environmental governance: Achieving collective action in social-ecological systems. *Science*, *357*(6352). https://doi.org/10.1126/SCIENCE.AAN1114/ASSET/DBFCD748-BF94-4CA9-8DF5-DDEA323453AA/ASSETS/GRAPHIC/357_AAN1114_F3.JPEG

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), 111–137. https://doi.org/10.1037/0033-2909.119.1.111

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a Neural Network centered on human lateral frontopolar cortex. *PLoS Biology*, *9*(6). https://doi.org/10.1371/JOURNAL.PBIO.1001093

Boyd, R., Richerson, P. J., McElreath, R., Henrich, J. P., Soltis, J. M., Gintis, H., Bowles, S., Mulder, M. B., Durham, W. H., & Bettinger, R. L. (2005). *The origin and evolution of cultures*. 456. https://global.oup.com/academic/product/the-origin-and-evolution-of-cultures-9780195181456

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290.

https://doi.org/10.1016/J.TICS.2010.04.004

Bryan, J. H., & Test, M. A. (1967). Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology*, *6*(4, Pt.1), 400–407. https://doi.org/10.1037/h0024826

Burkett, S. R. (1977). School Ties, Peer Influence, and Adolescent Marijuana Use. *Sociological Perspectives*, *20*(2), 181–202. https://doi.org/10.2307/1388930/ASSET/1388930.FP.PNG_V03

Byerly, H., Balmford, A., Ferraro, P. J., Hammond Wagner, C., Palchak, E., Polasky, S., Ricketts, T. H., Schwartz, A. J., & Fisher, B. (2018). Nudging pro-environmental behavior: evidence and opportunities. *Frontiers in Ecology and the Environment*, *16*(3), 159–168. https://doi.org/10.1002/FEE.1777

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). Parietal Cortex and Episodic Memory: An Attentional Account. *Nature Reviews. Neuroscience*, *9*(8), 613. https://doi.org/10.1038/NRN2459

Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., & McClintock, M. K. (2000). Multilevel integrative analyses of human behavior: social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin*, *126*(6), 829–843. https://doi.org/10.1037/0033-2909.126.6.829

Cacioppo, J. T., & Decety, J. (2011). Social neuroscience: challenges and opportunities in the study of complex behavior. *Annals of the New York Academy of Sciences*, *1224*(1), 162–173. https://doi.org/10.1111/j.1749-6632.2010.05858.x

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. https://doi.org/10.1016/j.cub.2010.04.055

Campbell-Meiklejohn, D. K., Simonsen, A., Jensen, M., Wohlert, V., Gjerløff, T., Scheel-Kruger, J., Møller, A., Frith, C. D., & Roepstorff, A. (2012). Modulation of Social Influence by Methylphenidate. *Neuropsychopharmacology*, *37*(6), 1517–1525. https://doi.org/10.1038/npp.2011.337

Carelli, R. M. (2002). The Nucleus Accumbens and Reward: Neurophysiological

Investigations in Behaving Animals. *Behavioral and Cognitive Neuroscience Reviews*, *1*(4), 281–296. https://doi.org/10.1177/1534582302238338

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220–238. https://doi.org/10.1111/j.1469-8986.2011.01293.x

Centola, D., Becker, J., Brackbill, D., & Baronchelli, A. (2018). Experimental evidence for tipping points in social convention. *Science*, *360*(6393), 1116–1119. https://doi.org/10.1126/SCIENCE.AAS8827/SUPPL_FILE/AAS8827_CENTOLA_SM.PDF

Chen, B., Zhu, Z., Wang, Y., Ding, X., Guo, X., He, M., Fang, W., Zhou, Q., Zhou, S., Lei, H., Huang, A., Chen, T., Ni, D., Gu, Y., Liu, J., & Rao, Y. (2018). Nature vs. nurture in human sociality: multi-level genomic analyses of social conformity. *Journal of Human Genetics*, *63*(5), 605–619. https://doi.org/10.1038/s10038-018-0418-y

Chen, Wu, Y., Tong, G., Guan, X., & Zhou, X. (2012). ERP correlates of social conformity in a line judgment task. *BMC Neuroscience*, *13*(1), 43. https://doi.org/10.1186/1471-2202-13-43

Chua, H. F., Liberzon, I., Welsh, R. C., & Strecher, V. J. (2009). Neural Correlates of Message Tailoring and Self-Relatedness in Smoking Cessation Programming. *Biological Psychiatry*, *65*(2), 165–168. https://doi.org/10.1016/j.biopsych.2008.08.030

Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Cohen, M., & Ranganath, C. (2007). Reinforcement Learning Signals Predict Future

Decisions. *Journal of Neuroscience*, *27*(2), 371–378. https://doi.org/10.1523/JNEUROSCI.4421-06.2007

Compton, R. J., Banich, M. T., Mohanty, A., Milham, M. P., Herrington, J., Miller, G. A., Scalf, P. E., Webb, A., & Heller, W. (2003). Paying attention to emotion: an fMRI investigation of cognitive and emotional stroop tasks. *Cognitive, Affective & Behavioral Neuroscience*, *3*(2), 81–96. https://doi.org/10.3758/cabn.3.2.81

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/NRN755

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist*, *10*(5), 191–198. https://doi.org/10.1037/h0040237

Danchin, E., Nöbel, S., Pocheville, A., Dagaeff, A.-C., Demay, L., Alphand, M., Ranty-Roby, S., van Renssen, L., Monier, M., Gazagne, E., Allain, M., & Isabel, G. (2018). Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions. *Science (New York, N.Y.)*, *362*(6418), 1025–1030. https://doi.org/10.1126/science.aat1590

Dawkins, R. (1989). *The selfish gene*. Oxford University Press. http://www.gurteen.com/gurteen/gurteen.nsf/id/X0001E4CE/

Day, R. L., MacDonald, T., Brown, C., Laland, K. N., & Reader, S. M. (2001). Interactions between shoal size and conformity in guppy social foraging. *Animal Behaviour*, *62*(5), 917–925. https://doi.org/10.1006/ANBE.2001.1820

Diedrichsen, J., Hashambhoy, Y., Rane, T., & Shadmehr, R. (2005). Neural correlates of reach errors. *Journal of Neuroscience*, *25*(43), 9919–9931. https://doi.org/10.1523/JNEUROSCI.1874-05.2005

Dietvorst, R. C., Verbeke, W. J. M. I., Bagozzi, R. P., Yoon, C., Smits, M., & Lugt, A. van der. (2009). A Sales Force—Specific Theory-of-Mind Scale: Tests of Its Validity by Classical Methods and Functional Magnetic Resonance Imaging. In *Journal of Marketing Research* (Vol. 46, pp. 653–668). Sage Publications, Inc. https://doi.org/10.2307/20618926

Dindo, M., Whiten, A., & de Waal, F. B. M. (2009). In-group conformity sustains different foraging traditions in capuchin monkeys (Cebus apella). *PloS One*, *4*(11), e7858. https://doi.org/10.1371/journal.pone.0007858

Doñamayor, N., Heilbronner, U., & Münte, T. F. (2012). Coupling electrophysiological and hemodynamic responses to errors. *Human Brain Mapping*, *33*(7), 1621. https://doi.org/10.1002/HBM.21305

Doñamayor, N., Marco-Pallarés, J., Heldmann, M., Schoenfeld, M. A., & Münte, T. F. (2011). Temporal dynamics of reward processing revealed by magnetoencephalography. *Human Brain Mapping*, *32*(12), 2228. https://doi.org/10.1002/HBM.21184

Doñamayor, N., Schoenfeld, M. A., & Münte, T. F. (2012). Magneto- and electroencephalographic manifestations of reward anticipation and delivery. *NeuroImage*, *62*(1), 17–29. https://doi.org/10.1016/J.NEUROIMAGE.2012.04.038

Dunning, J. P., & Hajcak, G. (2008). *See no evil: Directing visual attention within unpleasant images modulates the electrocortical response*. https://doi.org/10.1111/j.1469-8986.2008.00723.x

Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the crowd: brain substrates of long-term memory conformity. *Science (New York, N.Y.)*, *333*(6038), 108–111. https://doi.org/10.1126/science.1203557

Elliott, R., Agnew, Z., & Deakin, J. F. W. (2010). Hedonic and informational functions of the human orbitofrontal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(1), 198–204. https://doi.org/10.1093/CERCOR/BHP092

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010). Predicting Persuasion-Induced Behavior Change from the Brain. *Journal of Neuroscience*, *30*(25), 8421–8424. https://doi.org/10.1523/JNEUROSCI.0063-10.2010

Falk, E., & Scholz, C. (2018). Persuasion, Influence, and Value: Perspectives from Communication and Social Neuroscience. *Annual Review of Psychology*, *69*(1), 329–356. https://doi.org/10.1146/annurev-psych-122216-011821

Falk, Emily B., Morelli, S. A., Welborn, B. L., Dambacher, K., & Lieberman, M. D. (2013). Creating Buzz. *Psychological Science*, *24*(7), 1234–1242. https://doi.org/10.1177/0956797612474670

Falk, Emily B., O'Donnell, M. B., & Lieberman, M. D. (2012). Getting the word out: neural correlates of enthusiastic message propagation. *Frontiers in Human Neuroscience*, *6*, 313. https://doi.org/10.3389/fnhum.2012.00313

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Foster, K., & Ratnieks, F. (2005). A new eusocial vertebrate? *Trends in Ecology & Evolution*, *20*(7), 363–364. https://doi.org/10.1016/j.tree.2005.05.005

Friedman, D., Magnani, J., Paranjpe, D., Sinervo, B. (2017). Evolutionary games, climate and the generation of diversity. PLoS ONE 12(8): e0184052. https://doi.org/10.1371/journal.pone.0184052

Galef, B. G., & Whiskin, E. E. (2008). 'Conformity' in Norway rats? *Animal Behaviour*, *75*(6), 2035–2039. https://doi.org/10.1016/J.ANBEHAV.2007.11.012

Ganor-Stern, D., Gliksman, Y., Naparstek, S., Ifergane, G., & Henik, A. (2020). Damage to the Intraparietal Sulcus Impairs Magnitude Representations of Results of Complex Arithmetic Problems. *Neuroscience*, *438*, 137–144. https://doi.org/10.1016/J.NEUROSCIENCE.2020.05.006

Germar, M., Albrecht, T., Voss, A., & Mojzisch, A. (2016). Social conformity is due to biased stimulus processing: electrophysiological and diffusion analyses. *Social Cognitive and Affective Neuroscience*, *11*(9), 1449–1459. https://doi.org/10.1093/scan/nsw050

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social Influence and Perceptual Decision Making. *Personality and Social Psychology Bulletin*, *40*(2), 217–231. https://doi.org/10.1177/0146167213508985

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*(3), 472–482. https://doi.org/10.1086/586910

Gorin, A., Klucharev, V., Ossadtchi, A., Zubarev, I., Moiseeva, V., & Shestakova, A. (2021). MEG signatures of long-term effects of agreement and disagreement with the majority. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-82670-x

Groenewegen, H. J., Room, P., Witter, M. P., & Lohman, A. H. M. (1982). Cortical afferents of the nucleus accumbens in the cat, studied with anterograde and retrograde transport techniques. *Neuroscience*, *7*(4), 977–996. https://doi.org/10.1016/0306-4522(82)90055-0

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J. R., van Wassenhove, V., Wibral, M., & Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363. https://doi.org/10.1016/J.NEUROIMAGE.2012.10.001

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. (2000). Cognitive Response Profile of the Human Fusiform Face Area as Determined by MEG. *Cerebral Cortex*, *10*(1), 69–81. https://doi.org/10.1093/CERCOR/10.1.69

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*(1), 35–42. https://doi.org/10.1007/BF02512476

Hauber, W., & Sommer, S. (2009). Prefrontostriatal circuitry regulates effort-related decision making. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(10), 2240–2247. https://doi.org/10.1093/cercor/bhn241

Haun, D. B. M., Rekers, Y., & Tomasello, M. (2012). Majority-Biased Transmission in Chimpanzees and Human Children, but Not Orangutans. *Current Biology*, *22*(8), 727–731. https://doi.org/10.1016/j.cub.2012.03.006

Haun, D. B. M., Rekers, Y., & Tomasello, M. (2014). Children Conform to the Behavior of Peers; Other Great Apes Stick With What They Know. *Psychological Science*, *25*(12), 2160–2167. https://doi.org/10.1177/0956797614553235

Heyes, C. (2012). What's social about social learning? *Journal of Comparative Psychology*, *126*(2), 193–202. https://doi.org/10.1037/a0025180

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward Changes Salience in Human Vision via the Anterior Cingulate. *Journal of Neuroscience*, *30*(33), 11096–11103. https://doi.org/10.1523/JNEUROSCI.1026-10.2010

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., & Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 59–70. https://doi.org/10.3758/CABN.9.1.59

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/J.NEURON.2004.12.033

Huber, R. E., Klucharev, V., & Rieskamp, J. (2013). Neural correlates of informational cascades: Brain mechanisms of social influence on belief updating. *Social Cognitive and Affective Neuroscience*, *10*(4). https://doi.org/10.1093/scan/nsu090

Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., & Wagner, A. D. (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cerebral Cortex (New York, N.Y. : 1991)*, *24*(1), 49–66. https://doi.org/10.1093/CERCOR/BHS278

Izuma, K. (2013). The neural basis of social influence and attitude change. *Current Opinion in Neurobiology*, *23*(3), 456–462. https://doi.org/10.1016/J.CONB.2013.03.009

Izuma, K., & Adolphs, R. (2013). Social Manipulation of Preference in the Human Brain. *Neuron*, *78*(3), 563–573. https://doi.org/10.1016/j.neuron.2013.03.023

Johnston, K. L., & White, K. M. (2003). Binge-drinking: A test of the role of group norms in the theory of planned behaviour. *Psychology and Health*, *18*(1), 63–77. https://doi.org/10.1080/0887044021000037835

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative

conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012. https://doi.org/10.1177/01461672002610009

Kesebir, S. (2012). The Superorganism Account of Human Sociality. *Personality and Social Psychology Review*, *16*(3), 233–261. https://doi.org/10.1177/1088868311430834

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. https://doi.org/10.3758/s13415-011-0067-5

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, *61*(1). https://doi.org/10.1016/j.neuron.2008.11.027

Klucharev, V., Munneke, M. A. M., Smidts, A., & Fernández, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *Journal of Neuroscience*, *31*(33). https://doi.org/10.1523/JNEUROSCI.1869-11.2011

Klucharev, V., Smidts, A., & Fernández, G. (2008). Brain mechanisms of persuasion: How "expert power" modulates memory and attitudes. *Social Cognitive and Affective Neuroscience*, *3*(4). https://doi.org/10.1093/scan/nsn022

Klucharev, V., Zubarev, I., & Shestakova, A. (2014). Neurobiological mechanisms of social influence (Нейробиологические механизмы социального влияния). *Experimental Psychology*, *7*(4), 20–36.

Knudsen, E. B., & Wallis, J. D. (2022). Taking stock of value in the orbitofrontal cortex. *Nature Reviews Neuroscience 2022 23:7*, *23*(7), 428–438. https://doi.org/10.1038/s41583-022-00589-2

Knutson, B., & Wimmer, G. E. (2007). Splitting the difference: how does the brain code reward episodes? *Annals of the New York Academy of Sciences*, *1104*(1), 54–69. https://doi.org/10.1196/annals.1390.020

Konopasky, R. J., & Telegdy, G. A. (1977). Conformity in the Rat: A Leader's Selection of Door Color versus a Learned Door-Color Discrimination. *Perceptual and Motor*

Skills, *44*(1), 31–37. https://doi.org/10.2466/pms.1977.44.1.31

Krueger, F., & Hoffman, M. (2016). The Emerging Neuroscience of Third-Party Punishment. *Trends in Neurosciences*, *39*(8), 499–501. https://doi.org/10.1016/J.TINS.2016.06.004

Krugliakova, E., Gorin, A., Fedele, T., Shtyrov, Y., Moiseeva, V., Klucharev, V., & Shestakova, A. (2019). The Monetary Incentive Delay (MID) Task Induces Changes in Sensory Processing: ERP Evidence. *Frontiers in Human Neuroscience*, *13*, 382. https://doi.org/10.3389/fnhum.2019.00382

Levorsen, M., Ito, A., Suzuki, S., & Izuma, K. (2021). Testing the reinforcement learning hypothesis of social conformity. *Human Brain Mapping*, *42*(5), 1328–1342. https://doi.org/10.1002/HBM.25296

Liao, Y., Gramann, K., Feng, W., Deák, G. O., & Li, H. (2011). This ought to be good: brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology*, *48*(10), 1412–1419. https://doi.org/10.1111/J.1469-8986.2011.01205.X

Lin, L. C., Qu, Y., & Telzer, E. H. (2018). Intergroup social influence on emotion processing in the brain. *Proceedings of the National Academy of Sciences*, *115*(42), 10630–10635. https://doi.org/10.1073/pnas.1802111115

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), 9493–9498. https://doi.org/10.1073/PNAS.0608842104/SUPPL_FILE/IMAGE916.GIF

Louis, W., Davies, S., Smith, J., & Terry, D. (2007). Pizza and Pop and the Student Identity: The Role of Referent Group Norms in Healthy and Unhealthy Eating. *The Journal of Social Psychology*, *147*(1), 57–74. https://doi.org/10.3200/SOCP.147.1.57-74

Manning, M. (2009). The effects of subjective norms on behaviour in the theory of planned behaviour: a meta-analysis. *The British Journal of Social Psychology*, *48*(Pt 4), 649–705. https://doi.org/10.1348/014466608X393136

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience*, *10*(5), 647–656. https://doi.org/10.1038/nn1890

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic Value Underlies Asymmetric Updating of Impressions in the Morality and Ability Domains. *Journal of Neuroscience*, *33*(50), 19406–19415. https://doi.org/10.1523/JNEUROSCI.2334-13.2013

Menon, V. (2015). Salience Network. *Brain Mapping: An Encyclopedic Reference*, *2*, 597–611. https://doi.org/10.1016/B978-0-12-397025-1.00052-X

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a "Generic" Neural System for Error Detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798. https://doi.org/10.1162/jocn.1997.9.6.788

Montague, P. R., & Lohrenz, T. (2007). To Detect and Correct: Norm Violations and Their Enforcement. *Neuron*, *56*(1), 14–18. https://doi.org/10.1016/j.neuron.2007.09.020

Mulligan, E. M., & Hajcak, G. (2018). The electrocortical response to rewarding and aversive feedback: The reward positivity does not reflect salience in simple gambling tasks. *International Journal of Psychophysiology*, *132*, 262–267. https://doi.org/10.1016/J.IJPSYCHO.2017.11.015

Nadig, K. G., Jäncke, L., Lüchinger, R., & Lutz, K. (2010). Motor and non-motor error and the influence of error magnitude on brain activity. *Experimental Brain Research*, *202*(1), 45–54. https://doi.org/10.1007/S00221-009-2108-7

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/ANNUREV.NEURO.051508.135550

Nook, E. C., & Zaki, J. (2015). Social Norms Shift Behavioral and Neural Responses to Foods. *Journal of Cognitive Neuroscience*, *27*(7), 1412–1426. https://doi.org/10.1162/jocn_a_00795

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin, F. S., Crépin, A. S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., … De Zeeuw, A. (2016). Social norms as solutions. *Science*, *354*(6308), 42–43. https://doi.org/10.1126/SCIENCE.AAF8317/SUPPL_FILE/NYBORGSM.PDF

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. https://doi.org/10.1016/J.CONB.2004.10.016

O'Keefe, D. J. (2002). *Persuasion : theory & research*. Sage Publications. https://books.google.ru/books/about/Persuasion.html?id=e3V6Zen0UGwC&redir_esc=y

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, *39*(8), 3398. https://doi.org/10.1002/HBM.24184

Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., Rockström, J., Allerberger, F., McCaffrey, M., Doe, S. S. P., Lenferna, A., Morán, N., van Vuuren, D. P., & Schellnhuber, H. J. (2020). Social tipping dynamics for stabilizing Earth's climate by 2050. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(5), 2354–2365. https://doi.org/10.1073/PNAS.1900577117

Padoa-Schioppa, C., & Assad, J. A. (2006). *Neurons in the orbitofrontal cortex encode economic value*. https://doi.org/10.1038/nature04676

Parkinson, J. A., Willoughby, P. J., Robbins, T. W., & Everitt, B. J. (2000). Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: further evidence for limbic cortical-ventral striatopallidal systems. *Behavioral Neuroscience*, *114*(1), 42–63. http://www.ncbi.nlm.nih.gov/pubmed/10718261

Pike, T. W., & Laland, K. N. (2010). Conformist learning in nine-spined sticklebacks'

foraging decisions. *Biology Letters*, *6*(4), 466–468. https://doi.org/10.1098/rsbl.2009.1014

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The Transsituational Influence of Social Norms. *Journal of Personality and Social Psychology*, *64*(1), 104–112. https://doi.org/10.1037/0022-3514.64.1.104

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory (Eds Black AH, Prokasy WF) New York: Appleton Century Crofts* (pp. 64–99).

Richard Ridderinkhof, K., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2003). Empirical and Theoretical Perspectives on Animal Cognition. *14. R. Gelman, B. Butterworth, Trends Cognit. Sci*, *306*, 59. www.sciencemag.org

Rivis, A., & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology*, *22*(3), 218–233. https://doi.org/10.1007/S12144-003-1018-2

Rolls, E. T., Cheng, W., & Feng, J. (2020). The orbitofrontal cortex: reward, emotion and depression. *Brain Communications*, *2*(2). https://doi.org/10.1093/BRAINCOMMS/FCAA196

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156. https://doi.org/10.1016/J.TICS.2012.01.005

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, *11*(4), 168–176. https://doi.org/10.1016/J.TICS.2007.01.004

Sambrook, T. D., & Goslin, J. (2015). A Neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235. https://doi.org/10.1037/BUL0000006

Schnuerch, R., Koppehele-Gossel, J., & Gibbons, H. (2015). Weak encoding of faces

predicts socially influenced judgments of facial attractiveness. *Social Neuroscience*, *10*(6), 624–634. https://doi.org/10.1080/17470919.2015.1017113

Schnuerch, R., Richter, J., Koppehele-Gossel, J., & Gibbons, H. (2016). Multiple neural signatures of social proof and deviance during the observation of other people's preferences. *Psychophysiology*, *53*(6), 823–836. https://doi.org/10.1111/psyp.12636

Schnuerch, R., Schnuerch, M., & Gibbons, H. (2015). Assessing and correcting for regression toward the mean in deviance-induced social conformity. *Frontiers in Psychology*, *06*, 669. https://doi.org/10.3389/fpsyg.2015.00669

Schofffild, P. E., Pattison, P. E., Hill, D. J., Borland, R., John, D., & Ron, H. &. (2001). The influence of group identification on the adoption of peer group smoking norms. *Psychology and Health*, *16*(1), 1–16. https://doi.org/10.1080/08870440108405486

Scholz, C., Baek, E. C., O'Donnell, M. B., Kim, H. S., Cappella, J. N., & Falk, E. B. (2017). A neural model of valuation and information virality. *Proceedings of the National Academy of Sciences*, *114*(11), 2881–2886. https://doi.org/10.1073/pnas.1615259114

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). *The Constructive, Destructive, and Reconstructive Power of Social Norms*.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schultz, Wolfram. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*(1), 87–115. https://doi.org/10.1146/annurev.psych.56.091103.070229

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). https://doi.org/10.1093/scan/nss064

Smith, J. M., & Price, G. R. (1973). *The logic of Animal COnflict*.

Smith, J. R., & Louis, W. R. (2009). Teaching and Learning Guide for: Group Norms and the Attitude–Behaviour Relationship. *Social and Personality Psychology Compass*,

3(5), 850–854. https://doi.org/10.1111/J.1751-9004.2009.00200.X

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*(1), 185–196. https://doi.org/10.1016/J.NEURON.2007.09.011

Stallen, M., Smidts, A., Rijpkema, M., Smit, G., Klucharev, V., & Fernández, G. (2010). Celebrities and shoes on the female brain: The neural correlates of product evaluation in the context of fame. *Journal of Economic Psychology*, *31*(5). https://doi.org/10.1016/j.joep.2010.03.006

Sun, S., & Yu, R. (2016). Social conformity persists at least one day in 6-year-old children. *Scientific Reports*, *6*, 39588. https://doi.org/10.1038/srep39588

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning) (Adaptive Computation and Machine Learning series): Sutton, Richard S., Barto, Andrew G.: 9780262193986: Amazon.com: Books*. MIT press.

Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., & Dolan, R. J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. *Neuroimage*, *59*(1), 635. https://doi.org/10.1016/J.NEUROIMAGE.2011.06.051

Tamir, D. I., Zaki, J., & Mitchell, J. P. (2015). Informing others is associated with behavioral and neural signatures of value. *Journal of Experimental Psychology: General*, *144*(6), 1114–1123. https://doi.org/10.1037/xge0000122

Toelch, U., Pooresmaeili, A., & Dolan, R. J. (2018). Neural substrates of norm compliance in perceptual decisions. *Scientific Reports 2018 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-21583-8

Toelch, Ulf, Panizza, F., & Heekeren, H. R. (2018). Norm compliance affects perceptual decisions through modulation of a starting point bias. *Royal Society Open Science*, *5*(3), 171268. https://doi.org/10.1098/rsos.171268

Trautmann-Lengsfeld, S. A., & Herrmann, C. S. (2013). EEG reveals an early influence of social conformity on visual processing in group pressure situations. *Social Neuroscience*, *8*(1), 75–89. https://doi.org/10.1080/17470919.2012.742927

Trautmann-Lengsfeld, S. A., & Herrmann, C. S. (2014). Virtually simulated social pressure influences early visual processing more in low compared to high autonomous participants. *Psychophysiology*, *51*(2), 124–135. https://doi.org/10.1111/psyp.12161

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience and Biobehavioral Reviews*, *36*(8), 1870–1884. https://doi.org/10.1016/J.NEUBIOREV.2012.05.008

Wei, Z., Zhao, Z., & Zheng, Y. (2013). Neural mechanisms underlying social conformity in an ultimatum game. *Frontiers in Human Neuroscience*, *7*, 896. https://doi.org/10.3389/fnhum.2013.00896

Welborn, B. L., Lieberman, M. D., Goldenberg, D., Fuligni, A. J., Galván, A., & Telzer, E. H. (2016). Neural mechanisms of social influence in adolescence. *Social Cognitive and Affective Neuroscience*, *11*(1), 100–109. https://doi.org/10.1093/scan/nsv095

White, K. M., & Hyde, M. K. (2012). The Role of Self-Perceptions in the Prediction of Household Recycling Behavior in Australia. *Environment and Behavior*, *44*(6), 785–799. https://doi.org/10.1177/0013916511408069/ASSET/IMAGES/LARGE/10.1177_0013916511408069-FIG1.JPEG

Whiten, A., Horner, V., & de Waal, F. B. M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, *437*(7059), 737–740. https://doi.org/10.1038/nature04047

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the Theoretical Foundation of Sociobiology. *The Quarterly Review of Biology*, *82*(4), 327–348. https://doi.org/10.1086/522809

Wilson, E., & Wilson, D. (2008). Evolution "for the Good of the Group." *American Scientist*, *96*(5), 380. https://doi.org/10.1511/2008.74.380

Wu, H., Luo, Y., & Feng, C. (2016). Neural signatures of social conformity: A coordinate-based activation likelihood estimation meta-analysis of functional brain imaging

studies. *Neuroscience and Biobehavioral Reviews*, *71*, 101–111. https://doi.org/10.1016/j.neubiorev.2016.08.038

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, *22*(7), 894–900. https://doi.org/10.1177/0956797611411057

Zinchenko, O., & Klucharev, V. (2017). Commentary: The Emerging Neuroscience of Third-Party Punishment. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/FNHUM.2017.00512

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know Your Place: Neural Processing of Social Hierarchy in Humans. *Neuron*, *58*(2), 273. https://doi.org/10.1016/J.NEURON.2008.01.025

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*(JAN). https://doi.org/10.3389/fnins.2017.00010

Glossary

ACC — anterior cingulate cortex

DLPFC — dorsolateral prefrontal cortex

dMFC — dorsal medial frontal cortex

EEG — electroencephalography

ERPs — event-related potentials

fMRI — functional magnetic resonance imaging

FRN — feedback related negativity

MEG — magnetoencephalography

PCC — posterior cingulate cortex

pMFC — posterior medial frontal cortex, an area that encompasses a posterior portion of the anterior cingulate cortex and the pre-supplementary motor area, plays a critical role in adaptive, goal-directed behavior

RPE — reward prediction error, the difference between expected rewards and the reward that has just been experienced

TMS — transcranial magnetic stimulation

TPJ — temporoparietal junction

VMPFC — ventromedial prefrontal cortex

VTA — ventral tegmental area

Keywords

Social conformity, fMRI, TMS, MEG, EEG, reinforcement learning, neuroimaging, cognitive mechanisms, social influence, medial prefrontal cortex, ventral striatum, dopamine system