

Neural Mechanisms of Social Influence

Malia F. Mason and Rebecca Dyer

Columbia Business School

Michael I. Norton

Harvard Business School

Address Correspondence To:

Dr. Malia F. Mason
Columbia Business School
720 Uris Hall
3200 Broadway
New York, NY 10027

Email: maliamason@columbia.edu

Phone: (212) 854-1070

Abstract

The present investigation explores the neural mechanisms underlying the impact of social influence on preferences. We socially tagged symbols as valued or not – by exposing participants to the preferences of their peers – and assessed subsequent brain activity during an incidental processing task in which participants viewed popular, unpopular, and novel symbols. The medial prefrontal cortex (mPFC) differentiated between symbols that were and were not socially tagged – a possible index of normative influence – while aspects of the striatum (the caudate) differentiated between popular and unpopular symbols – a possible index of informational influence. These results suggest that integrating activity in these two brain regions may differentiate objects that have become valued as a result of social influence from those valued for non-social reasons.

Keywords: Social Neuroscience; Social Influence; Social Norms; Preferences; Brain; Consumer Behavior; Normative Influence; Informational Influence

The iconic Lacoste green crocodile logo – initially created in honor of the 1920's French tennis star Rene Lacoste – has waxed and waned in popularity over the years, enjoying enormous popularity in the United States in the late 1970's and early 1980's while nearly disappearing from sight in the 1990's, only to reemerge in this decade as a desired status symbol. One reason to buy clothing featuring this crocodile might be to emulate Lacoste himself, of course, but we would be surprised if American teenagers have any awareness of the origins of the logo. Instead, such trends are often driven by the adoption – and rejection – of products by others: The value of little green crocodiles depends critically on the value that others attach to that symbol.

In this paper, we model the process by which social influence impacts preferences in a one-hour experimental session, using a paradigm in which we train participants to see symbols as socially valued or not by providing them with feedback about the preferences of others. We then examine the impact of this social feedback on the brain activity that participants exhibit while viewing objects that have been endorsed or rejected by their peers, exploring the neural processes underlying changes in valuation due to social influence.

Mechanisms of Social Influence

The notion that humans are influenced in their beliefs, preferences, and behaviors by the beliefs, preferences, and behaviors of others has become nearly axiomatic across the social sciences; the sheer number of terms used to describe this process is indicative of its ubiquity, from social influence to social proof to peer pressure to bandwagon effects to conformity to herding (Abrahamson, 1991; Asch, 1951; Banerjee, 1992;

Bikhchandani, Hirshleifer, & Welch, 1992; Sherif, 1936; see Cialdini & Goldstein, 2004, for a review). Indeed, even non-human primates are quick to adhere to social norms (Whiten, Horner, & de Waal, 2005). The impact of social influence has been demonstrated in countless domains, including pain perception (Craig & Prkachin, 1978), littering (Cialdini, Reno, & Kallgren, 1990), voting (Gerber, Green, & Larimer, 2008), donating to charities (Reingen, 1982), expressing prejudice (Apfelbaum, Sommers, & Norton, 2008), choosing jobs (Higgins, 2001; Kilduff, 1990), investing in the stock market (Hong, Kubik, & Stein, 2004), and, most relevant to the current investigation, both adoption and rejection of consumer products (Berger & Heath, 2007). Judgment and decision-making researchers have also increasingly recognized the fundamental impact of social factors on human behavior: Investigations of such “social decision-making” (Sanfey, 2007) have the potential to speed the integration of psychologists with both neuroscientists and game theorists, offering a more complete account of human decision-making (Camerer, 2003).

From an early stage in research exploring the effects of social factors on behavior, researchers focused their attention on a fundamental question: the extent to which behavior influenced by peers indicated a true change in attitude, such that social influence changed people’s minds, or merely a desire to be publically consistent with the viewpoints of others. Following on Asch’s (1951) classic conformity studies, in which participants gave obviously wrong answers to a simple line judgment task when confederates had given those wrong answers before them, Deutsch and Gerard (1955) varied whether responses to such tasks were visible to others or not; they showed that conformity was at its highest when responses were seen by others – due to what they

termed normative influence – but that even responses given in private could be influenced by the behavior of others – what they termed informational influence. In this and subsequent investigations, the extent to which behavior was influenced by true changes in belief as opposed to public desires to conform must be inferred comparing behavior that is influenced only by private factors (anonymous behavior) from behavior that is influenced by both private and public factors; subtracting the former from the latter results in the amount of private attitude change. While this strategy is experimentally elegant, the ideal evidence for this theory would be to assess both normative and informational and influence simultaneously, to show that each has a separate and dissociable impact.

Brain imaging research offers just this potential to offer support for this two-factor theory of the impact of social influence on behavior. One of the key theoretical questions lingering from Asch's initial investigation, for example, was whether participants were merely conforming to the confederate's answers, or whether they came to literally see the wrong answers as correct. In a recent functional magnetic resonance imaging (fMRI) investigation of conformity in which confederates gave false answers to a mental rotation task, Berns et al. (2005) demonstrated that the erroneous responses of others altered activity in brain regions implicated in mental rotation, suggesting not just a social impact of conformity pressures, but a true change in perception (see Sherif, 1936). These results offer evidence that social influence can be indexed at the level of the brain, and that examining different regions known to be implicated in different processes may be a fruitful avenue to test the two-factor theory of social influence.

While the research reviewed above has focused primarily on conformity to tasks (line judgments and mental rotations), we use brain imaging to explore these dynamics in the domain of preferences, examining how the brain responds when participants are confronted by symbols that they previously learned were socially valued or socially rejected by others. In the spirit of previous studies on social influence, we consider both normative and informational aspects of such pressures, assessing brain activity in a) regions involved with processing the opinions and mental states of others – the normative aspects of social influence and b) regions implicated in experienced utility or reward – the informational aspects of social influence. By indexing activity in these different regions, we hope to offer evidence in support of the theory that informational and normative influence have dissociable impact on changes in preferences due to social pressure.

Neural Mechanisms of Social Influence

Normative Influence. We expected the processing of “socially tagged” objects – objects about which participants had seen others express an opinion (i.e., endorse or reject) – to be associated with significantly greater activity in regions that play a role in representing and decoding the mental states (e.g., beliefs, desires and preferences) of other people. A large body of neuroimaging research points to the central role of the medial prefrontal cortex (mPFC) in intuiting other people’s mental states and in social thinking more generally (for reviews see Amodio & Frith, 2006; Blakemore & Decety, 2001; Gallagher & Frith, 2003; Mason & Macrae, 2008). For example, enhanced activity in this region has been detected when people consider the motivations underlying other’s

actions (de Lange, Spronk, Willems, Toni, & Bekkering, 2008; Mason, Banfield, & Macrae, 2004), their feelings (Ochsner et al., 2004), and their fears (Olsson, Nearing, & Phelps, 2007). The region is also active when people reflect on the impression they have made on others (Izuma, Saito, & Sadato, 2008). Particularly relevant to the present investigation is functional brain imaging (fMRI) evidence that the mPFC plays a role in the processing of social norms, including perceiving norm violations (Berthoz, Armony, Blair, & Dolan, 2002) and experiencing embarrassment or guilt for the social norm transgressions of others (Takahashi et al., 2004). Dovetailing with these brain imaging findings is neuropsychological evidence that patients with damage to regions of the mPFC exhibit social disinhibition due to a lack of awareness or concern for social norms (Beer et al., 2003; Eisenberg, 2000).

Taken together, previous research suggests that the mPFC plays a central role in representing and processing information about the beliefs, attitudes and feelings of other people. We therefore predicted that people would exhibit greater recruitment of the mPFC when exposed to objects that they had learned were socially valued or devalued, compared to objects that had not been similarly socially tagged.

Informational Influence. We also explored whether social influence impacts the experienced utility – the reward value – of socially tagged objects. Brain imaging research (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Elliott, Friston, & Dolan, 2000; Knutson, Adams, Fong, & Hommer, 2001; McClure, Laibson, Loewenstein, & Cohen, 2004; O’Doherty, Deichmann, Critchley, & Dolan, 2002) and electrophysiology work with non-human primates (Schultz, Apicella, Scarnati, & Ljungberg, 1992; Tremblay, Holerman, & Schultz, 1998) converge on a central role of the striatum – the nucleus

accumbens (nACC), putamen, and caudate – in representing the value of primary reinforcers, or stimuli with inherent reward such as cocaine, but also secondary reinforcers, stimuli that are rewarding by association or that have acquired value through acculturation, such as art (Vartanian & Goel, 2004). Thus striatal regions are active not just when people view inherently attractive faces (Aharon et al., 2001), but also when they view their significant other, whether conventionally attractive or not (Aron et al., 2005).

Particularly relevant to the present investigation is recent evidence that recruitment of these regions is impacted by exposure to products that people have come to value as the result of marketing actions, even holding the objective utility of those objects constant. For instance, people exhibit significantly greater activity in reward regions when they consume a wine they believe to be expensive relative to when they consume that same wine priced more modestly (Plassman, O’Doherty, Shiv & Rangel, 2008) due to their having learned the general rule that price is a signal of quality (Rao & Monroe, 1989). In addition, reward regions are more active when people consume drinks that they believe are made by a company whose brand they have learned to value than when they consume that same drink under a different brand name (McClure et al., 2004). These results suggest that reward regions code not only for the inherent value of a product but also for the utility from learned associations like price and brand information. While both price and brand are in part socially constructed – people’s preferences for expensive branded products is related to a desire for social utility – we explore whether these cortical areas are sensitive to value created solely and specifically from learning the opinions of others. Thus while we predicted that activity in the mPFC would track with

whether objects had been socially *tagged* or not, we predicted that activity in reward areas such as the striatum (nACC, putamen, and caudate) would vary as a function of the valence of that tagging – whether participants had learned that objects were socially *valued* or not.

Overview of the Experiment

Prior to collecting functional imaging data, we exposed participants to information about others' preferences for abstract symbols in a “social influence” phase. During this phase, participants learned that some of the abstract symbols were “popular” (preferred by others on 90% of trials) while others were “unpopular” (preferred on just 10% of trials). We then assessed brain activity using fMRI when people were exposed to these socially tagged (popular and unpopular) symbols, as well as new symbols about which they had not been provided social information. We expected that regions implicated in social processing (mPFC) and reward (striatum, including nACC, putamen, and caudate) would be differentially active depending both on whether symbols had been socially tagged and on the valence of that tagging.

Method

Participants and Design. Twelve male participants from a university community completed the experiment for course credit or monetary compensation. All participants were strongly right-handed as measured by the Edinburgh handedness inventory (Raczkowski, Kalat, & Nebes, 1974), reported no significant abnormal neurological history, and had normal or corrected-to-normal visual acuity.

The experiment had a single factor (symbol type: popular, unpopular, new) repeated-measures design. Each participant completed a pre-scanning “social influence” phase and was then scanned while performing a delayed match-to-sample task with the three symbol types as target items.

Stimulus Materials. We selected 30 abstract symbols for use in the experiment (see Figures 1A and 1B for examples). Of the 30 symbols, we randomly selected twenty for use in the pre-scanning portion of the experiment – 10 for the popular condition and 10 for the unpopular condition. The remaining 10 were used in the new symbol condition and thus were not seen by participants until they performed the delayed match-to-sample task while in the fMRI magnet.

We also created a set of 200 faces to use in the pre-scanning “social influence” phase.

Pre-Scanning Phase: Instantiating Social Influence. Participants were told that hundreds of people had been asked to make preference judgments on pairs of abstract symbols, and that they would see pictures of these people accompanied by a visual depiction of the choice each of these people made (see Figure 1A). Participants were asked to form an impression of which symbols were liked by which individuals, and were told that, though they would not explicitly be tested on the information, it was important for them to pay close attention to the task.

During this “social influence” phase, symbols from the popular set were depicted as being chosen (as indicated by a green box that appeared around the chosen symbol) 90% of the time, while symbols from the unpopular set were depicted as being chosen 10% of the time. On half of the trials the chosen item appeared on the left-hand side of

the screen while on the other half, the chosen item appeared on the right. This design allowed us to control for familiarity – popular and unpopular symbols were presented with the same frequency – and manipulate only the social value attached to those symbols, by varying only which symbol was highlighted.

Participants completed a total of 200 trials that lasted 2500ms in duration. Each trial began with a face – randomly selected from the set of 200 – displayed at the center of the screen. After 750ms, two symbols appeared below the face; a green box indicating the person's preferred symbol appeared around one of the symbols 250ms later. The face and symbols then remained on the screen for an additional 1500ms (see Figure 1A).

Participants were instructed to indicate whether the chosen symbol appeared on the left or right via a key press.

Scanning Phase: Assessing Social Influence.

Participants were told that they would see symbols – some that they had seen previously, some new – appear at the center of the screen, interspersed with fixation crosses, and that their task was to press the button when they saw the same symbol twice in a row (a standard delayed match-to-sample task).

Participants were scanned in two event-related functional (EPI) runs. To bring the spin system into a steady state, four dummy shots were acquired at the start of each of the two runs. A total of 147 volumes were collected within each EPI run. Each experimental trial lasted for a total of 2500ms. Trials were pseudo-randomized within each of the two runs and the run order presentation was counterbalanced across participants (see Figure 1B). In each of the two EPI runs, participants saw new symbols on 36 trials, popular symbols on 27 trials, and unpopular symbols on 27 trials. Nine of these 90 trials – or 10%

– involved symbol repeats (participants seeing the same symbol two trials in a row) and thus elicited button responses from the participants. The remaining 57 EPI volumes were jittered catch trials (i.e., fixation symbols, “+”) used to optimize estimation of the event-related BOLD response. The stimuli were presented using PsyScope (version 1.2.5) and back projected with an Epson (ELP-7000) LCD projector onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Pillow and foam cushions were placed within the head coil to minimize head movements.

Image Acquisition. All images were collected using a 1.5T GE Signa scanner with standard head coil. T1- weighted anatomical images were collected using a 3-D sequence (SPGR; 128 sagittal slices, TR = 7 ms, TE = 3 ms, prep time = 315 ms, flip angle = 15°, FOV = 24 cm, slice thickness = 1.2 mm, matrix = 256 x 192). Functional images were collected with a gradient echo EPI sequence (each volume comprised 25 slices; 4.5 mm thick, 1mm skip; TR = 2500 ms, TE = 35 ms, FOV = 24 cm, 64 x 64 matrix; 90° flip angle).

fMRI Analysis. Functional MRI data were analyzed using Statistical Parametric Mapping software (SPM5, Wellcome Department of Cognitive Neurology, London, UK; Friston et al., 1995). For each functional run, data were preprocessed to remove sources of noise and artifact. Preprocessing included slice timing and motion correction, coregistration to each participant’s anatomical data, normalization to the ICBM 152 brain template (Montreal Neurological Institute), and spatial smoothing with an 8 mm (full-width-at-half-maximum) Gaussian kernel. Analyses took place at two levels: formation of statistical images and regional analysis of hemodynamic responses.

For each participant, a general linear model with 8 regressors was specified. The model included regressors specifying the three conditions of interest – popular, unpopular, and new symbols (modeled with functions for the hemodynamic response), a fourth regressor modeling the button responses/repeat trials (modeled with functions for the hemodynamic response), a fifth regressor to distinguish between the two EPI runs (modeled with a constant), two regressors modeling scanner drift (modeled as linear trends), and the constant term that SPM automatically generates and includes in the model. The general linear model was used to compute parameter estimates (β) and t -contrast images for each comparison at each voxel. These individual contrast images were then submitted to a second-level, random-effects analysis to obtain mean t -images.

Results

Pilot Behavioral Study

To confirm that our social tagging manipulation impacted liking for the stimuli, we conducted a behavioral version of the pre-scanning social influence phase described above, in which participants ($N = 32$) rated all symbols on a 5-point scale both before and after the social influence phase. To measure the change in liking as a result of the manipulation, we calculated difference scores by subtracting pre-ratings from post-ratings, such that positive numbers indicate an increase in liking and negative numbers a decrease. As expected, popular symbols were rated significantly higher after the social influence phase than before ($M_{\text{difference}} = .39$, $SD = .82$), $t(31) = 2.67$, $p < .02$, while attitudes towards unpopular symbols showed a marginally significant decrease in the post influence phase ($M_{\text{difference}} = -.29$, $SD = .82$), $t(31) = 1.99$, $p = .056$, such that the change

in liking between popular and unpopular symbols was significantly different, $t(31) = 4.78, p < .001$.

fMRI Analyses – Direct Comparisons

To identify brain regions that were sensitive to stimuli that had previously been socially tagged, we conducted a direct contrast between symbols that were viewed during the social influence phase – both popular and unpopular – compared to the new symbols that had not been socially tagged. A single cluster in an aspect of the medial prefrontal cortex (mPFC; BA10) exhibited significantly greater activity ($p < .001$; $k = 10$) when participants viewed symbols which had been socially tagged relative to when they viewed new symbols (Table 1). The opposite contrast – comparing new to socially tagged symbols – revealed significantly greater activity ($p < .001$; $k = 10$) in a cluster that spanned the right superior occipital gyrus and precuneus (BAs7, 19, 31).

Next, we explored differences in brain activity specifically between popular and unpopular symbols. Results of a direct comparison between these two conditions revealed significantly greater activity in the right caudate – part of the striatum – when participants viewed popular relative to unpopular symbols ($p < .001$, $k = 10$; Table 2). The opposite contrast – unpopular symbols $>$ popular symbols – revealed no cortical regions that exhibited greater signal to unpopular symbols at the same threshold.

Region of Interest (ROI) Analyses

To explore our effects further, we conducted follow-up ROI analyses, by building and dropping 10 mm spheres over the coordinates of interest, extracting the % signal

change with the tools provided by the SPM5 MarsBar utility (Brett, Anton, Valabregue, & Poline, 2002) for each of the conditions in each ROI, and then averaging these values across all participants.

We first examined the difference in recruitment of mPFC we observed in the direct comparison of symbols from the social influence phase – both popular and unpopular – to the new symbols that had not been socially tagged. As is characteristic of this brain area (Shulman et al., 1997), the ROI analysis revealed that mPFC results were driven by a difference in signal decrease relative to baseline (Figure 2). The signal in this region decreased significantly less to socially tagged symbols than to new symbols, $t(11) = 3.35, p < .01$.

Next, we explored the difference in recruitment of the caudate we had observed between popular and unpopular symbols. The ROI analysis demonstrated that activity in the right caudate was significantly greater for popular symbols relative to unpopular symbols, $t(11) = 3.19, p < .01$, while there was no significant difference in signal between popular symbols and new symbols observed in this area, $t(11) < 1, ns$ (Figure 2).

Whereas the results obtained in our pilot behavioral study demonstrated that participants liked popular symbols more than unpopular symbols after the social influence phase, results from the brain imaging portion of the experiment – which allow us to compare reward value for popular and unpopular symbols to reward value for novel symbols – suggest that the post social influence difference that emerged was driven more by a decrease in the value of unpopular shapes than an increase in value of popular shapes.

General Discussion

We explored the neural mechanisms of social influence, examining how manipulating the popularity of abstract symbols – our rough proxy for the kinds of stimuli that are socially tagged in the real world, like the Lacoste crocodile – impacted brain activity when viewing those symbols.

As expected, mPFC, a brain region involved in thinking about the attitudes and preferences of others (see Amodio & Frith, 2006), was more active when participants viewed symbols that had been socially tagged than symbols for which they had no prior social information, suggesting a possible index of normative influence at the level of the brain. Interestingly, activity in this region did not differentiate between popular and unpopular symbols, but rather only between socially tagged and new symbols. This suggests that this area plays a general role in tracking whether preferences are socially relevant or not; indeed, knowing whether an item is disliked by one's peers is as important as knowing whether an item is liked. Also as predicted, we found that a region involved in the experience of reward – the caudate, part of the striatum – exhibited greater activity in response to popular symbols relative to unpopular symbols, providing a possible index of informational influence at the level of the brain.

These results suggest that predicting whether a symbol is socially valued may require integrating data from both the mPFC and the caudate. Looking solely at the mPFC reveals only whether a symbol is socially tagged or not, while looking solely at the caudate reveals only whether a symbol is liked or disliked; only by looking at both regions together can we identify those symbols that have become valued as a result of social influence. In addition, hearkening back to the models of conformity put forth by

Asch (1951) and Deutsch and Gerard (1955), these results support the notion that informational and normative influence are distinct components of the overall impact of social influence on preferences.

Although activity in the caudate distinguished between symbols that were popular and unpopular, it is worth highlighting the nature of this difference. Results of our ROI analyses indicated that the signal discrepancy was largely driven by an attenuation of activity in the caudate when participants encountered unpopular symbols; no significant difference between new and popular symbols emerged in this region. First, these results address a possible critique of the experimental paradigm that we employed in the present investigation. It is possible that by highlighting popular objects in the social influence phase, we drew participants' attention to popular symbols more than to unpopular symbols, an account that would imply that responses to unpopular symbols – those that participants ignored and thus were not exposed to – would be similar to those of new symbols; these results, however, suggest this is not the case.

More broadly, these results imply that the caudate may be more attuned to information about what is rejected than what is preferred by others, a finding with important implications for researchers of consumer behavior. First, this attunement may translate into an asymmetry in the speed with which people come to exhibit a preference for a socially endorsed item and the rate at which people come to reject a socially undesirable item over time, which can inform existing models of product adoption and abandonment (Rogers, 1962). Second, marketers are increasingly relying on social networks as part of the marketing mix, as evidenced by the increased popularity of word-of-mouth consulting services such as BzzAgent and similar in-house services such as

Proctor and Gamble's Tremor. Most of these campaigns take the form of social endorsement of products, rather than denigration of competing products. More broadly, assessing the effectiveness of such social networking campaigns is notoriously difficult, since the recipient of word-of-mouth can wait for days, weeks, or even months before making a purchase. Our results suggest that marketers might bring consumers into an imaging session and expose them to products, using activity in reward regions (such as the caudate) to assess pure liking but also using activity in mPFC to assess the extent to which those products have been socially tagged or not – a possible index of the effectiveness of word-of-mouth marketing.

We have focused on two particular functions of the mPFC and caudate – in representing social information and reward, respectively – but both regions of course subserve a number of other functions. Most salient as an alternative account for our brain imaging results, both regions consistently exhibit greater BOLD activity during easy versus difficult tasks. The mPFC has been shown to be more active when people engage in tasks with minimal processing demands (Buckner et al., 2008; Gusnard & Raichle, 2001; McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003; Shulman et al., 1997), such as overlearned button press tasks (Mason et al., 2007); if processing socially tagged symbols is easier than processing new symbols, then our mPFC results might be an artifact of task difficulty. Similarly, some research points to a role for the caudate in the control and coordination of movements, including learning the relationship between a triggering stimuli and its appropriate response (e.g., Seger & Cincotta, 2006); if learning about popular symbols is easier than learning about unpopular symbols, then our caudate results might be an artifact of this form of task difficulty.

We believe this alternative explanation is unlikely to account for our results for two reasons. First, from a methodological standpoint, the task participants performed in the imaging portion of the experiment was minimally demanding across all types of symbols, consisting primarily of passively viewing symbols and responding with a button press on only a small percentage (10%) of the trials. Indeed, button press response times did not differ across the three symbol types ($M_{unpopular} = 881\text{ms}$, $M_{popular} = 865\text{ms}$, $M_{new} = 884\text{ms}$), $F(2, 30) < 1$, suggesting that the symbols elicited similar levels of processing effort. Second, we observe differences in caudate recruitment between popular and unpopular symbols, but no difference in mPFC recruitment between these two types of symbols: If task difficulty underlies caudate differences between popular and unpopular symbols, we might expect to observe similar differences in mPFC recruitment for these two types of symbols, but do not (Figure 2). Following the same logic, if differences in mPFC recruitment between socially tagged and novel symbols are due to task difficulty, we might expect to observe similar differences in caudate recruitment, but again do not (Figure 2). Thus while we cannot conclusively rule out the possibility that task difficulty accounts for our effects, the pattern of results elicited by the three symbol types across the caudate and mPFC more closely matches our account.

Finally, we have drawn sharp distinctions between normative and informational influence throughout this paper, and stressed the specificity of the brain regions underlying these different forms of social influence. In practice, of course, these two kinds of influence are frequently interlinked (opinions expressed to conform to the attitudes of others can over time become internalized preferences, for example), and the brain regions underlying these processes likely work in concert as well. Indeed, the fact

that we observe activation in the caudate as opposed to other reward-related regions such as the nucleus accumbens points to the possible interrelatedness of the systems: Several recent investigations have provided evidence for a role of the striatum in representing reward that is socially derived (see Sanfey, 2007). King-Casas et al. (2005), for example, demonstrated that the expectation of fair treatment from an interaction partner in an ultimatum game was associated with increased activity in the caudate (see also Fliessbach et al., 2007; Tabibnia, Satpute, & Lieberman, 2008). Similarly, some evidence suggests that aspects of the mPFC may be involved in coding for reward (e.g., McClure et al., 2004). Izuma et al. (2008) provide evidence of the separable but interrelated roles of caudate and mPFC: In their investigation, activity in the caudate tracked with the magnitude of both social and monetary reward, but the mPFC was selectively activated when participants processed feedback about the impression they made on others (i.e., considered their perspective), regardless of valence.

Like our investigation which modeled the formation of social preferences in a time-limited laboratory session, these other investigations are similarly short in duration, precluding an exploration of the time course of the interplay of brain regions over time. Future research should explore the process by which learning the attitudes of other people translates into stable preferences in memory: Brain regions involved in episodic and short-term memory might play a dominant role during early stages of the social influence process, while regions involved in storing semantic, general knowledge about the world might be more involved as those novel preferences become more stable.

One intriguing possibility would be to explore a role for areas of the right hippocampus implicated specifically in memory for social information (e.g., Kirwan & Stark, 2004; Somerville, Wig, Whalen, & Kelley, 2006).

Conclusion

We close by highlighting how the minimalistic nature of our paradigm speaks to the robustness of the power of social influence. First, we observed our effects – greater recruitment of the mPFC to socially tagged symbols and greater caudate activity to popular symbols – despite the fact that participants were never explicitly directed to judge the symbols during the scanning session, instead performing an incidental processing task. This suggests that the impact of social influence is present and detectable even in the absence of any conscious intention on the part of the participant to consider the value of the target item (e.g., “Do I like this symbol?”) or any instruction to consider the previous social context in which this information was encountered (e.g., “How popular was this symbol?”). These results are particularly notable in light of previous evidence that relevant norms must be salient to elicit norm-congruent behavior (Cialdini, Kallgren, & Reno, 1991). Second, our effects emerged at the level of the brain after just a short training duration, involving abstract symbols tagged by the opinions of strangers, and in a setting that was decidedly not naturalistic (an MR scanner). We can only imagine the magnitude of our effects in more familiar contexts, using more meaningful stimuli, and involving social tagging by non-strangers, all important areas for future research.

References

- Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review*, *16*, 586-612.
- Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, *32*, 537-551.
- Amodio, D. M. & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268-277.
- Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology*, *95*, 918-932.
- Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L.L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology*, *94*, 327-337.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (ed.) *Groups, Leadership and Men*. Pittsburgh, PA: Carnegie Press.
- Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, *103*, 797-818.
- Beer, J. S., Heerey, E. H., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*, 594-604.

- Berger, J. & Heath, C. (2007). Where consumers diverge from others: Identity-signaling and product domains. *Journal of Consumer Research*, 34, 121-134.
- 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, 245-253.
- Berthoz, S., Armony, J. L., Blair, R. J. J., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, 125, 1696-1708.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural-change as informational cascades. *Journal of Political Economy*, 100, 992-1026.
- Blakemore, S. J. & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561-567.
- Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002, August). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default system: Anatomy, function, and relevance to disease. In A. Kingstone & M. Miller (Eds.), *The Year in Cognitive Neuroscience 2008* (Vol. 1124, pp. 1-38), Ann, NY: New York Academy of Sciences.
- Camerer, C.F. (2003). *Behavioral game theory*. Princeton, NJ: Princeton University Press.

Cialdini, R. B. & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591-621.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 201-234). New York: Academic Press.

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*, 1015-1026.

Craig, K. D. & Prkachin, K. M. (1978). Social modeling influences on sensory decision theory and psychophysiological indexes of pain. *Journal of Personality and Social Psychology, 36*, 805-813.

Delgado, M. R., Nystrom, L. E., Fissell, K., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses for reward and punishment in the striatum. *Journal of Neurophysiology, 84*, 3072-3077.

Deutsch, M. & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology, 51*, 629-36.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Reviews in Psychology, 51*, 665-697.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience, 20*, 6159-6165.

- Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, *318*, 1305-1308.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. J. S. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.
- Gallagher, H. L. & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences*, *7*, 77–83.
- Gerber, A.S., Green, D.P., & Larimer, C.W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, *102*, 33–48.
- Gusnard, D. A. & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Neuroscience*, *2*, 685-694.
- Higgins, M. C. (2001). Follow the leader?: The effects of social influence on employer choice. *Group & Organization Management*, *26*, 255-282.
- Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock market participation. *The Journal of Finance*, *59*, 137-163.
- Izuma, K., Saito, D.N., & Saldato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*, 284-294.
- Kilduff, M. (1990). The interpersonal structure of decision making: A social comparison approach to organizational choice. *Organizational Behavior and Human Decision Processes*, *47*, 270-288.

- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*, 78-83.
- Kirwan, C. B. & Stark, C. E. (2004). Medial temporal lobe activation during encoding and retrieval of novel face–name pairs. *Hippocampus, 14*, 919–930.
- Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience, 21*, RC159.
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I. & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology, 18*, 454–457.
- Mason, M. F., Banfield, J. & Macrae, C. N. (2004). Thinking about actions: The neural substrates of person-related action knowledge. *Cerebral Cortex, 14*, 209-214.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315*, 393-395.
- Mason, M. F. & Macrae, C. N. (2008). Perspective-taking from a social neuroscience standpoint. *Group Processes & Intergroup Relations, 11*, 215-232.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science, 306*, 503-507.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*, 379-387.

- McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation: An fMRI study. *Journal of Cognitive Neuroscience, 15*, 394-408
- O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33*, 815– 826.
- Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., & Mackey, S. C. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*, 1746–1772.
- Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fear by observing others: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*, 3-11.
- Plassman, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, 105*, 1050-1054.
- Raczkowski, D., Kalat, J. W., Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia, 12*, 43-47.
- Rao, A. & Monroe, K.B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research, 26*, 351–357.
- Reingen, P. H. (1982). Test of a list procedure for inducing compliance with requests. *Journal of Consumer Research, 5*, 96-102.
- Rogers, E. M. (1962). *Diffusion of Innovations*. The Free Press. New York.

- Sanfey, A.G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*, 598-602.
- Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *Journal of Neuroscience*, *12*, 4595-4610.
- Seger, C. A. & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, *16*, 1546-1555.
- Sherif, M. (1936). *The psychology of social norms*. New York: Harper Collins.
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience*, *9*, 648-663.
- Somerville, L.H., Wig, G.S., Whalen, P.J., & Kelley, W.M. (2006). Dissociable medial temporal lobe contributions to social memory. *Journal of Cognitive Neuroscience*, *18*, 1253-1265.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*, 339-347.
- Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: An fMRI study. *NeuroImage*, *23*, 967-974.
- Tremblay, L., Hollerman, J. R., Schultz, W. (1998). Modifications of reward expectation-related neuronal activity during learning in primate striatum. *Journal of Neurophysiology*, *80*, 964-977.

Vartanian, O. & Goel, V. (2004). Neuroanatomical correlates of aesthetic preferences for paintings. *NeuroReport*, *15*, 893-897.

Whiten, A., Horner, V., & de Waal, F. B. M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, *437*, 737-740.

Table 1. All clusters that exhibited differential activity when participants viewed socially tagged relative to new symbols. Coordinates are reported in Talairach space. The displayed t-values are associated with each area's peak hemodynamic response, $k = 10$; $p < .005$. BA = Brodmann Area; B = bilateral; R = right; L = left; * = regions that emerged at a slightly reduced threshold of $k = 10$; $p < .001$.

Socially Tagged Symbols > New Symbols

k	Anatomical Location	BA	coordinates			t-value
			x	y	z	
53	(R) medial frontal gyrus	10	6	47	9	5.46*

New Symbols > Socially Tagged Symbols

k	Anatomical Location	BA	coordinates			t-value
			x	y	z	
16	(L) superior occipital gyrus	19	-30	-69	28	6.67
87	(R) precuneus	7	18	-68	34	6.34*
	(R) superior occipital gyrus	31	24	-69	26	5.11
18	(R) inferior temporal gyrus	37	45	-67	-2	4.11
10	(R) caudate body		18	-16	20	3.47

Table 2. All clusters that exhibited differential activity when participants viewed popular relative to unpopular symbols. Coordinates are reported in Talairach space. The displayed t-values are associated with each area's peak hemodynamic response, $k = 10$; $p < .005$.

BA = Brodmann Area; B = bilateral; R = right; L = left; * = regions that emerged at a slightly reduced threshold of $k = 10$; $p < .001$.

Popular Symbols > Unpopular Symbols

k	Anatomical Location	BA	coordinates			t-value
			x	y	z	
33	(R) caudate		12	-2	19	5.17*
15	(L) superior frontal gyrus	10	-21	65	11	4.15

Unpopular Symbols > Popular Symbols

k	Anatomical Location	BA	coordinates			t-value
			x	y	z	
None.						

Figure 1. (A) Timeline depicting an experimental trial during the “social influence” pre-scanning phase. Each trial began (at 0ms) with a face displayed at the center of the screen. At 500ms, the fixation cross was replaced by a target face. At 1250ms, two symbols appeared on the screen below the face. One of the symbols – the target’s preferred symbol – was surrounded by a green box. The face and the symbols remained on the screen for 1250ms until the end of the trial.

(B) Timeline depicting an fMRI experimental run. Participants pressed the button when they saw the same symbol depicted twice in a row (delayed match-to-sample task). To tease apart the signal associated with each of the experimental conditions (new, popular and unpopular shapes), fixation crosses were “jittered” among the shapes.

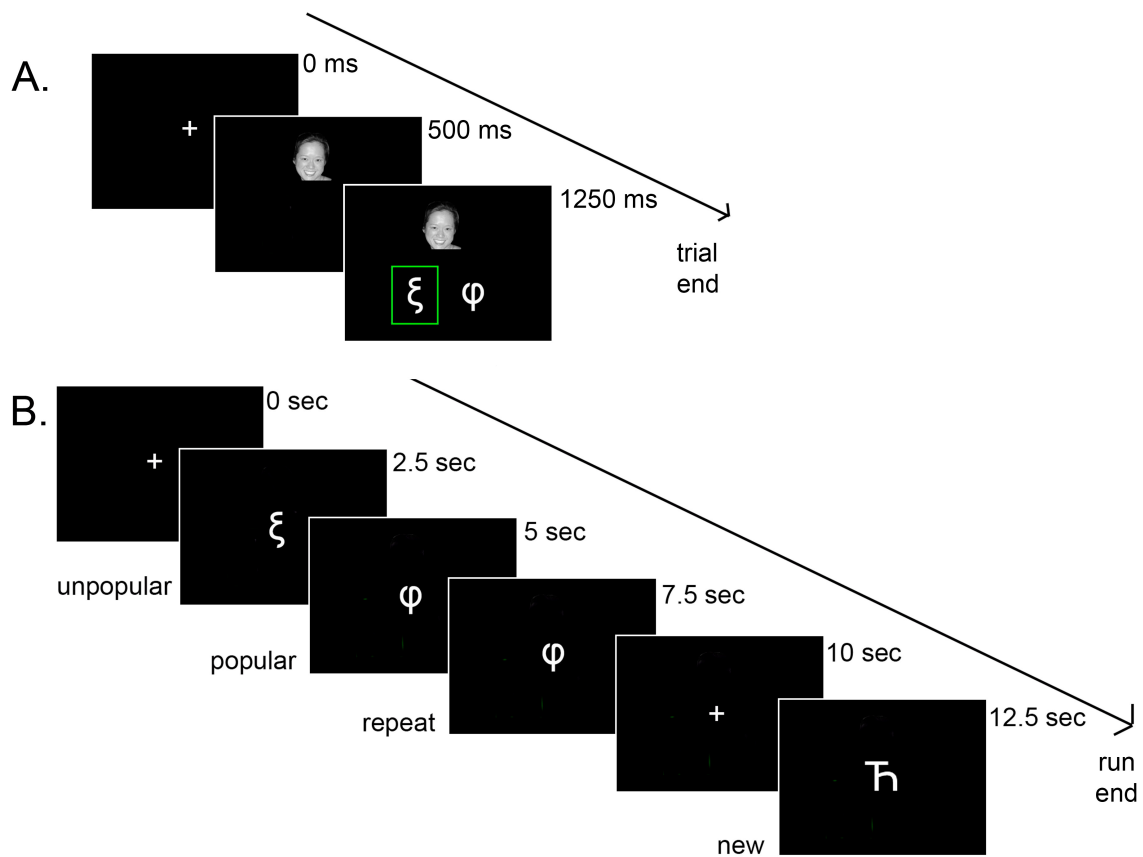


Figure 2. Results of ROI analyses in: bilateral medial prefrontal cortices (top) and the right caudate (bottom). Values were computed by dropping a 10 mm sphere and pulling the % signal change, and then averaging across all participants. Images depict the results of direct comparisons: “socially tagged symbols > new symbols” (top) and “popular > unpopular symbols” (bottom), displayed on the average anatomical high-resolution image in neurological convention.

