

Sticking with the nice guy: Trait warmth information impairs learning and modulates person perception brain network activity

Victoria K. Lee · Lasana T. Harris

Published online: 13 May 2014
© Psychonomic Society, Inc. 2014

Abstract Social learning requires inferring social information about another person, as well as evaluating outcomes. Previous research shows that prior social information biases decision making and reduces reliance on striatal activity during learning (Delgado, Frank, & Phelps, *Nature Neuroscience* 8 (11): 1611–1618, 2005). A rich literature in social psychology on person perception demonstrates that people spontaneously infer social information when viewing another person (Fiske & Taylor, 2013) and engage a network of brain regions, including the medial prefrontal cortex, temporal parietal junction, superior temporal sulcus, and precuneus (Amodio & Frith, *Nature Reviews Neuroscience*, 7(4), 268–277, 2006; Haxby, Gobbini, & Montgomery, 2004; van Overwalle *Human Brain Mapping*, 30, 829–858, 2009). We investigate the role of these brain regions during social learning about well-established dimensions of person perception—trait warmth and trait competence. We test the hypothesis that activity in person perception brain regions interacts with learning structures during social learning. Participants play an investment game where they must choose an agent to invest on their behalf. This choice is guided by cues signaling trait warmth or trait competence based on framing of monetary returns. Trait warmth information impairs learning about human but not computer agents, while trait competence information produces similar learning rates for human and computer agents. We see increased activation to warmth information about human agents in person perception brain regions.

Both authors contributed equally to the design, analysis, and writing of this project.

V. K. Lee (✉) · L. T. Harris
Department of Psychology and Neuroscience, Duke University,
Durham, NC, USA
e-mail: victoria.k.lee@duke.edu

L. T. Harris
Center for Cognitive Neuroscience, Duke University, Durham, NC,
USA

Interestingly, activity in person perception brain regions during the decision phase negatively predicts activity in the striatum during feedback for trait competence inferences about humans. These results suggest that social learning may engage additional processing within person perception brain regions that hampers learning in economic contexts.

Keywords Social learning · Person perception · Social cognition · Trait warmth/competence

Learning from the physical world and learning from the social world involve different levels of complex processing. Suppose that a person inserts one dollar into a slot machine that operates on probabilistic outcomes. There are two possible outcomes of this action—a positive gain of three dollars or a negative loss of the dollar. In this economic scenario, determining whether the action taken was a good decision depends singly on learning the probabilities of the slot machine, a process that will be guided by the person's goal to maximize profit. Now imagine a social interaction that depends on the same economic principles. Instead of a slot machine, a person gives another person (the receiver) one dollar. The possible economic outcomes are the same; the receiver can either return three dollars to the person or return nothing and the person loses the dollar. Although the premise of these two economic scenarios is the same, in the social scenario, the inferred traits of *both* parties in the exchange can influence the interpretation of the outcome (gain of three dollars or loss of one), making social learning more complex, or multiply determined.

Learning whether the investment was a good decision in the social scenario relies not only on the economic outcome of the exchange, but also on social factors relevant to the receiver, including his or her identity, goals, and inferred traits. Suppose the receiver is homeless versus one's friend; suddenly gaining three dollars may appear more or less positive, and losing one dollar may appear more or less negative, depending

on the inferred traits of the receiver (e.g., a friend may be perceived as more trustworthy than a homeless person). Inferred traits provide a decision-maker with a concise schema for predicting another person's behavior and have the ability to change the interpretation of the outcome from singular to polymorphic (Uleman, Newman, & Moskowitz, 1996). Thus, the complexity of social learning arises because multiple sources of information must be considered. How, then, can people learn to make decisions in such complex social interactions, incorporating both economic outcomes and social factors into their decision-making processes? It may be that the brain relies on additional processing within the person perception brain network that interacts with traditional learning structures during social learning.

The idea that social and nonsocial learning may differ is supported by other dual-systems approaches to learning. In recent years, researchers have found support for models of “hot” automatic versus “cool” controlled decision making (Bernheim & Rangel, 2004), model-free versus model-based learning (Gläscher, Daw, Dayan, & O’Doherty, 2010), and associative versus prediction error learning signals (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011). Here, we propose that a dual-systems approach may allow us to better understand how social learning differs from nonsocial learning. While nonsocial learning may rely on decision-making structures including the striatum and medial prefrontal cortex (MPFC), social learning may engage additional brain processing within the person perception network.

As a caveat, we are suggesting not that two separate systems exist but, rather, that an additional network of brain regions is engaged when one thinks about the intentions of the other person and changes the mode of processing from nonsocial to social. Numerous studies have shown that the striatum and MPFC are engaged during social decision making (see below). Why, then, would we suspect that additional brain processing is necessary for social decisions? Social decisions often contradict economic models that attempt to predict social behavior, suggesting that simple reinforcement learning models by themselves are not sufficient to explain this complex social behavior (Lee, McGreevy, & Barraclough, 2005). Studies have repeatedly shown that other important social factors, including one's reputation and perceived moral character (i.e., inferred traits), also influence decision making (Delgado, Frank, & Phelps, 2005; King-Casas et al., 2005; Phan, Sripada, Angstadt, & McCabe, 2010). The social neuroscience literature on person perception has largely investigated how people infer traits and form impressions of others and has identified a network of brain regions engaged in these processes. However, until recently, few studies have looked at how person perception brain regions respond during social learning (see below). Evidence from these previous studies suggests that these social brain regions may influence reward processing and valuation.

Social decision making engages decision-making and person perception brain regions

A growing body of research has explored how decision-making brain regions respond to social stimuli. Social rewards, including being labeled trustworthy by another person (Izuma, Saito, & Sadato, 2008), gaining social approval by donating money to charity in the presence of others (Izuma, Saito, & Sadato, 2010), and viewing pictures of smiling people (Lin, Adolphs, & Rangel, 2012), reliably activate regions of the striatum that are also active to monetary rewards. From this and other studies, it has been suggested that social learning may be similar to basic reinforcement learning (Jones et al., 2011). The striatum's role in reward prediction error is well established in the literature and has become a clear measure of learning in reward paradigms (McClure, Berns, & Montague, 2003; Schultz, Dayan, & Montague, 1997; Schultz & Dickinson, 2000). This learning signal is also exhibited in economic games requiring cooperation and trust (see Sanfey, 2007, for a review) and when a person's behavior violates social expectations (Harris & Fiske, 2010), further supporting the notion that social learning relies on decision-making processes.

The argument for a common valuation scale (Montague & Berns, 2002) also supports the idea that some of the same brain structures are engaged during social and nonsocial learning. When decisions include different types of future rewards, whether they are of a social nature or not, the brain must convert value signals into a “common currency” that will guide future behavior. The ventral MPFC (VMPFC) has been implicated in creating such subjective value signals for both social and monetary rewards (Lin et al., 2012; Smith et al., 2010). Furthermore, activity in this region correlates with the subjective value of making voluntary charitable donations (Hare, Camerer, Knoepfle, O’Doherty, & Rangel, 2010), a paradigm that is often considered social in nature.

However, social learning may not simply be another type of reinforcement learning. Studies have shown that reward and value signals involved in learning are influenced by the social context. For example, prior social information about an investment partner reduces reliance on feedback mechanisms in the striatum (Delgado et al., 2005). Similarly, when one shares rewards, striatal activity is modulated by the identity of the other person and ratings of social closeness (Fareri, Niznikiewicz, Lee, & Delgado, 2012). Interestingly, in a study of vicarious reward, connectivity between the anterior cingulate (ACC) and ventral striatum was positively correlated with similarity ratings with another person (Mobbs et al., 2009), suggesting that the experience of reward may be modulated by connections from other brain regions to the striatum. Furthermore, functional connectivity analyses show that value signals in the VMPFC may integrate information from brain regions implicated in person perception, such as the anterior

insula and posterior superior temporal cortex (Hare et al., 2010). Together, these studies provide growing support for the involvement of person perception brain regions in social decision making and social learning.

But what exactly are these person perception brain regions representing, and how do these representations affect learning signals? A rich literature on person perception suggests that there are different types of social information that may be extracted when viewing another person. These mental state inferences rely on a network of brain regions including the MPFC, superior temporal sulcus (STS), temporal parietal junction (TPJ) and precuneus, as well as the regions investigated in the above-mentioned studies. By examining how these brain regions respond during a social-learning task, we may gain insight into how they affect decision making in a social context.

Linking person perception to social learning

A multitude of psychological processes occur when another person is encountered. Decades of social psychological research on person perception suggest that there are important social factors that should be taken into account when human beings are used as stimuli in psychological experiments (Fiske & Taylor, 2013; Ponkanen, Alhoniemi, Leppanen, & Hietanen, 2011). People extract social category information such as age, race, and gender (Fiske, 1998; Ito, Thompson, & Cacioppo, 2004; Ito & Urland, 2003), as well as assessments of a person's trustworthiness (Willis & Todorov, 2006) and other spontaneous trait inferences (Ambady & Rosenthal, 1992; Todorov & Uleman, 2003) with minimal exposure to other people. This information facilitates an impression of the person—a social-learning process that allows one to make predictions about behavior in future social interactions.

This social learning is supported by mental state inferences—a form of social cognition—and a network of brain regions involved in person perception including, but not limited to, the MPFC, STS, TPJ, and precuneus (Frith & Frith, 2003; Gallagher & Frith, 2003; Haxby, Gobbini, & Montgomery, 2004; van Overwalle, 2009; Young & Saxe, 2009). Although these brain regions work together as a network in social cognition tasks, previous research has highlighted the differential roles each of these regions may play in social cognition. For instance, the MPFC may be engaged when inferring enduring dispositions of others and the self, as well as more temporal states (van Overwalle, 2009). The TPJ has recently been implicated in creating a social context relevant for future behaviors (Carter, Bowling, Reeck, & Huettel, 2012; Carter & Huettel, 2013). The STS is implicated in the perception of socially relevant information (Allison, Puce, & McCarthy, 2000; Zilbovicius et al., 2013). And the precuneus is implicated in self-processing, along with

other mental tasks (Cavanna & Trimble, 2006). Given that impressions are used to generate predictions, it may be that brain regions implicated in person perception interact with brain regions that support learning during social decision making.

Evolutionary pressures have prepared people with the ability to predict the nature of an impending social interaction in order to differentiate friend from foe (Brown & Brüne, 2012) and navigate our complex social world (Fiske & Taylor, 2013). Accordingly, we have developed a keen ability to assess different types of social information that allow for these predictions. The primary dimensions of person perception—trait warmth and trait competence (Asch, 1946; Fiske, Cuddy, & Glick, 2007; Rosenberg, Nelson, & Vivekananthan, 1968)—satisfy this evolutionary motive. While trait warmth describes a person's intentions—whether they are good or ill—trait competence describes a person's ability to carry out those good or ill intentions (Fiske, Cuddy, Glick, & Xu, 2002).

Research suggests that these two types of inferred traits are often assessed together but are not treated equally. Warmth information is judged before competence information and carries more weight when impressions of others are formed (Asch, 1946; Fiske et al., 2007; Wojciszke, Bazinska, & Jaworski, 1998). Evolutionarily, this makes sense: A perceiver must know whether another person intends harm before deciding whether that person is capable of carrying out that harm. Furthermore, asymmetries in the diagnosticity of warmth and competence information may place different weight on positive and negative information in these domains (Reeder, 1993; Reeder & Brewer, 1979; Skowronski, 2002). Negative warmth information (e.g., stealing a wallet) is more informative about a person's traits than is positive warmth information (e.g., returning a lost wallet), because people low in trait warmth can manipulate warmth information in order to be seen positively (immoral people can act morally in a socially desirable context) but only people low in trait warmth will behave negatively in the warmth domain (moral people tend not to act immorally; Reeder & Spores, 1983). On the other hand, positive competence information (e.g., hitting a home run) is more informative than negative competence information (e.g., striking out at the plate), because people high in trait competence show a wide range of behaviors (competent baseball players sometimes strike out), while people low in trait competence rarely show such range in behavior (incompetent baseball players do not hit home runs; Reeder, 1979). Therefore, the number of behavioral incidences needed to change an impression of a person may differ in the warmth and competence domains (Tausch, Kenworthy, & Hewstone, 2007). However, despite the primacy of trait person perception dimensions, few studies (Baron, Gobbini, Engell, & Todorov, 2011; Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009) have investigated how person perception and the associated brain regions are modulated during social learning.

The hypothesis that these learning processes will be influenced by the social context is not unjustified. Previous research documents a variety of heuristics often used in situations of uncertainty that bias decision making (Tversky & Kahneman, 1974, 1986). Likewise, when a person is encountered for the first time—a situation bound with uncertainty—assessments of social information such as trait warmth and competence may also function as a heuristic that influences social learning. For instance, a person perceived as high in trait warmth may be predicted to be amicable in a social interaction. This prediction is based on a heuristic (stereotype or schema) about warm people. If that heuristic leads to an incorrect prediction and the interaction is not amicable, prediction error signals update this information and incorporate it into future decision making (Harris & Fiske, 2010). However, evidence suggests that this learning process may be impaired for social contexts; previous work has shown that trait warmth information affects feedback processing in the striatum, resulting in biased decision making (Delgado et al., 2005). An interesting question yet to be explored is when a person uses heuristics about the warmth and competence of a social target, how do person perception brain regions respond in this social learning context.

The present study

Here, we build on the social-learning literature, examining whether brain regions underlying person perception are modulated during social learning. We investigated this research question, looking specifically at how inferences of trait warmth and competence about a financial partner affect these brain regions within a social-learning context. In a novel investment task, participants selected both human and computer agents to invest in a stock market on their behalf. We framed returned profit (i.e., feedback) in either the trait warmth or trait competence domain for each agent, thus requiring participants to incorporate the social information about the agent into their decision making. We tested how the use of this social information affects participants' ability to learn an underlying rule of the game: unbiased agent selection. Importantly, feedback was programmed so that all agents were equally likely to return a large profit. Therefore, in a rational economic sense, profit-maximizing participants should learn to be unbiased in their agent selection, not relying on inferences of trait warmth or competence to guide agent selection. However, given that the brain evolved in a social context, we hypothesized that social information (trait inferences) would serve as heuristics that biased learning and decision making in this social economic task. Furthermore, given the primacy of warmth information in impression formation, there might be differential effects of warmth, not competence inferences about human actors, on learning. These effects might not

apply to the computer conditions, because unlike people, computers do not have minds, making mental state inferences unnecessary (although computers may be anthropomorphized at times).

Method

Participants

Behavioral data were collected for 50 individuals (34 female, 16 male; age: $M = 26.56$, $SD = 8.38$). Of those 50 participants, fMRI data were collected for 30 right-handed individuals (23 female, 7 male; age: $M = 27.07$, $SD = 8.89$) recruited to participate in this study through Duke University's Brain Imaging Analysis Center participant pool. Five were removed from the sample due to an excessive number (over 25 %) of missed trials. The remaining 45 behavioral participants (25 fMRI participants) were included in data analysis. All participants gave informed consent. Brain-imaging participants received \$30 for their participation, plus any additional winnings, while behavioral participants received \$10 plus any additional winnings.

Procedure and experimental design

Participants were told to imagine that they were interested in investing in the stock market and decided to go to an investment firm where they could select an agent to invest in the stock market on their behalf. The present task allowed them to view all agents in the firm before making their final investment decision. In order to manipulate trait person perception dimensions (warmth and competence), we created two types of investment firms. For *warmth* firms, feedback represented the agents' *generosity* when dividing the profit between themselves and the participants. All agents in these firms were considered high in the competence domain because they all made an equally large amount of profit, which was to be divided at the agents' discretion. In the *competence* firms, feedback represented the agents' *ability* to make good investments in the market. All agents in this firm were considered high in the warmth domain because all profit was returned to the participants.

Both types of firms had human and computer agents. Computer agents were described as computer programs used for reading the stock market and were represented by geometric symbols (e.g., circle, square, triangle, etc.). In the competence condition, the computer programs used algorithms that were used to read the market. A "competent" computer program would be one that read the market well and returned a large profit. In the warmth condition, the computer programs used algorithms that divided the profit, returning some portion of the profit to the participant. A "warm" computer program

would therefore be one that returned a large profit to the participants. In other words, in the warmth condition, the computer programs could be thought to use algorithms that relied on different distributions, and a warm computer would one that used a negatively skewed distribution, returning a higher profit more often than a lower profit. In reality, all the computer programs were programmed to return an average profit, suggestive of a normally distributed distribution.

Within each condition, participants saw the same three agents on each trial, allowing them to tie previous outcomes (feedback) to the agents. Twelve agents were presented in total (three in the human warmth, three in the human competence, three in the computer warmth, and three in the computer competence conditions). We used a miniblock design to present all four conditions (human warmth, human competence, computer warmth, computer competence) within each run, with a total of seven runs in the study (see Fig. 1). An instruction screen at the beginning of each miniblock indicated which condition participants would be playing. Within the miniblock, each trial began with a 4-s decision screen asking participants to select one of three agents to invest on their behalf by pressing the corresponding button. After a jittered anticipation period (2–8 s; see Ollinger, Corbetta, & Shulman, Ollinger et al. 2001a; Ollinger, Shulman, & Corbetta, 2001b), a feedback screen appeared showing the earnings on that round of investment with the agent selected, along with fictive error information about how much they could have earned if they had selected a different agent (Lohrenz, McCabe, Camerer, & Montague, 2007). Finally, a jittered intertrial interval (2–8 s) preceded the next trial.

The feedback screen always presented a high, medium, and low reward. Importantly, this feedback was programmed so that all agents were equally likely to return each reward on the first trial and each agent returned each reward twice throughout a miniblock. Hence, participants should learn that there is no “warm” or “competent” agent on the basis of feedback, suggesting that agent selection should be unbiased. All four conditions were identical (including feedback), with the exception of the identity of the agents, which were unique to each condition. Therefore, any differences in behavior can be attributed to our manipulation of the perception of the social context that originates the reward (i.e., warmth or competence) and the type of agent (i.e., human or computer).

In order to minimize the number of missed trials in the scanner, participants completed a mock version of the task (eight trials) for practice before entering the scanner. The agents used in this mock version were not included in the scanner task, in order to ensure that there would be no preexisting beliefs about the agents seen in the scanner (e.g., that one agent was better than the others). We simply used this mock task to familiarize participants with the task structure. To assess individual differences in preferences for warmth or competence information and human or computer agents, participants completed a questionnaire before and after the scanning session where they reported their preferences for agent type and trait information. They also rated each of the respective agents on either warmth or competence (on a scale of 1 to 7) after the scanning session. They were then thanked for their participation and paid from one trial selected at random, in addition to the hourly rate.

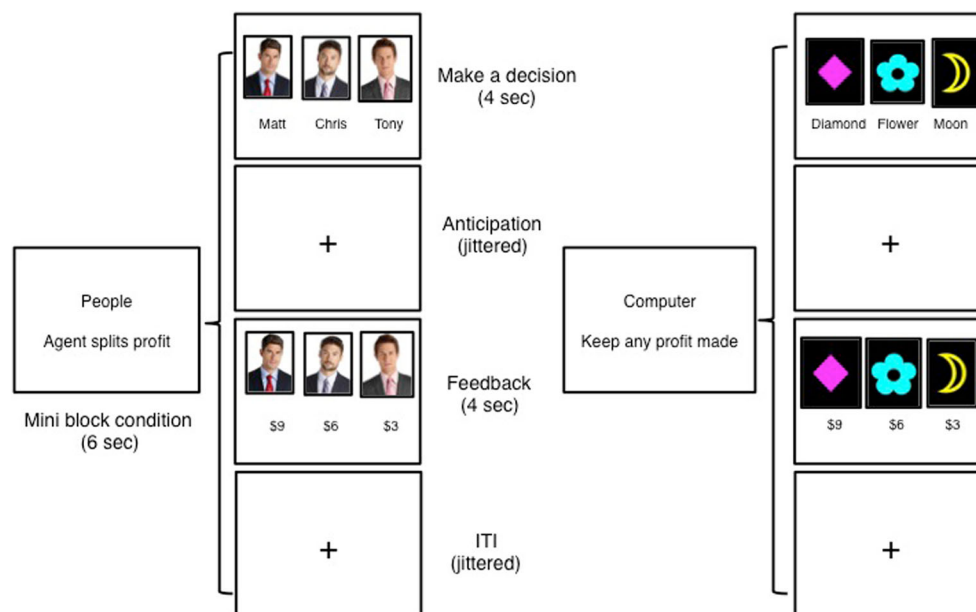


Fig. 1 Experimental design. Two (human warmth and computer competence) of four conditions are shown for display purposes. All four conditions are presented in miniblocks within each run. Each miniblock begins with an

instruction screen indicating the condition, followed by trials consisting of a decision phase where an agent is selected, followed by a jittered anticipation period and by a feedback phase with a jittered intertrial interval to end the trial

Behavioral analysis

We used two different approaches to assess participants' learning—each measuring the extent to which participants *did* or *did not* learn the equated feedback structure of the game. The first approach uses two behavioral measures (bias index and fictive error index) that are inversely related to learning, measuring the extent to which participants used suboptimal strategies when deciding with which agent to invest, suggesting that they did not learn. These two behavioral measures depend on inferences about the mental states of the agents and can be considered more social models. The second approach uses two behavioral measures (last-trial accuracy and running average) that are directly related to learning, measuring the extent to which participants' decisions were consistent with the behavior of a person who learned the feedback structure of the game. Importantly, these two approaches test statistical and economic models of learning, respectively, not social models. Together, these measures provide converging evidence that trait warmth information about people hinders learning.

Bias index

We first assessed learning by looking at the extent to which participants were biased in their agent selection. This analysis relies on the fact that returned profit was equated across the three agents. If participants learned that all agents were equally likely to return a large profit (hence, there is no “warm” or “competent” agent), their agent selection should be random, and they should be less likely to predominantly choose one agent over the other two. In other words, they should choose equally among the three different agents within a condition. Our bias index measures the extent to which participants did not follow this strategy, suggesting that they did not learn the contingencies of the game, instead relying on trait inferences. Within a condition (e.g., human warmth), we calculated the frequency with which the participant selected agent A, agent B, and agent C. The standard deviation of these three frequencies represents how biased participants were when selecting an agent and serves as our bias index. For example, if a participant learned that feedback was equated and was completely unbiased when selecting an agent, he or she would select each agent 14 times in that condition, resulting in a standard deviation of 0. However, if a participant did not learn and was biased toward selecting one agent over the other two (e.g., selecting agent A 17 times, agent B 14 times, and agent C 11 times), the resulting standard deviation would be higher (e.g., 3). This analysis ignores whether participants were choosing the agent that returned the highest or lowest profit and simply measures the extent to which the same agent was selected throughout the game.

One could argue that selecting the same agent throughout the experiment is just as optimal as selecting at random; the expected value of each strategy is equal, because we programmed feedback to be equal across all three agents. While this is true, selecting the same agent throughout the experiment would suggest that participants completely ignored feedback on every trial and blindly chose the agent that they first selected (regardless of whether the agent returned a high, medium, or low reward on this first trial), suggesting that they did not learn the contingencies of the game. None of our participants used the latter strategy throughout the entire experiment.

Fictive error index

Another method we used to measure social learning relies on fictive error information—information about how much could have been earned if a different agent was chosen (Lohrenz et al., 2007). The feedback screen of each trial always presented the profit returned by the agent selected, as well as feedback from the two agents not selected. Therefore, participants always saw the agent who returned the high, medium, and low profit for that trial. For this learning analysis, we looked at whether fictive error information influenced decisions on the following trial. Specifically, we examined whether participants chose the agent that returned the highest profit on the previous trial. This strategy is less than optimal because each agent was programmed to return the highest profit only twice within the six trials of a miniblock. Therefore, if agent A returned the highest profit on trial 1, the probability of agent A returning the highest profit on the next trial was lower than the two agents that had yet to return a high profit. The only time this strategy would be optimal is when an agent returned a high profit on two consecutive trials. However, because we randomized the trial order, the occurrence of such an event would have been by chance. Therefore, adopting a strategy where the participant chose the agent that previously returned the highest profit would suggest that he or she did not learn the structure of the feedback, relying instead on a trait inference relevant to the previous trial.

Last-trial accuracy

We next turned to an analysis that measures the extent to which participants learned the structure of the feedback in the game. Within a miniblock, we programmed each of the three agents to return the highest profit twice within the six trials. Therefore, by the last trial of the miniblock, two agents had already returned the highest profit twice, and one agent had yet to return the highest profit twice. If participants learned that feedback was equated, they should be able to predict which agent had yet to return the highest profit a second time and choose that agent. We measured the extent

to which participants were accurate on the last trial of each miniblock. Higher last-trial accuracy suggests that participants learned the feedback structure of our game and chose the agent that had yet to return the highest profit. This tests a statistical model of learning based on probabilities.

Running average

Alternatively, participants may have strictly relied on feedback (returned profit from the agents) in a purely economic sense, completely ignoring trait inferences that may have resulted from the returned profit. This analysis measures the extent to which participants ignored trait inferences and relied on feedback to calculate a running average of the returned profit from each agent and chose the agent with the highest average. This strategy would suggest that participants took a purely economic approach to playing the game, relying only on returned profit to make their decisions.

fMRI acquisition and data analysis

A 3.0 Tesla GE Signa Excite whole-body scanner was used to collect structural images (T1-weighted MPRAGE: 256×256 matrix; FOV = 256 mm; 116 one-mm sagittal slices) followed by functional images (EPI sequence: TR = 2,000 ms; TE = 25 ms; FOV = 192 cm; flip angle = 75° ; echo spacing = 0.29 ms; 39 slices; voxel size: $3 \times 3 \times 3$ mm³). A computer presented the stimuli projected to a screen mounted at the rear of the scanner bore. Stimuli were reflected through a filter and a mirror, which participants viewed while supine.

BOLD data preprocessing

Both image preprocessing and statistical analysis used Brain Voyager QX (<http://www.brainvoyager.de>). Before statistical analysis, image preprocessing consisted of (1) slice acquisition order correction, (2) 3-D rigid-body motion correction, (3) voxelwise linear detrending across time, (4) temporal band-pass filtering to remove low- and high-frequency (scanner and physiology related) noise. Distortions of EPI images were corrected with a simple affine transformation. Functional images were registered to the structural images and interpolated to cubic voxels. After coregistering participants' structural images to a standard image using a 12-parameter spatial transformation, their functional data were similarly transformed, along with a standard moderate degree of spatial smoothing (Gaussian 8-mm FWHM).

BOLD data analysis

Data analysis used the general linear model available on the Brain Voyager QX software package. The time course was normalized using *z*-transformation, controlling for differences

in signal-to-noise ratio across different brain regions. We conducted a random-effects GLM analysis on BOLD signal during the learning paradigm with separate stick-function predictors during the decision, anticipation, and feedback phases for each of the four trial types: human warmth, human competence, computer warmth, and computer competence. We also added predictors for motion correction to the model. The predictors were convolved with a standard canonical hemodynamic response function. We transformed structural and functional data of each participant to standard Talairach stereotaxic space (Talairach & Tournoux, 1988).

Main effect contrasts

We performed whole-brain contrasts for each of the main effects (*agent*: human > computer; *trait*: warmth > competence) during the decision, anticipation, and feedback phases. To correct for multiple comparisons, we used the cluster-level statistical threshold estimator plug-in in the BrainVoyager QX software (see Forman et al., 1995). This plug-in is designed to perform a Monte Carlo simulation to determine the number of contiguous voxels needed in order to correct for multiple comparisons. We set our initial statistical parametric maps (SPMs) to $p < .01$ and cluster corrected at a false positive rate of $p < .05$, using the recommended 1,000 iterations. The determined cluster sizes and significant regions are reported in Table 1.

Anatomically defined regions of interest

To examine responses in individual person perception brain regions and help us to understand the brain mechanisms of the processes of interest, we anatomically defined regions of interest (ROIs)—specifically, the anterior and posterior rostral MPFC (as defined in Amodio & Frith, 2006), bilateral STS, bilateral TPJ, and bilateral precuneus. The STS, TPJ, and precuneus were anatomically defined on both the right and left sides of the brain. Patterns of activation for right and left are reported separately in the article. We also anatomically defined the left and right striatum. This is a conservative approach to our analyses; anatomically defining large regions such as the MPFC increases the number of voxels examined, adding noise to the data from nonactive voxels. Therefore, any significant results obtained from these regions are agnostic to “sweet-spots” within the brain region.

Regressions

Using linear regression, we tested whether activity in person perception brain regions during the decision phase predicts activity in the striatum during feedback. We entered beta values from the person perception brain regions into a linear

Table 1 Main effect contrasts: Peak coordinates for two main effect contrasts at three different phases of the task (decision, anticipation, and feedback)

Contrast	Phase	Determined Cluster Threshold	Region	BA	Talairach Coordinates (x, y, z)	Voxels	t-Value	p-Value	Effect Size
Human > Computer	Decision	45	Interior Frontal Gyrus	9	55, 21, 26	1,404	3.85	.0008	.38
	Anticipation	60	Anterior Cingulate	24	1, 24, -1	2,690	4.52	.0001	.46
				Interior Frontal Gyrus	46	-47, 39, 11	8,171	5.13	.00003
	Feedback	52	Does not survive						
Warmth > Competence	Decision	49	Superior Frontal Gyrus	8	-14, 45, 44	1,741	-4.04	.0005	.40
	Anticipation	46	Middle Temporal Gyrus	39	-35, -72, 29	2,874	3.72	.001	.37
	Feedback	52	Posterior Cingulate	31	16, -51, 20	1,605	-5.10	.00003	.52
				Precentral Gyrus	6	-14, -9, 35	2,347	-3.97	.0006

Note. We acquired these regions using a cluster level statistical threshold estimator based on 1,000 iterations in a Monte Carlo simulation. Uncorrected SPMs were set at $p < .01$ and cluster corrected at a false positive rate of $p < .05$. We report regions surviving correction

regression as the predictor and used the bilateral striatal activity during feedback as the outcome measure. To avoid multicollinearity between our predictors, we performed separate regressions for each person perception brain region and condition and corrected for multiple comparisons using a Bonferroni correction.

We also tested for brain–behavior relationships—specifically, whether the learning models correlated with brain activity during feedback. Using parametric analysis, we created additional GLMs, following the procedure described above. We regressed each participant’s bias index, fictive error index, last-trial choice, and running average onto brain activity. Resulting maps were thresholded at $p < .005$ and cluster corrected using the Monte Carlo simulation described above.

Three versus one contrasts

We conducted planned 3 versus 1 contrasts for both our behavioral and imaging data. This analysis compares the data for a focal cell within the design with those for the other three cells. We hypothesized that the human warmth condition would produce less learning (higher bias and fictive error use; lower last-trial accuracy and running average) and greater activity in person perception brain regions than would the other three conditions. To test this, we performed a *t*-test on this focal cell and the average of the three remaining cells. We also tested the orthogonal contrasts, testing each condition as a focal cell to determine whether the human warmth condition was the only deviant cell. Significant results are reported in the article.

As a final descriptive analysis, we examined the effect sizes associated with the three versus one contrasts in the person perception brain regions. Using a data-driven approach, we specifically looked in the brain regions showing significant interaction or main effects in the ANOVA. These brain regions include the left STS, right TPJ, and posterior rostral medial

prefrontal cortex (PRMFC). We rank ordered the effect sizes for each of the conditions within a brain region to discern whether agent, trait, or the combination of the two has the biggest effect in the brain region.

Manipulation check

We collected an independent sample of participants ($N = 400$; 44 % female; age: 41 %, 18–29 years old; 31 %, 30–39 years old; 11 %, 40–49 years old; 10 %, 50–59 years old; 7 %, over 60 years old) using Amazon’s Mechanical Turk (MTurk) to test how participants may have interpreted the manipulations present in our 2×2 design (i.e., the meanings of the manipulations of warmth and competence as applied to human and computer agents). The study was advertised as a 5-min psychology survey that paid \$0.50 for completion.

We assigned participants to one of the eight conditions in our 2 (agent: human, computer) $\times 2$ (trait: warmth, competence) $\times 2$ (returned profit: high, low) between-subjects design. After providing informed consent, participants read a description of the manipulations used in the study regarding the type of agent and firm (trait manipulation). Participants then read a sentence telling them how much returned profit they received from the agent (manipulated to be a high or low amount). We then asked participants to make an attribution (answer whether the agent, stock market, circumstance, or a combination of the above had the most influence on the returned profit), to state whether they would reinvest with the agent, and to rate on a scale of 1 to 5 (*very unlikely* to *very likely*) whether they would expect to receive the same amount of profit if the agent was in a variety of situations (using the trait-situation rating form; Church, Katigbak, & del Prado, 2010).

We performed chi square tests on the attribution and reinvest questions, assuming equal frequencies, to test whether the different manipulations resulted in different attributions and behavior (see Table 2). We found that participants made more

Table 2 Chi square values obtained for attribution question assuming equal frequencies in each condition. All values are significant using $\alpha = .005$, $df = 3$, corrected for multiple comparisons

	Computer High	Computer Low	Human High	Human Low
Attribution χ^2 values				
Warmth	45.68	38.48	30.96	64.08
Competence	41.36	39.76	70.8	38.64

attributions to the agent than to the stock market, circumstance, or a combination of information during the warmth descriptions, regardless of the type of agent. Conversely, they made more attributions to a combination of the agent, the circumstance, and the stock market in the competence conditions. Participants also made fewer attributions to the circumstance across all conditions (see Fig. 2a). Participants were also more likely to reinvest in all of the conditions (see Table 3), except for the two low-warmth conditions, where they were not more likely to invest (see Fig. 2b).

We took an average of each of the situations in the trait-situation rating form to test the extent to which participants interpreted the manipulations as stable traits. Higher averages represent a higher dispositional (trait) attribution. The averages were entered into a $2 \times 2 \times 2$ between-subjects ANOVA. We observed a main effect of trait, $F(1, 392) = 7.87$, $p = .005$, partial $\eta^2 = .02$, $\Omega^2 = .80$, such that participants showed greater trait attributions to competence ($M = 3.42$, $SD = 0.30$) than to warmth ($M = 3.24$, $SD = 0.35$) information. There was no main effect difference for agent, and no interaction between agent and trait, $p > .05$. This suggests that regardless of the type of agents, participants' trait attributions were influenced by the traits themselves. Specifically, participants perceived competence as a more enduring trait, stable across situations, and less variable across situations than warmth.

Results

Learning behavior

We ran a 2 (agent) \times 2 (trait) ANOVA to examine whether our experimental conditions led to participants earning different amounts of money. There were no significant main effects or interactions (all $ps > .05$), suggesting that participants did not earn more money in any of the conditions. Therefore, differences in strategy per condition discussed below are not due to participants earning more money during a particular condition. Stated differently, participants did not gain an advantage (they were equally competent) at earning money playing in any one of our experimental conditions.

Bias index

We examined participants' agent selection to see whether the four conditions differentially hindered learning. Our bias index measures whether participants were more likely to select one agent over the other two, which is inversely related to participants' learning. We calculated this bias index for each of the four conditions and conducted a planned 3 versus 1 contrast to determine whether the human warmth cell deviated from the other three cells. We find that participants showed significantly more bias in the human warmth¹ condition, as compared with the other three cells, $t(44) = 2.33$, $p = .024$, partial $\eta^2 = .11$.

We also entered our bias index into a 2 (agent: human, computer) \times 2 (trait: warmth, competence) repeated measures ANOVA. We found a significant agent \times trait interaction, $F(1, 44) = 6.60$, $p = .014$, partial $\eta^2 = .13$, $\Omega^2 = .71$. To further probe the interaction, we performed follow-up post hoc simple effect tests. We found that warmth information specifically about humans ($M = 1.35$, $SD = 0.66$), but not computers ($M = 1.08$, $SD = 0.49$), leads to increased biased agent selection, $t(44) = 2.67$, $p = .011$, partial $\eta^2 = .14$ (see Fig. 3a).²

Fictive error

We next tested whether participants used fictive error information to guide their decisions. This measure is again inversely related to learning, since participants should learn that choosing the agent that was previously the highest lowered their chances of receiving the high reward on the current trial (see the Method section). We find similar results with the 3 versus 1 contrast; participants used fictive error information more often in the human warmth³ condition, as compared with the average of the other three conditions, $t(44) = 2.25$, $p = .03$, partial $\eta^2 = .10$.

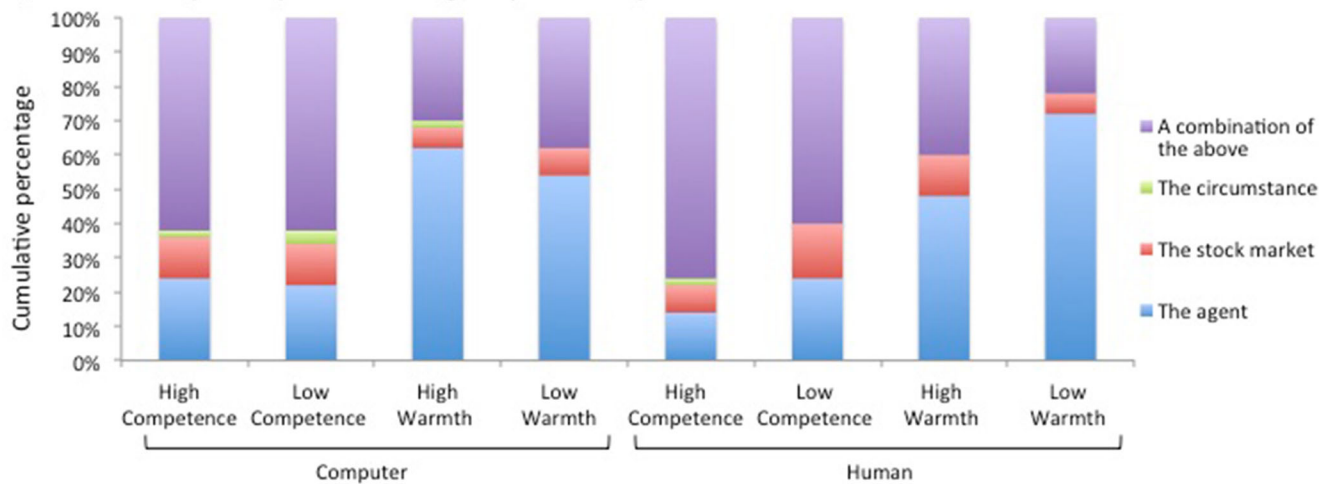
We also entered the fictive error measure into a 2 (agent) \times 2 (trait) repeated measures ANOVA, which revealed a significant main effect of agent where participants were more likely to use fictive error for human ($M = 0.25$, $SD = 0.13$) than for computer ($M = 0.23$, $SD = 0.13$) agents, $F(1, 44) = 4.43$,

¹ The computer warmth condition also deviates from the average of the other three conditions, resulting in less bias than in the other three conditions, $t(44) = 2.62$, $p = .012$, partial $\eta^2 = .14$.

² Participants were marginally more biased for the human agent when using warmth ($M = 1.35$, $SD = 0.66$), as compared with competence ($M = 1.25$, $SD = 0.62$), information, $t(44) = 1.79$, $p = .080$, partial $\eta^2 = .07$, but showed the opposite pattern for computer agents, with a greater bias for competence ($M = 1.19$, $SD = 0.48$) than for warmth ($M = 1.08$, $SD = 0.49$) information, $t(44) = -1.81$, $p = .077$, partial $\eta^2 = .07$.

³ The computer warmth condition also deviates from the average of the other three, resulting in less fictive error use than in the other three conditions, $t(44) = 2.21$, $p = .032$, partial $\eta^2 = .10$.

a) Percent of participants making responsibility attributions



b) Percent of participants reporting they would or would not reinvest

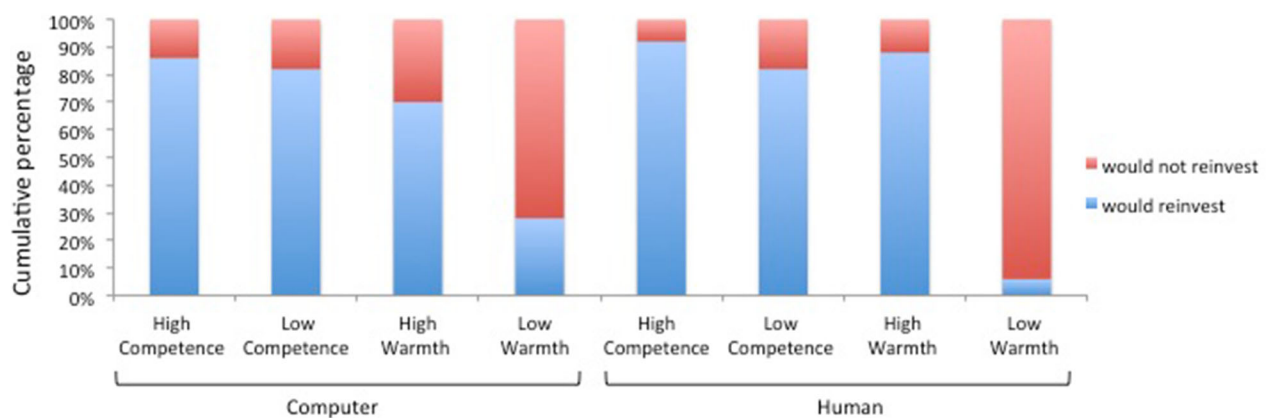


Fig. 2 Attributions for behavior. **a** Percentages of participants who attributed responsibility for investment behavior to the agent, the stock market, circumstance, or a combination of these three. **b** Percentages of participants who reported that they would or would not reinvest with the agent

$p = .041$, partial $\eta^2 = .09$, $\Omega^2 = .54$. To probe our hypotheses further, we tested the simple effects using post hoc t -tests. We again find that participants were more likely to use fictive error information for human ($M = 0.26$, $SD = 0.14$) than for computer ($M = 0.22$, $SD = 0.13$) agents within the warmth domain, $t(44) = 2.68$, $p = .01$, partial $\eta^2 = .14$ (see Fig. 3b).⁴

Last-trial accuracy

We turned next to looking at whether there was evidence that participants *did* learn the feedback structure of the game. Given that the last trial of each miniblock was deterministic, we examined whether participants were accurate on this last trial, choosing the agent that had yet to return the highest profit (see the Method section for details). Our 3 versus 1 contrast

⁴ The ANOVA also revealed a marginally significant agent \times trait interaction, $F(1, 44) = 3.23$, $p = .079$, partial $\eta^2 = .07$, $\Omega^2 = .42$.

did not reveal any deviant cells for this behavioral measure, all $ps > .05$.

We also entered participants' last-trial accuracy into a 2 (agent) \times 2 (trait) repeated measures ANOVA. We did not find any significant results.⁵

Running average

We also tested whether participants ignored trait inferences and interpreted feedback from the agents in a purely economic sense. We measured the extent to which participants calculated a running average of the amount of profit returned by each agent and chose the agent that had the highest average. Our 3

⁵ We did observe a marginally significant main effect of trait, with participants showing greater accuracy for the competence ($M = 0.39$, $SD = 0.14$) than for the warmth ($M = 0.34$, $SD = 0.12$) information, $F(1, 44) = 3.71$, $p = .061$, partial $\eta^2 = .078$, $\Omega^2 = .469$. These results suggest that participants may have learned the feedback structure better for competence than for warmth information.

Table 3 Chi square values obtained for reinvest question assuming equal frequencies in each condition. All values are significant using $\alpha = .005$, $df = 1$, corrected for multiple comparisons

	Computer High	Computer Low	Human High	Human Low
Reinvest χ^2 values				
Warmth	8.0	9.68	28.88	38.72
Competence	25.92	20.48	35.28	20.48

versus 1 contrast did not reveal any deviant cells for this behavioral measure, all $ps > .05$.⁶

We entered participants' highest average agent accuracy scores into a 2 (agent) \times 2 (trait) repeated measures ANOVA. We did not observe any significant results (all $ps > .05$), suggesting that participants were not using this strategy differentially for the different conditions.

Experimental questionnaires

We tested whether participants had any clear preference for trait and agent conditions before and after participating in the study. Frequencies are reported in Table 4. We performed a one-way χ^2 goodness-of-fit test (assuming equal frequencies) on pre- and postexperimental preferences separately. Our results indicate that participants' trait preferences were not equally distributed before participating, $\chi^2(2) = 13.33$, $p = .001$, or after participating, $\chi^2(2) = 8.40$, $p = .015$. Before participating, 56 % of participants preferred the trait competence condition, 33 % preferred the trait warmth condition, and 11 % were undecided. After participating, 47 % preferred competence, 40 % preferred warmth, and 13 % were undecided. Similarly, agent preferences were not equally distributed before participating, $\chi^2(2) = 19.60$, $p < .001$, or after participating, $\chi^2(2) = 15.60$, $p < .001$. Before participating, 62 % of participants preferred human agents, 29 % preferred computer agents, and 9 % were undecided. After participating, 60 % preferred human agents, 27 % preferred computer agents, and 13 % were undecided.

We performed a McNemar–Bowker test of symmetry (Bowker, 1948; McNemar, 1947) to test whether participants switched preferences after participating. We find that preferences stayed the same for the type of trait, $\chi^2(3) = 2.00$, $p > .05$, as well as type of agent, $\chi^2(3) = 2.80$, $p > .05$, suggesting that agent behavior did not affect participants' preferences for type of agent or type of firm.

⁶ However, we observed a marginally significant 3 versus 1 contrast in which participants showed higher reliance on the running average in the computer competence condition than in the other three conditions, $t(44) = 1.81$, $p = .076$, partial $\eta^2 = .07$.

Imaging analysis

Because we were interested in how brain regions implicated in person perception are modulated during social learning, we examined activation in anatomically defined person perception brain regions. Below, we describe the pattern of activation within each of these regions. All reported patterns of activation in the person perception brain regions are for the decision phase of the task. Additionally, we examined activation in the striatum during the feedback phase to compare our results with established findings in the neuroscience of decision making literature.

3 versus 1 contrasts

To test our hypothesis that the human warmth condition would increase activation in person perception brain regions, as compared with the average of the other three conditions, we performed 3 versus 1 contrasts for each of the ROIs. We find that within the left STS, the human warmth⁷ condition increased activation, as compared with the other three conditions, $t(24) = 2.17$, $p = .040$, partial $\eta^2 = .16$; however, we did not find significant effects of human warmth information in the other brain regions.⁸

ANOVAs

We also performed omnibus ANOVAs and simple effect tests for the individual brain regions. Significant results are reported below.

Superior temporal sulcus

We examined activation in the left and right STS independently. In the left STS, we observed a significant main effect of trait, $F(1, 24) = 4.32$, $p = .048$, partial $\eta^2 = .15$, $\Omega^2 = .51$, such that warmth ($M = 0.02$, $SD = 0.06$) engaged this region more than did competence ($M = -0.001$, $SD = 0.02$). This main effect is qualified by a significant agent \times trait interaction, $F(1, 24) = 5.00$, $p = .035$, partial $\eta^2 = .17$, $\Omega^2 = .57$. Probing this

⁷ The human competence condition was also deviant from the other cells, showing less activity for the human competence condition than for the other three conditions, $t(24) = 2.29$, $p = .031$, partial $\eta^2 = .18$.

⁸ We find similar but marginal results for the human warmth condition in the right TPJ, $t(24) = 1.95$, $p = .063$, partial $\eta^2 = .14$, and in the PRMFC, $t(24) = 2.04$, $p = .052$, partial $\eta^2 = .15$, both showing increased activation to the human warmth condition, as compared with the other three cells. In the right TPJ, the human competence cell showed significantly less activation than the other three cells, $t(24) = 2.57$, $p = .017$, partial $\eta^2 = .22$, and the computer warmth condition was marginally significant, again showing less activation than the other three cells, $t(24) = 1.75$, $p = .092$, partial $\eta^2 = .11$.

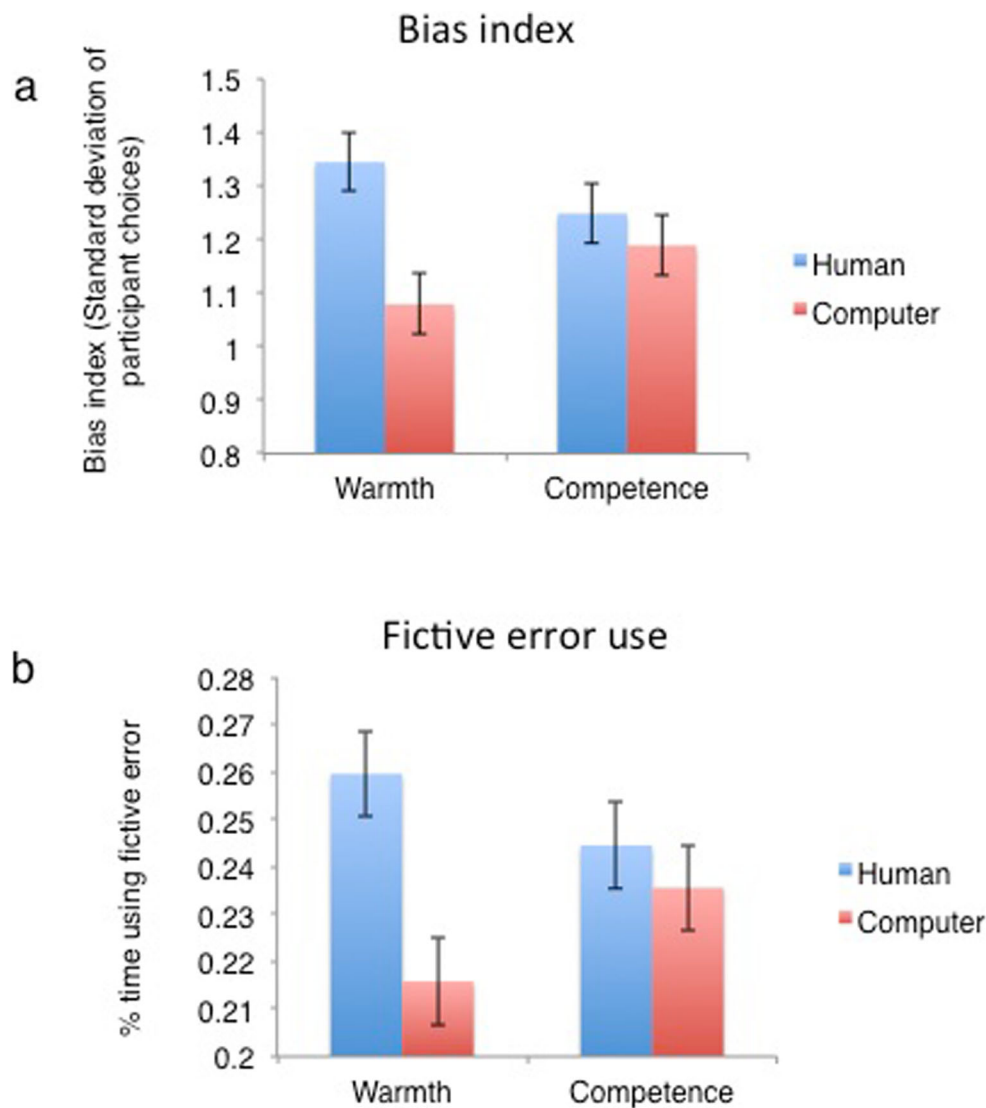


Fig. 3 Behavioral measures. **a** Bias index calculated from agent selection using standard deviation of participant choices. Higher bias index (scores on the y-axis) represents reduced learning. Error bars represent the

standard error of the mean. **b** Fictive error use. Higher percentages on the y-axis represent reduced learning. Error bars represent the standard error of the mean

interaction, we found that the human warmth condition ($M = 0.04$, $SD = 0.10$) engaged the left STS more than did the human competence condition ($M = -0.002$, $SD = 0.03$), $t(24) = 2.31$, $p = .030$, partial $\eta^2 = .18$, or the computer warmth condition ($M = -0.002$, $SD = 0.03$), $t(24) = 2.11$, $p = .045$, partial $\eta^2 = .16$ (see Fig. 4).

We found a similar pattern of activity in the right STS, although the main effects and interaction were not statistically significant (all $ps > .05$).

Temporal parietal junction

Similarly, we examined activation in the left and right TPJ. In the right TPJ, we observed a significant agent \times trait interaction, $F(1, 24) = 6.30$, $p = .019$, partial $\eta^2 = .21$, $\Omega^2 = .67$, in which the human warmth condition ($M = 0.03$, $SD = 0.10$)

again increases right TPJ activity, as compared with the human competence condition ($M = -0.01$, $SD = 0.03$), $t(24) = 2.23$, $p = .035$, partial $\eta^2 = .17$ (see Fig. 4).⁹

We observed a similar pattern of activity in the left TPJ, although the main effects and interaction were not statistically significant (all $ps > .05$).¹⁰

⁹ The human warmth condition is marginally different, as compared with the computer warmth condition ($M = -0.01$, $SD = 0.04$), $t(24) = 1.99$, $p = .058$, partial $\eta^2 = .14$

¹⁰ In the left TPJ, the interaction was marginally significant, $F(1, 24) = 3.16$, $p = .088$, partial $\eta^2 = .12$, $\Omega^2 = .40$. The simple effects of interest are also marginal, with the human warmth condition ($M = 0.03$, $SD = 0.11$) more active than the human competence condition ($M = -0.004$, $SD = 0.03$), $t(24) = 1.67$, $p = .108$, partial $\eta^2 = .10$, or computer warmth condition ($M = -0.01$, $SD = 0.03$), $t(24) = 1.76$, $p = .092$, partial $\eta^2 = .11$.

Table 4 Preference switching: number of participants (out of 45) who changed their reported preferences on pre- and post questionnaires

Trait preference switching

		Post questionnaire			Total
		Competence	Warmth	Undecided	
Pre questionnaire	Competence	18	4	3	25
	Warmth	1	12	2	15
	Undecided	2	2	1	5
	Total	21	18	6	45

Agent preference switching

		Post questionnaire			Total
		Human	Computer	Undecided	
Pre questionnaire	Human	22	2	4	28
	Computer	4	8	1	13
	Undecided	1	2	1	4
	Total	27	12	6	45

Note. The rows represent participants' preferences in the prequestionnaire before interacting with the agents. The columns represent participants' preference after interacting with the agents. Each cell represents the number of participants who switched from one preference (in the prequestionnaire) to another (in the postquestionnaire). Along the diagonals (shaded gray) are the number of participants who did not change their preferences from pre- to postquestionnaire. The bottom row represents postquestionnaire totals for each condition. The last column on the right represents prequestionnaire totals for each condition

Medial prefrontal cortex

Using the functional divisions created by Amodio and Frith (2006), we examined activation in two regions of the MPFC, as well as the combination of these two regions. While the posterior rostral MPFC is implicated in cognitive tasks, the anterior rostral MPFC is implicated in emotional tasks. In the posterior rostral MPFC, we observed a significant main effect of agent, $F(1, 24) = 4.44, p = .046$, partial $\eta^2 = .16, \Omega^2 = .53$, with greater activation to human ($M = 0.02, SD = 0.06$) than to computer ($M = -0.01, SD = 0.02$) agents (see Fig. 4).¹¹ The human warmth condition ($M = 0.04, SD = 0.10$) increased activity in the prMPFC, as compared with the computer warmth condition ($M = -0.01, SD = 0.04$), $t(24) = 2.22, p = .036$, partial $\eta^2 = .17$.

In the anterior rostral MPFC, we observed a similar pattern of activation, although the ANOVA results were not significant (all $ps > .05$).

¹¹ The agent \times trait interaction was marginally significant, $F(1,24) = 4.15, p = .053$, partial $\eta^2 = .15, \Omega^2 = .50$. The simple effect test for the human warmth ($M = 0.04, SD = 0.10$), as compared with the human competence ($M = 0.002, SD = 0.03$), condition was marginally significant, $t(24) = 1.88, p = .072$, partial $\eta^2 = .13$.

Precuneus

Although a pattern of activation similar to that in the other regions described above was displayed, we only observed trending main effects and interactions in the precuneus.

Striatum

We examined activation in the left and right striatum during feedback. ANOVA results and the 3 versus 1 contrasts were not significant for the left or right signal from the striatum (all $ps > .05$).

Effect sizes

As an exploratory analysis, we examined the effect sizes associated with each of the 3 versus 1 contrasts in the brain regions showing significant differences in the ANOVAs above (i.e., left STS, right TPJ, and PRMFC). Within each brain region, we rank ordered the conditions according to effect size to determine whether the specific brain region responds most to agent, trait, or a combination of the information. We take a descriptive approach to describing the results. In the left STS, the greatest effect sizes were for the human competence and human warmth conditions, suggestive of a main effect of agent (larger effect for human agents). The PRMFC shows a similar pattern, with greater effect sizes for human warmth and computer warmth, again suggestive of a main effect of trait information (larger effect sizes for warmth information). In the right TPJ, we observed the greatest effect size for the human competence condition, suggesting that the right TPJ combines both agent and trait information (see Fig. 5).

Regressions

We tested whether activation in person perception brain regions listed above (left STS, right TPJ, and PRMFC) during the decision phase would predict activity in the striatum during the feedback phase. To avoid multicollinearity between our predictors, we entered beta values from the each of the person perception brain regions into separate linear regression as the predictor and used the bilateral striatal activity during feedback as the outcome measure. We also applied a Bonferroni correction for multiple comparisons. We find that activity in the left STS during the decision phase *negatively* predicts activity in the striatum during feedback for the human competence condition, $\beta = -0.66, F(1, 23) = 10.73, p = .003$, adj. $R^2 = .29$. Participants who engaged this person perception to a lesser extent at decision showed greater striatal activity during feedback for this condition. Interestingly, these results suggest that in a context where the mental state of the agent is not as relevant (i.e., competence inferences about people), person perception brain regions are less engaged and predict the extent to which the striatum is more

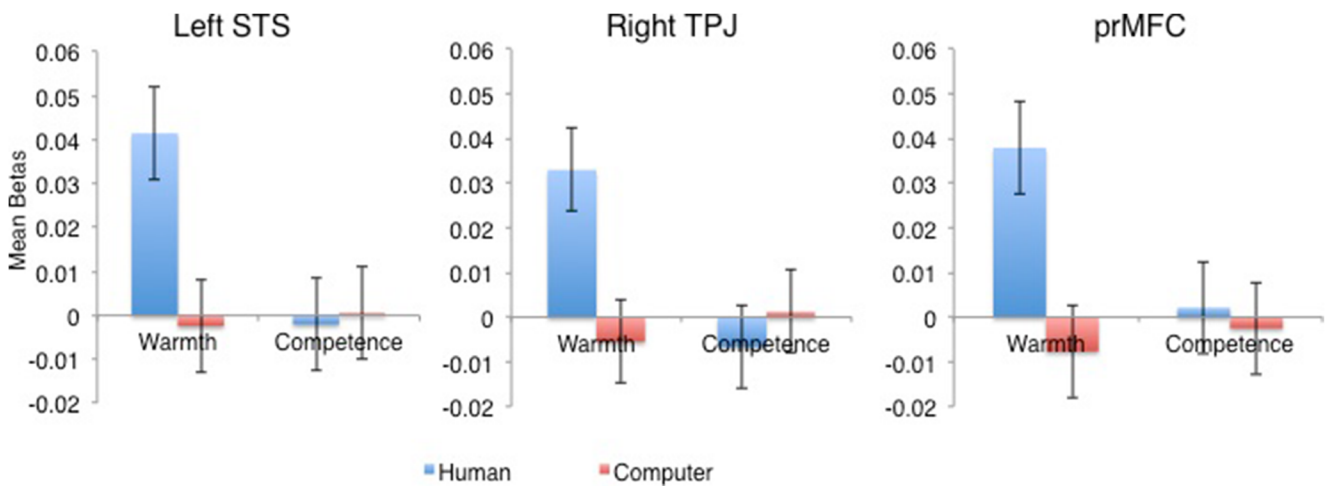


Fig. 4 Activation in person perception brain regions during the decision phase. Beta values extracted from regions of interest show a significant agent \times trait interaction, with increased activation for the human warmth condition. Error bars represent the standard error of the mean

active. We observed similar results in the right TPJ, although it does not survive correction, and did not observe significant results for the PRMFC.

Brain-behavior correlations

We tested whether our bias index and fictive error measures correlate with activity in any brain regions during feedback.

Significant regions are reported in Table 5. In particular, we find that during feedback, the bias index for human and computer agents is negatively correlated with activity in the striatum—specifically, the caudate [human, $x, y, z = -26, -42, 11, t(24) = -4.19, p = .0003$; computer, $x, y, z = 13, 27, 11, t(24) = -4.8, p = .00007$, and $x, y, z = -20, 15, 20, t(24) = -5.11, p = .00003$]. Similarly, the fictive error measure in the warmth domain is negatively correlated with activity in

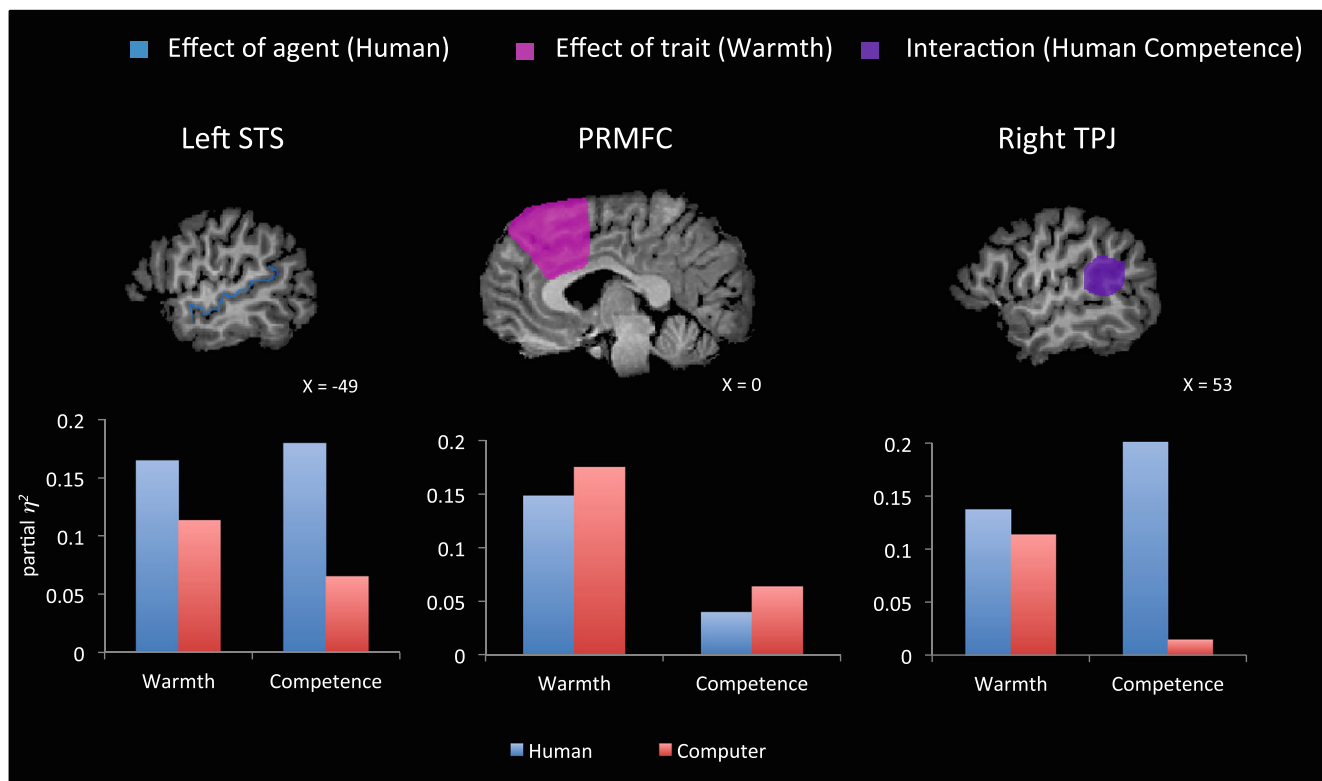


Fig. 5 Exploratory effect size analysis. Effect sizes associated with each of the 3 versus 1 contrasts in brain regions showing significant differences in the ANOVAs. The left STS responds to agent information, the PRMFC

responds to trait information, and the right TPJ responds to a combination of trait and agent information

Table 5 Brain–behavior correlations

Main effect (FB)	Region	Range (mm)	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>	<i>p</i>
Brain regions correlated with bias index during feedback							
Human	Caudate	3	−26	−42	11	−4.19	.00033
	Precuneus	5	−26	−69	26	−4.30	.00024
Computer	Caudate	5	13	27	11	−4.80	.00007
	Thalamus	4	−2	−6	20	−4.70	.00009
Warmth	Caudate	3	−20	15	20	−5.11	.00003
	Posterior cingulate	5	−29	−60	11	−5.36	.00002
Competence	Precentral gyrus	2	−23	−18	53	−5.87	.00001
	Precentral gyrus	0	61	6	32	−5.01	.00004
Brain regions correlated with fictive error during feedback							
Human	Middle temporal gyrus	6	−26	−57	26	−4.64	.00010
Computer	Cingulate gyrus	2	−5	−15	26	−5.30	.00002
Warmth	Caudate	3	13	12	20	−4.77	.00008
	Cingulate gyrus	4	1	−12	23	−4.65	.00010
Competence	Doesn't survive						

Note. Peak coordinates for brain regions correlated with bias index and fictive error during feedback. Uncorrected statistical parametric maps were set at $p < .005$ and cluster corrected at a false positive rate of $p < .05$, using a Monte Carlo stimulation with 1,000 iterations. We report regions surviving correction

the caudate [$x, y, z = 13, 12, 20; t(24) = -4.77, p = .00008$]. These results are consistent with the idea that our behavioral measures are assessing the extent to which learning is hindered.

Discussion

The present study investigated the role of the person perception brain network during social learning—specifically, the STS, TPJ, MPFC, and precuneus. Using a novel investment task, we showed that trait warmth information specifically about people, but not computers or competence information about either agent, impairs learning. Using two different behavioral measures, we showed that participants were more likely to be biased in their agent selection and were more likely to use fictive error information in the human warmth condition, suggesting that they did not learn the contingencies of the game. These results are consistent with previous work showing that warmth information impairs social decision making in a trust game (Delgado et al., 2005). It appears that participants were not simply relying on feedback to guide their decisions in the human warmth condition. Instead, perhaps they used trait inferences about people's warmth as a heuristic to guide their decisions. The brain data show that activity in person perception brain regions increases when decisions about human agents are made in the warmth domain, but not in the competence domain and not for computers.

Interestingly, the extent to which a participant engages person perception brain regions during the decision phase negatively predicts the extent to which the striatum is active

during feedback for conditions in which the mental states of the agent may be less relevant (human competence inferences). Therefore, it may be that in a more nonsocial context when person perception processes are less engaged, participants are more reliant on feedback processing in the striatum, and participants make better choices suggestive of learning. However, in a more social context such as the human warmth condition, participants may become more reliant on trait inferences about the agents to guide their decisions. This idea is consistent with previous research suggesting that activity in person perception brain regions hampers basic economic processing in social contexts (Harris, Lee, Capetany, & Cohen, 2014). This makes a case for a dual-systems approach in which social decision making within the warmth domain may rely on additional brain processing outside of the traditional decision-making network to guide decisions. When decisions are made in a social context, it would be important to engage processes associated with person perception to predict whether the agent has good or bad intentions. And given that the brain evolved in a social, not monetary, context, it follows that these processes have been given precedence when put in a social context.

Our exploratory effect size analysis highlights that different parts of the person perception network may be engaged for different types of social information. Consistent with previous literatures, the STS seems to be mainly concerned with identifying human agents (Allison et al., 2000; Zilbovicius et al., 2013). Similarly, the MPFC seems to be mainly concerned with warmth information, consistent with the idea that this region is involved in inferring mental states of others when in

a social interaction (van Overwalle, 2009). The right TPJ is the only region that appears to integrate both agent and trait information, with the largest effect size for the human competence condition. This is consistent with previous literature showing that the TPJ provides unique social information for agents who are both social and relevant for future behavior (Carter et al., 2012).

Social psychological theory suggests that human and computer agents differ in fundamental ways (Fiske & Taylor, 2013). An interested reader may therefore ask whether warmth information is relevant for computer agents. Our manipulation check suggests that despite the learning differences we observed in our study, there is no conscious report of computer warmth being very different from human warmth; participants made attributions to the agent at similar rates in the warmth conditions, regardless of type of agent. We observed similar results for competence information, with attributions to a combination of the agent, stock market, and circumstance. Furthermore, the trait-situation rating form suggests that the participants made trait inferences dependent not on the type of agent but, rather, on the type of social information (warmth or competence).

These results also suggest there is an important distinction to make when studying social decision making. We showed that specific types of social information affect learning and brain regions differently, consistent with previous works highlighting warmth and competence differences (Asch, 1946; Fiske et al., 2007; Wojciszke et al., 1998). While trait competence information produced similar learning rates for humans and computers, trait warmth information about humans impaired learning. Additionally, our results show that this effect is not due to the effect of trait warmth information in general but is limited to the perception of human agents. Interestingly, it is the combination of human–warmth that produces increased activation in person perception brain regions despite the fact that the human competence condition also presented human faces, suggesting that this network may be engaged only when the social context warrants it. This should serve as a cautionary tale to researchers interested in social decision making; all social contexts are not the same.

This study suggests that person perception brain regions may supply the additional computational processing necessary to make decisions in complex social environments—specifically, within the warmth domain. Although these processes may lead to less than optimal decisions in an economic context, they are generally useful in other social contexts, such as when meeting a person for the first time. Therefore, it is reasonable that these processes are recruited when forming impressions of the agents in this study. These impressions may serve as predictions or heuristics for future social interactions but may not be updated in the same way using feedback mechanisms.

Acknowledgments Funding for this project was provided by startup funds given to L.T.H. by Duke University. We would like to thank those who provided helpful comments on early versions of this project at the Social and Affective Neuroscience Society conference (2012), including Mauricio Delgado.

Competing Financial Interests The authors declare no competing financial interests.

References

- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267–278.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychological Bulletin*, 111, 256–274.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268–277.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Social Psychology*, 42, 258–290.
- Baron, S. G., Gobbini, M. I., Engell, A. D., & Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, 6, 572–582.
- Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 1558–1590.
- Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43(244), 572–574.
- Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, 6, 1–19.
- Carter, R. M., Bowling, D. L., Reeck, C., & Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science*, 337(6090), 109–111.
- Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, 17(7), 328–336.
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129(3), 564–583.
- Church, A. T., Katigbak, M. S., & del Prado, A. M. (2010). Cultural similarities and differences in perceived affordances of situations for Big Five behaviors. *Journal of Research in Personality*, 44(1), 78–90.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8(11), 1611–1618.
- Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., & Delgado, M. R. (2012). Social network modulation of reward-related signals. *The Journal of Neuroscience*, 32(26), 9045–9052.
- Fiske, S. T. (1998). Stereotypes, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *4th ed* (Handbook of Social Psychology, Vol. 2, pp. 357–411). New York: McGraw-Hill.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social perception: Warmth and competence. *Trends in Cognitive Science*, 11, 77–83.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878.

- Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture (2/e)*. London: Sage.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*, 636–647.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459–473.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77–83.
- Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.
- Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience*, *30*(2), 583–590.
- Harris, L. T., & Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. *Social Neuroscience*, *5*, 76–91.
- Harris, L. T., Lee, V. K., Capestany, B. H., Cohen, A. O. (2014). Assigning economic value to people results in dehumanization brain response. *Journal of Neuroscience, Psychology, and Economics*.
- Haxby, J. V., Gobbini, M. I., & Montgomery, K. (2004). Spatial and temporal distribution of face and object representations in the human brain. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 889–904). Cambridge, MA: MIT Press.
- Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the timecourse of social perception: the effect of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*, 1267–1280.
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616–626.
- Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*(2), 284–294.
- Izuma, K., Saito, D. N., & Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience*, *22*(4), 621–631.
- Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., & Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *The Journal of Neuroscience*, *31*(37), 13039–13045.
- 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*(5718), 78–83.
- Lee, D., McGreevy, B. P., & Barraclough, D. J. (2005). Learning and decision making in monkeys during a rock–paper–scissors game. *Cognitive Brain Research*, *25*(2), 416–430.
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*, 1250–1252.
- Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*(3), 274–281.
- 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*, *104*(22), 9493–9498.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–346.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157.
- Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., & Dalgleish, T. (2009). A key role for similarity in vicarious reward. *Science*, *324*(5929), 900.
- Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, *36*(2), 265–284.
- Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001a). Separating processes within a trial in event-related functional MRI I: The method. *NeuroImage*, *13*, 210–217.
- Ollinger, J. M., Corbetta, M., & Shulman, G. L. (2001b). Separating processes within a trial in event-related functional MRI II: Analysis. *NeuroImage*, *13*, 218–229.
- Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences*, *107*(29), 13099–13104.
- Ponkanen, L. M., Alhoniemi, A., Leppanen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, *6*, 486–494.
- Reeder, G. D. (1979). Context effects for attributions of ability. *Personality and Social Psychology Bulletin*, *5*, 65–68.
- Reeder, G. D. (1993). Trait-behavior relations and dispositional inference. *Personality and Social Psychology Bulletin*, *19*, 586–593.
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*, 61–79.
- Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*, 736–745.
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294.
- Sanfey, A. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*, 598–602.
- Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, *12*, 508–514.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.
- Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500.
- Skowronski, J. J. (2002). Honesty and intelligence judgments of individuals and groups: The effects of entity-related behavior diagnosticity and implicit theories. *Social Cognition*, *20*(2), 136–169.
- Smith, D. V., Hayden, B. Y., Truong, T. K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of Neuroscience*, *30*(7), 2490–2495.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: an approach to cerebral imaging*. New York: Thieme Medical publishers.
- Tausch, N., Kenworthy, J. B., & Hewstone, M. (2007). The confirmability and disconfirmability of trait concepts revisited: Does content matter? *Journal of Personality and Social Psychology*, *92*(3), 542.
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *29*, 549–562.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*, 1124–1131.
- Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, *59*, S251–S278.

- Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–280.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*, 829–858.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mental state after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263.
- Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405.
- Zilbovicius, M., Saitovitch, A., Popa, T., Rechtman, E., Diamandis, L., Chabane, N., & Boddaert, N. (2013). Autism, social cognition and superior temporal sulcus.