INTRODUCTION

How do we make decisions? How do we judge what is right or wrong and how does this judgment translate to behavior? Over the last decade, research on the human brain has begun to shed light on such questions. Those research efforts build on a strong foundation of animal research responsible for the delineation of neural circuitry involved in processing information about rewards and punishments. Animal research also provided for the development of an understanding of how such circuitry operates during simple decision-making, such as pressing a lever to receive a reward. Advances in technology, chiefly the advent of neuroimaging techniques such as functional magnetic resonance imaging (fMRI), have allowed researchers to investigate similar matters regarding the operation of the human brain. Furthermore, research using these techniques may be extended in new directions to address questions not easily explored in animals, such as those involving the more complex decisions that occur in human society, (e.g., trusting an individual during a business transaction). These research efforts have introduced interdisciplinary collaborations and considerations, ranging from philosophy to economics, into the field of cognitive neuroscience. The contributions of these diverse fields do much to shape current thinking on human decision-making. In this paper, we will discuss how social information can modulate traditional ways of thinking about rational and economic decision-making specifically by affecting the neural systems of reward. First, we present an overview of the
neural circuitry underlying human reward systems. Next we present a description of an experiment where social impressions affect trust judgments and decision-making. Finally, we address potential implications of the experimental findings to the legal field and discuss the potential of future interdisciplinary collaborations across law and neuroscience.

THE REWARD SYSTEM OF THE BRAIN

A “reward” may be operationally defined as any stimulus with desirable properties that can drive behavior. Based on this definition, it is postulated that the purpose of rewards in the environment is to (1) induce hedonic feelings that (2) encourage exploratory behavior and (3) shape learning to ensure exploitation of previously rewarded behaviors.\(^1\) For example, a laboratory rat placed in an operant conditioning chamber (e.g., a “Skinner box” containing a lever that releases food pellets) will be driven to explore the environment and press an available lever because of the potential for a reward.\(^2\) After gaining a food pellet, the rat learns to associate behavior and reward (action-outcome), resulting in an increase in the frequency or intensity of lever-pressing.\(^3\) In 1954, Olds and Milner conducted “self-stimulation” experiments where, instead of a food reward, animals were electrically stimulated in the medial forebrain bundle (a group of neural fibers containing dopamine that connect midbrain dopaminergic centers such as the ventral tegmental area with forebrain structures such as the nucleus accumbens) during the pressing of a lever.\(^4\) Those authors found that the reinforcing properties of the electrical stimulation led the animals to increase their response frequencies.\(^5\)

Such self-stimulation studies, in addition to various


\(^3\) Id. at 279-80.


\(^5\) Id. at 423-25.
pharmacological experiments, lent credence to the “dopamine hypothesis of reward,” which postulated that reinforcing effects in the brain of various stimuli (e.g., addictive drugs) were caused by the release of the neurotransmitter dopamine in the brain. Electrophysiological recordings of neuronal activity of these dopaminergic cells, helped refine the relationship between dopamine and reward. While recording in the non-human primate brain, Wolfram Schultz and colleagues made the observation that dopaminergic cells in the midbrain, specifically in two distinct nuclei called the substantia nigra and the ventral tegmental area, were active when an unexpected reward was delivered. In the experiment, these cells displayed bursts of activity once a monkey received an unexpected reward, such as a drop of juice. However, once a light cue predicted the delivery of the juice (an instance of classical conditioning), the dopaminergic neurons no longer displayed the burst of activity at the time of reward. Instead, after learning of the association, bursts of activity by the dopaminergic cells were observed at the time of the light cue—in other words, the earliest predictor of the reward. Finally, if an expected reward failed to occur (e.g., if juice was not delivered following presentation of the light cue), then a depression in the activity of the dopaminergic cells was observed, signaling a prediction error. Dopaminergic neurons are therefore thought to aid in reward-related learning by providing a prediction error that can adjust expectations and guide behavior.

Some of the primary targets of dopaminergic neurons include prefrontal cortical regions and the striatum, a structure located deep in the brain below the cortex and known for its heterogeneity in connectivity and functionality. For example, research has implicated the

7. Id.
8. Id.
9. Id.
10. Id.
11. See P. Read Montague & Gregory S. Berns, Neural Economics and the Biological Substrates of Valuation, 36 NEURON 265, 265–72 (2002); Schultz, supra note 1, at 243–44.
striatum in motor, cognitive, and motivational processes, suggesting that the striatum may be a potential venue for the integration of movement and motivational information. While most of this research was conducted in animals, recent investigations of striatal function have extended to the human brain. Neuropsychological and neuroimaging techniques represent two primary methods used to probe the human striatum. Neuropsychological research programs take advantage of pharmacological or anatomical lesions of brain regions and allow for an investigation of the necessity of the structure for proper performance to occur. For instance, patients suffering from Parkinson’s disease display striatal dysfunction due to a deterioration of the dopaminergic projection to the striatum. This diminished dopaminergic input into the striatum leads to well-characterized motor deficits such as prominent tremors, as well as to problems in cognitive processes such as learning from feedback in typical trial and error tasks. Such behavioral results in humans mirror the electrophysiological data from animals, which suggest that dopamine processes a prediction error signal that impacts learning.

The other methodology commonly used to investigate the functionality of the human striatum (as well as other brain regions) is neuroimaging, including both positron emission tomography (PET) and fMRI. Recent studies using PET, which allows imaging of neurotransmitter levels in the brain during cognitive tasks, have demonstrated that dopamine is released in the striatum during highly arousing situations that elicit motivation,

---

13 See, e.g., Daphna Shohamy et al., Role of the Basal Ganglia in Category Learning: How Do Patients with Parkinson’s Disease Learn?, 118 BEHAV. NEUROSCIENCE 676, 676 (2004); Nora Volkow et al., “Nonhedonic” Food Motivation in Humans Involves Dopamine in the Dorsal Striatum and Methylphenidate Amplifies this Effect, 44 SYNAPSE 175, 176 (2002).
14 See, e.g., Shohamy et al., supra note 13, at 676.
15 Id. at 683.
16 Id.
17 Id. at 682.
18 Schultz et al., supra note 6, at 1594.
such as food delivery when one is hungry\(^1\) or while playing a video game for monetary rewards.\(^2\) Another technique to study the human brain is fMRI, which takes advantage of the magnetic properties of blood to allow experimenters to measure the brain’s activity indirectly through an assessment of the hemodynamic responses, also known as the blood-oxygenated level dependent (BOLD) response.\(^3\) Briefly, fMRI builds on the idea that changes in blood flow in a specific brain region correlate with neural activity in that same region,\(^4\) although a direct correlation is still under investigation.\(^5\)

Using fMRI, investigators have been able to extend findings in animals to humans using both primary rewards (e.g., juice)\(^6\) and secondary rewards (e.g., money).\(^7\) For example, increases in oxygen demand, and thus brain activity, have been observed in the human striatum while subjects are anticipating a potential juice reward\(^8\) or a potential monetary reward.\(^9\) The striatum has also been associated with coding the differential response between a positive and negative consequence of an action, that is, whether the action led to a reward or a punishment.\(^10\) Those data suggest that the striatum, specifically the dorsal region of the striatum called the caudate nucleus,
rather than processing the reward per se, may be involved in learning associations between behaviors and potential rewards.\textsuperscript{29} The idea that the human striatum is important for reward-related learning is consistent with the findings from animal studies. In fact, the prediction error signal communicated by dopaminergic neurons\textsuperscript{30} has been observed in the human striatum in fMRI paradigms,\textsuperscript{31} further linking the human striatum with learning the value of stimuli or actions that predict rewards.\textsuperscript{32}

THE SOCIAL MIND: HOW SOCIAL FACTORS MAY MODULATE NEURAL SYSTEMS OF REWARD

While early neuroimaging studies confirmed animal studies and extended those concepts into the human brain, research has now started to focus on the various ways in which social factors can contribute to reward processing in humans. Suffice it to say that this literature is beyond the scope of the present paper. A more comprehensive review on both the evolution of neuroeconomics\textsuperscript{33} and social neuroscience\textsuperscript{34} can be found elsewhere. The focus of this paper is the transition from simple processes performed by the striatum, such as learning that an action leads to a reward, to more complex processes observed in typical human society, such as learning that an individual predicts a potential reward during a business or legal proceeding.

Social stimuli are known to engage the brain’s reward system, be they beautiful faces,\textsuperscript{35} money\textsuperscript{36} or status

\textsuperscript{29} Mauricio Delgado et al., An fMRI Study of Reward-Related Probability Learning, 24 NEUROIMAGE 862, 862 (2005).
\textsuperscript{30} Schultz et al., supra note 6, at 1594.
\textsuperscript{32} Montague & Berns, supra note 11, at 265.
\textsuperscript{35} Itzhak Aharon et al., Beautiful Faces Have Variable Reward Value: fMRI and Behavioral Evidence, 32 NEURON 537, 537 (2001).
\textsuperscript{36} John O’Doherty et al., Abstract Reward and Punishment
symbols such as extravagant sports cars. Notably, modulation of the activity of brain regions such as the striatum is observed during social interaction. For instance, activation of the striatum is particularly prominent during cooperation of two individuals during the so-called prisoner’s dilemma game, where the two people interact and can either cooperate or defect toward a reward that varies in size according to their respective choices. Striatum activation also increases when individuals are presented with the faces of previous cooperators and even when exacting revenge on defectors, an activity that was interpreted by the authors as a rewarding feeling resulting from the punishment of perceived unfairness.

One interesting experiment of social interaction involved a game known in economics as the “trust game.” In a typical trust game, an investor is faced with a choice of how much money to transfer to another player, the trustee. The transferred money gets tripled and the trustee can either defect from the interaction and keep the investment or he can send back some of the money to the investor, thus ensuring a profitable transaction for both players. In multi-round exchanges, a reputation for players is built, thus each move has to be considered carefully. In an elegantly designed experiment, King-Casas and colleagues found that
reciprocity inspires learning during transactions, which leads to formation of reputations.\textsuperscript{47} The authors found that an intention to trust signal was being computed in the striatum.\textsuperscript{48} That is, subjects learned that a player was trustworthy and that was reflected as an investment in the next trial. Interestingly, the intention to trust signal shifted in time as learning progressed, similar to the temporal prediction error signal exhibited by dopamine, further suggesting that the striatum is involved in learning about reward-related stimuli in a social setting.\textsuperscript{49}

Thus, research suggests that social factors map cleanly onto the existing knowledge of basic reward circuits. It is unclear, however, how information such as moral beliefs or social perceptions influence decision-making and the neural circuitry of reward processing. It has been observed, for example, that people are willing to forgo part of wages if they believe that their employer’s mission is praiseworthy.\textsuperscript{50} This type of behavior cannot be accounted for by rational theory,\textsuperscript{51} and other variations (e.g., choosing to drive to a more distant department store because one does not like the policies of the nearest store) have been replicated elsewhere.

In order to study this phenomenon, we conducted a study where participants were instructed they would play a variation of the trust game with three different partners portrayed as having different levels of moral aptitude.\textsuperscript{52} In our variation of the trust game, participants were given a dollar and told that they could either keep (i.e., defect) or share (i.e., invest) the money. If they chose to share it with a partner, the money would be tripled, and the participant would then receive feedback on whether the

\textsuperscript{47} Brooks King-Casas et al., \textit{Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange}, 308 \textit{Science} 78, 82 (2005).
\textsuperscript{48} Id. at 81.
\textsuperscript{49} Id. at 82.
\textsuperscript{50} ROBERT FRANK, \textit{WHAT PRICE THE MORAL HIGH GROUND? ETHICAL DILEMMAS IN COMPETITIVE ENVIRONMENTS} 76-84 (2004).
\textsuperscript{51} Id. at 77 (“From the perspective of strict economic theory . . . the price one actually pays should not depend . . . on the prices one is able to pay.”); see also ROBERT FRANK, \textit{PASIONS WITHIN REASON: THE STRATEGIC ROLE OF THE EMOTIONS} 178 (1988).
\textsuperscript{52} See Mauricio Delgado et al., \textit{Perceptions of Moral Character Modulate the Neural Systems of Reward During the Trust Game}, 8 \textit{Nature Neuroscience} 1611, 1611 (2005).
partner shared back and split the profits (positive feedback) or defected (negative feedback). The subjects were instructed they would play with three fictional partners twenty-four times each. The partners were portrayed as having a certain personality that may or may not match the actual behavior displayed by the partner. Participants saw a computer screen with the face and name of a partner and were given the decision to keep or share money. After registering their choice, participants were then presented with the feedback from the partner (positive or negative) before a new trial would commence.

After being allowed practice trials to facilitate understanding of the rules and operation of the trust game, participants were given three bios that included a photograph (counterbalanced across the study), a name, and a blurb that described the partner's moral aptitude, as well as a recent newspaper article detailing an event in the partner's life. For example, one partner was described as a volunteer who had recently saved a woman from a club fire, which suggested that this partner was morally praiseworthy (“good” partner). Another partner was a business school graduate who attempted to sell heat-insulating tiles from the space shuttle Columbia on internet auction sites (“bad” partner). The third partner was involved in a similarly arousing story (e.g., supposed to be in a plane that crashed but he missed the flight), although it contained no information to form biases regarding expected moral behavior (“neutral” partner). The bios, created by economist Robert Frank, were extremely effective at creating social expectation and irrational impressions of each partner's behavior. However, despite their apparently disparate moral aptitudes, the partners all played with the same reinforcement schedule (50%). That is, they all shared or kept game money at the same frequency. Thus, based on outcomes alone, participants should have learned over time to adjust their expectations for the fictional partner's moral behavior and adapt decision-making appropriately.

A manipulation check (e.g., questionnaire asking how trustworthy a partner was perceived to be) showed that subjects learned at some level that all three partners were
essentially equivalent in their behavior. Specifically, during a pre-experimental session, the subjects rated the partners on moral aptitude differentially and according to the fictional bios. However, following the twenty-four trials with the partners, those differential ratings were abolished. Nevertheless, during game play the participants were still more trusting of the “good” partner, making more share decisions with the “good” rather than the “bad” partner. This observation was maintained across the entire game and was still evident in the trials as the game was nearing an end. The behavioral result suggests that social perceptions are strong modulators of behavior and can significantly influence economic decision-making. Using concurrently performed fMRI, we were also able to pinpoint the biological mechanisms that led to this behavior.

As discussed above, the striatum plays an important role in reward-related learning. Thus, it is no surprise that striatum activation was observed when participants were receiving feedback from partners; such feedback was being used to guide future decision-making. When participants were interacting with the neutral partner, for instance, increases in BOLD signal were observed in the striatum following positive feedback, while a decrease was observed following negative feedback. This pattern mimics the previously characterized signal in the striatum that differentiates between rewards and punishments, suggesting that the human striatum is valuating the current feedback to guide future decision-making. In the case of the neutral partner, there was no information that could bias behavior. Participants thus had to learn through trial and error what the outcome of interacting with the neutral partner would be. As a result, participants shared and kept about 50% with the neutral partner, as one would expect based on random sampling.

During trials in which subjects had been provided the fictional biographies, however, the differential signal in the striatum was not observed. This suggests that the brain’s trial and error learning system may have been inhibited during the game by the availability of prior social information. Participants may have bypassed the current feedback (e.g., good partner did not share with me on this trial) due to the overwhelming prior social information.
Consequently, subjects did not update their decision-making, preferring instead to conform to their original biases created by irrational social expectations. Thus, perceptions of moral character can influence the neural systems of reward and learning by creating social expectations that are more difficult to update. This finding raises the following important consideration for studies in the social domain that involve interactions: humans have biases, and the strength of those biases can modulate how we make decisions. Future designs must take such issues into account.

**IMPLICATIONS FOR THE LAW**

A fundamental consideration in many, if not all, legal undertakings is the assessment of the credibility of evidence. In many circumstances, evidence is presented to a judge or jury through a witness to aid in the explanation. The behavioral results reported here parallel a well-known phenomenon—namely, that the perception of the personal attributes of a witness directly impacts their credibility. As the presentation of a series of events or facts fundamentally shapes the case before a judge or jury, the credibility of the witness used to communicate those details is central to the task of the trial attorney. While the experiments reported here were conducted in a controlled laboratory setting, they may nonetheless remind legal professionals of the significance of moral perception in legal proceedings.

The present experiments emphasize the importance of establishing the credibility of the witness early in their presentation to a court. The previously published results demonstrate that the pre-established perception of the moral aptitude of a game character influences the behavior of human participants over an extended period of time. Indeed, if the game character is perceived as trustworthy, then the human subjects continued to trust the character in the form of investments (i.e., “share” decisions) throughout the course of the game. Extending this observation to the court room, once a perception of a witness is established with a jury or judge, it subsequently impacts all later interpretations of that witness’s

\[54\]. Id. at 1611.
testimony and behavior. As is well known by trial attorneys, the cultivation of a witness’s image early in a proceeding as a trustworthy, moral, and upright individual may reap rewards throughout the trial. Indeed, the moral aptitude of a witness is often central to a criminal trial where the actions of the accused are being evaluated for criminal intent. Similarly, attorneys should not neglect the cultivation of the image of scientific or technical experts in civil trials, such as patent infringement cases.

While the presentation of a witness provides the attorney with the opportunity to shape perceptions in a courtroom, the individual members of a jury also arrive at court with preconceived notions. The data presented here emphasize that such preconceptions may have a lingering effect on the later behavior of the jury. A voir dire examination provides the attorney with the opportunity to assess those biases and to plan for trial accordingly. A voir dire examination usually refers to the examination by the court or by attorneys of prospective jurors to determine their qualifications for jury service and whether cause exists to excuse particular jurors, as well as to provide information about the jurors to the attorneys. When performed before a trial, a carefully crafted voir dire examination allows the attorney to determine what perceptions the potential member of a jury possesses before presenting any portion of the case. By recognizing that the preconceived notions have a long-term impact on the actions and perceptions of the potential jurors, an attorney may well invoke his right to excuse jurors whose preconceived notions would directly impact the heart of the case to be presented. By excluding such individuals from the jury pool, the eventual reception of the attorney’s case would thereby be improved.

FINAL THOUGHTS AND FUTURE DIRECTIONS WITH RESPECT TO THE LAW

With certain variations, the experiments reported here may be used in the future to help establish strategies for the presentation and cross-examination of witnesses. In the results presented here, the human subjects behaviorally maintained their trust throughout the entire
course of the experiments. Future experiments could investigate what behavior on the part of the fictional character would be required to best establish this trust and credibility. The fictional characters used in the experimental paradigms could even be crafted to mimic the actual participants in a trial, including their actual physical, mental, and personality traits. Subsequently, attorneys could experiment with different ways of presenting the witness to see if the various approaches would impact the credibility and moral aptitude of the witness as perceived by experimental subjects acting as a mock jury.

Additionally, the fictional character may be fashioned after one of the witnesses to be used by opposing counsel. Attorneys could then employ a variety of approaches in the context of the trust game to investigate how most effectively to disrupt the credibility of the witness. The attorney might then draft a cross-examination of that witness according to the most effective approach. By performing these preliminary experiments in the controlled (and relatively inexpensive) setting of a laboratory, the attorney would gain a wealth of information about potential jury perceptions and thus be able to prepare for a mock trial or actual trial more effectively.