NEURAL AND BEHAVIORAL INVESTIGATIONS OF SOCIAL REWARD PROCESSING

Thesis by

Alice Lin

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2012

(Defended March 26, 2012)

ABSTRACT

Neural and Behavioral Investigations of Social Reward Processing

Alice Lin

Despite an extensive literature on the neural substrates of reward, relatively little is known about how social interactions modify decision-making. Here I present three experiments that examine the neural basis of social reward processing both in neurotypicals and individuals with autism spectrum disorder (ASD), a neuropsychiatric syndrome associated with social cognition impairments. Using functional magnetic resonance imaging (fMRI), I recorded brain activity during a probabilistic reward learning task with either social (smiling/frowning faces) or monetary (gaining/losing money) rewards. I found substantial overlap in the neural circuitry associated with social and non-social reward processing, suggesting that social rewards are processed similarly to other types of rewards. In contrast, individuals with ASD showed behavioral impairments in social reward processing, both in probabilistic reward learning and in an ecologically valid charitable donation task. Exploratory neuroimaging in ASD showed hypoactivation of key reward areas during decision-making. Taken together, these findings support the idea of a "common neural currency" in decision-making but also suggest the construction of accurate social reward value signals relies on recruitment of additional regions known to process social information.

Thesis Advisors: Prof. Antonio Rangel & Prof. Ralph Adolphs

dedicated to Jessica Herman ('89 - '08),

my little sister next door

ACKNOWLEDGEMENTS

I must first thank my graduate advisors, Prof. Antonio Rangel and Prof. Ralph Adolphs, for leading me into a fascinating area of research at the intersection of two exciting fields. Their appreciation for beautiful data and insight into the right questions to ask has indelibly shaped this thesis and inspired me continually to be a better scientist. I have always told people how lucky I was to not only find one, but two amazing advisors. I'm grateful for their great mentorship and guidance --- even outside of the lab. I now know the difference between Star Wars and Star Trek thanks to Antonio and have a slightly greater appreciation for hoppy beers thanks to Ralph. I also thank my committee members, Prof. Shinsuke Shimojo, Prof. John O'Doherty, and Prof. Peter Bossaerts, for their support, advice, and availability.

From the time I arrived in the Rangel and Adolphs labs until now, I have greatly benefited from the generosity of others. I would like to thank especially: Todd Hare, my first officemate who taught me everything I know about SPM and still entertains my questions even 6000 miles away; Cendri Hutcherson for her advice and insight on everything big and small; Lynn Paul and Dan Kennedy for their willingness to answer every question I had about autism diagnostics; Catherine Holcomb, without whom I wouldn't have been able to complete any of the work done with the ASD population; Ian Krajbich for always introducing the right amount of levity, and who was for a long time the other Caltech-lifer; my SURF students Chris Li and Karin Tsai; and my officemates Shabnam Hakimi and Vanessa Janowski, who are a constant source of encouragement.

On a personal level, I'd like to thank Joanna Hsu, Kai Shen, Nikki Sullivan, Stella Hu, Lea and Hubert Shen, Michael Chang, Isaac Gremmer, and my MBCLA and PCC small groups, who were my faithful cheerleaders along this marathon. Appreciation also goes out to my little brother, Edward, who really isn't so little any more, and whom I increasingly look up to more and more. Last of all, I want to thank my dad, who always thought I should get a PhD instead of "making pennies" in the real world, and my mom, who always just wanted me to go outside and play more.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

Chapter 1: **Introduction**

Facebook, a social networking site, is preparing for a $100 billion IPO this year; it will be the biggest of any tech company in history. Alongside Facebook is a growing cadre of social-media sites and mobile apps helping us to connect to one another. What has made companies like Facebook, Vineloop, and Simple Energy successful is their ability to tap into our interest in other people --- social stimuli. None of these companies need to pay us to click a button to find out whether the guy who lived next door to us in college has gotten engaged. We perform tasks like these because we are intrinsically motivated by social rewards, social stimuli and information that are rewarding.

Some have argued that we are evolutionary pre-disposed to be social creatures. And one reason for our big brains is to accommodate our social nature (Allman, 1999). The work of Reader and Laland more recently has added evidence to this idea by suggesting the expansion of the primate cortex relates to our greater capacity for social learning (Reader & Laland, 2002).

Researchers have been intrigued by social stimuli for some time. Darwin, while crafting his theory on evolution and natural selection, was at the same time conducting experiments to understand whether the expression of emotion is innate. His experiments were simple. He showed subjects a series of photographs of human faces, some with muscles artificially contracted by electric probes, and asked his subjects what emotion they thought the photographs conveyed (Darwin, 1872). Almost a hundred years later, Ekman concluded that facial expressions of emotions are not culturally determined, but universal across human cultures and thus biological in origin (Ekman & Sorenson, 1969).

Probably the best evidence that the foundations of face processing and our interaction with social stimuli are hardwired is the fact that within minutes after birth, newborns orient towards face-like stimuli, despite having no prior relevant visual experiences (Goren & Sarty, 1975). Clearly some aspects of our social nature are grounded in our biology, so a natural place to turn for study is inside the brain.

There is now a large body of work investigating the neural basis of processing socially relevant stimuli such as faces (Haxby, Horwitz, & Ungerleider, 1994; N Kanwisher, 2006; Tsao, Freiwald, & Tootell, 2006) and body positions (Desimone, Albright, & Gross, 1984; Peelen & Downing, 2007). More recently, economists have begun studying the neural basis of more complex constructs in the social domain, like altruism and trust, with the dictator and ultimatum game respectively. The dictator game is a simple game played between two people. One player, "the dictator", is given an amount of money (e.g., $10), and decides how much he wants to share with the second player. The second player has no input to the decision and must accept whatever the dictator proposes. Basic economics would predict in this scenario that the dictator would keep the full amount, as there are no repercussions. Experimental evidence, however, defies these predictions: subjects around the world give on average much more than $0 (Henrich et al., 2005). Computations are no longer straightforward when we introduce other people into the situation.

A rich history of social psychology and behavioral economics studies have repeatedly shown that social rewards are special and can cause people to act in "irrational" ways. However, few have investigated the basic reward properties and why.

This thesis studies some very basic questions about social rewards:

1. What is a social reward?
2. What makes them rewarding?
3. How do they compare to other types of rewards?

Along the way, we also explore what it informs us about autism spectrum disorder. To address these questions, the thesis uses a combination of neuroimaging and behavioral methods.

In Chapter 2, I provide a working definition for social rewards in this thesis and review several that have been studied in the literature. I select one of these and describe how we verified its rewarding nature. I end the chapter with a conceptual framework for thinking about the stages of social reward processing.

In Chapter 3, I use the social reward we selected in Chapter 2 to probe how the brain processes social rewards. I model subject behavioral choices to this social reward and compare it to their choices with monetary rewards. We searched for areas of the brain that respond to the value of these rewards in a parametric way and found overlap with the reward processing network cited in the neuroimaging literature.

In Chapter 4, I explore social reward processing in a clinical population that has deficits in social cognition. If social reward processing is identical to processing for other types of rewards, then it should be impossible to selectively knock out social reward processing. Yet autism spectrum disorder (ASD) is a case where this dissociation does seem to be present. In the first study I present, I use a charitable donation to assess preferences across a range of stimuli in the ASD population, and find a clear, specific domain impairment in social cognition. In the second study, I return to the paradigm used

in Chapter 3 to look at neural processing in the ASD population. People with ASD seem to perform comparably to neurotypicals in the monetary condition but not in the social condition. Yet imaging analysis reveal reduced activity in key reward areas in the social condition compared to the monetary condition.

In Chapter 5, I summarize my findings and discuss the contributions of my work to our understanding of social rewards and decision-making. Additionally, I suggest some open questions for the field.

References

Allman, JM. (1999). Evolving Brains. (pg. 173). New York, NY: Sci Am Library.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. Chicago, IL: University of Chicago Press.

Desimone, R., T. Albright, Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4, 2051-2062.

Ekman, P., Sorenson, E., and Friesen, W. (1969). Pan-cultural elements in facial displays of emotion. Science 164, 86-88.

Goren, C., Sarty, M. and Wu, P. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics 56, 544-549.

Haxby, J., Horwitz, B. et al. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci 14, 6336- 6353.

Henrich, J., Boyd, R. et al. (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav Brain Sci 28, 795-855.

Kanwisher, N. and Galit, Y. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361, 2109-2128.

Peelen, M. V. and Downing, P. E. (2007). The neural basis of visual body perception." Nat Rev Neurosci 8, 636-648.

Reader, S. M. and Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. Proc Natl Acad Sci USA 99, 4436-4441.

Tsao, D., Freiwald, W., Tootell, R and Livingstone, M. (2006). A cortical region consisting entirely of face-selective cells. Science 311, 670-674.

CHAPTER TWO

Are social rewards motivating?

Chapter 2: **What is a social reward and how is it processed?**

2.1 What is a social reward?

Defining social rewards is not easy. Social rewards can take different forms, can involve one or more sensory modalities, and even when it has the same perceptual properties, it may not be a social reward in different contexts. They can be evoked by the most basic social stimuli (like images of eyes or smiling faces) to more complex constructs (social rewards, like reputation and fairness) that involve complicated situations and multiple social parties.

2.1.1 Definition of social rewards

There are two ways of thinking about rewards in general. There is the traditional behaviorist definition that describes a reward as that which reinforces behavior (Skinner, 1935). Positive rewards induce approach behavior and negatives rewards induce avoidance behavior. Then there is the common sense definition, which links rewards to a hedonic experience and says a reward is something that one finds pleasurable. With advances in neuroimaging techniques and tools, there is now an increasing corpus of neuroscience data that also associates rewards with a specific set of brain structures, together comprising a reward-processing network.

Social rewards are no different. We can find social stimuli that reinforce, modify, and influence social behavior. People will also report that they find social rewards enjoyable. Whether social rewards activate the brain's reward system is a hypothesis that I will be testing in this thesis.

The working definition I use in my research is:

A social reward is a social interaction that people will seek out or work for.

I have intentionally chosen not to include social stimuli and social information in the definition. Not because they don't meet the criteria for rewards, but quite the opposite: social stimuli and social information are some of the most rewarding and salient stimuli.

While there is considerable overlap between social stimuli (like faces) and social information, I exclude social stimuli and social information in order to narrow the focus of this thesis.

2.1.2 Examples of social rewards

Researchers have tested a broad range of social rewards: smiling attractive faces (O'Doherty et al., 2003), pleasant touch (E. Rolls et al., 2003), giving to others (T Hare, Camerer, & Knoepfle…, 2010), acceptance/inclusion (Eisenberger, Lieberman, & Williams, 2003), compliments signaling approval/validation (K Izuma, 2008), revenge – punishment of unfair partners (De Quervain, 2004), and reputation/status (Zink et al., 2008). *Researchers have found that all these examples have reward value and change people's behavior from what is predicted by rational choice.*

Social rewards are impactful even when represented by simplified cues. Gold stars used in primary school symbolize high achievement and can confer increases in reputation and status. The commonly used thumbs-up sign signals social approval from

others. Bateson et al. (2006) found that images of eyes had a significant increase on the contributions to an honesty box used to collect money for drinks in a university coffee room compared to images of a flower (Figure 2.1). The authors theorized that cooperative behavior could have been induced in the participants by the perception of being watched and therefore reputational concerns. This process could be mediated through neurons in the human perceptual system that are selectively responsive to stimuli involving faces and eyes (Emery, 2000; Haxby, Hoffman, & Gobbini, 2000). This is one of many examples of weak, automatic, subconscious cues playing at some level on our desire for social rewards.

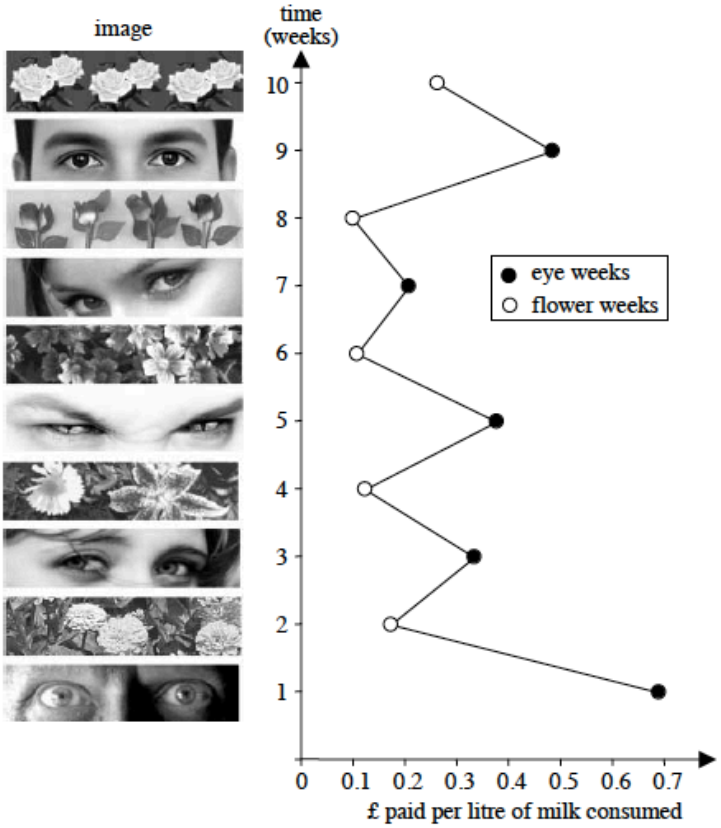


Figure 2.1 Pounds paid per liter of milk consumed as a function of week and image type. Study by Bateson et al. measuring effect of images of eyes on cooperation

In this thesis, I begin with a basic social reward –- positive, neutral, and negative faces with matching sound effects. What I mean by basic is that we expect it to be universally processed and experienced without any prior learning or context. All people should be spontaneously capable of experiencing this social reward.

2.2 How do we measure social rewards?

An important question before we begin our research is how do we know if a social stimulus, or any stimulus for that matter, is rewarding?

We must have a method for testing the reward properties of our stimulus. The key test is whether it motivates approach behavior for reward attainment. There are many different ways to behaviorally measure reward value. Aharon (2001), in a study looking at the reward value of aesthetic faces, used a "keypress" task to measure the amount of work subjects performed in order to change the relative duration they viewed different face images. The keypress task was able to evaluate how rewarding different categories of faces were (Aharon et al., 2001).

In our study, we operationalize social rewards with performance on a simple instrumental learning task. We asked participants to choose among slot machines associated with a distribution of differently valued outcomes. On every trial, participants were presented with two slot machines. Once he/she selected one, they were shown the outcome of their choice. Some slot machines were associated with positively valenced outcomes. We found subjects quickly learned to select these consistently over a slot machine associated with negatively valenced outcomes. Consistent selection of a slot machine indicated high motivation and reward for the stimuli associated with that slot machine.

2.3 How are social rewards processed?

Lastly I propose a framework for thinking about social rewards in decision-making to guide our discussion. The figure below lays out the general stages of social reward processing.

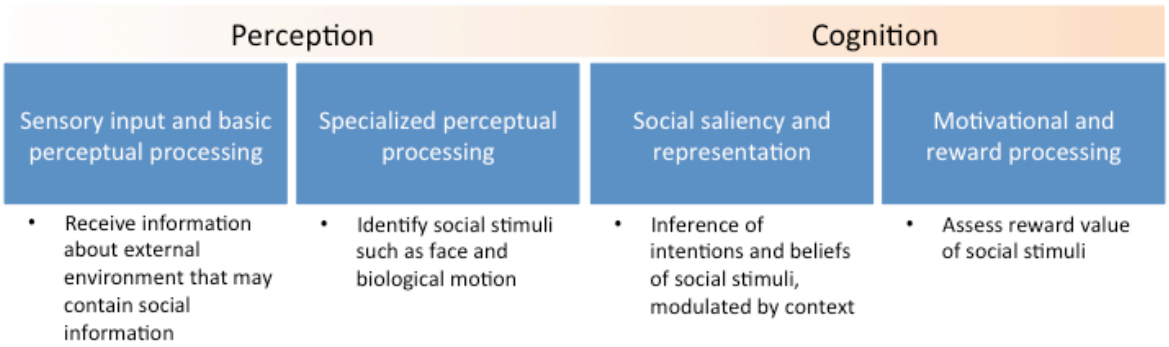


Figure 2.2 General stages of social reward processing

Social reward processing draws on many of the same brain structures involved in perception, cognition, and behavior more generally, but I highlight a few in each of the stages that consistently appear specialized in social cognition.

1. Sensory input and basic perceptual processing: How are socially relevant stimuli and signals perceived?

   Perception of social stimuli begins with the same sensory transduction as used with nonsocial objects. For example, social stimuli perceived through the visual system proceed through the same basic edge and feature extraction as nonsocial objects.

2. Specialized perceptual processing

Processing of social stimuli is already specialized at the level of perception. In the visual system, there is evidence that regions of higher-order visual cortex are disproportionately engaged by faces or by biological motion.

fusiform face area (FFA) --- a region of ventral temporal cortex that has larger responses to faces than to any other visual object category (N. Kanwisher, McDermott, & Chun, 1997).

3. Representation: Interpreting social stimuli

After creating a perception and representation of, say, a face, we go on to make inferences about emotions, intentions, and beliefs of the other person with contributions from the following regions:

medial prefrontal cortex (mPFC) – an area consistently activated when we think about other people's minds (Amodio & Frith, 2006; R. Saxe, 2006)

insula – an interoceptive somatosensory cortex involved in representing our own somatic states (Singer, Seymour, O'Doherty, & Kaube…, 2004)

amygdala – a structure providing rapid and automatic emotional processing for social cognition (Kluver & Bucy, 1997)

right temporal parietal junction (rTPJ) – an area implicated by many studies in attributing beliefs to others (R Saxe & Kanwisher, 2003)

posterior Superior Temporal Sulcus (STS) – a region that has been associated with interpreting the motions of a human body in terms of the person's goals (R.

Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004; J. Schultz, Imamizu, Kawato, & Frith, 2004)

4. Motivational and reward processing

A key input for social decision-making is attributing a motivational value to the social representation.

ventromedial prefrontal cortex(vmPFC) – this is a key valuation area for rewards and punishments (Ongur & Price, 2000). In fact, vmPFC damage associated with impairment in social behavior has been well documented in patients. One example is EVR, who had most of his vMPFC lesioned in a tumor resection procedure. Though the surgery was successful in removing the tumor, it brought profound changes to his personality that manifested in inappropriate social conduct. Despite changed social conduct and decision-making, neuropsychological testing showed no change in EVR's intellectual abilities (Saver & Damasio, 1991). vmPFC patients also experience diminished emotional arousal before making risky choices (Bechara, Tranel, Damasio, & Damasio, 1996). These results gave rise to a theory about the role of emotion in decision-making called the somatic marker hypothesis (Damasio, 1996), which argues that emotional signals guide decision-making, including those in the social domain. For example, damage to vMPFC appears to result in an inability to recognize social faux pas and reduces empathic concern for others (Shamay-Tsoory, Tomer, Berger, & Aharon-Peretz, 2003).

An important thing to keep in mind is that, while I laid out this simple framework linearly, in reality the flow of social information is multidirectional and recursive. There is extensive feedback everywhere in the brain. Interpreting social rewards, particularly complex ones, depends critically on context and intention. One way to conceptualize this is to imagine an initial feed-forward sweep driven by sensory areas that is rapid and automatic, followed by cycles of additional processing that progressively recruit additional regions of the cortex that are biased by the first information and top-down effects incorporating controlled processing and conscious intentions (Adolphs, 2009).

References

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. Annu. Rev. Psychol. 60, 693-716.

Aharon, I., N. Etcoff, et al. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron 32, 537-551.

Amodio, D. M. and C. D. Frith (2006). Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7, 268-277.

Bateson, M., D. Nettle, et al. (2006). Cues of being watched enhance cooperation in a real-world setting. Biol Lett 2, 412-414.

Bechara, A., D. Tranel, et al. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb Cortex 6, 215-225.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 351, 1413-1420.

De Quervain, D. J.-F. (2004). The Neural Basis of Altruistic Punishment. Science 305, 1254-1258.

Eisenberger, N. I., M. D. Lieberman, et al. (2003). Does rejection hurt? An FMRI study of social exclusion. Science 302, 290-292.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24, 581-604.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci *30*, 583-590.

Haxby, J., E. Hoffman, et al. (2000). The distributed human neural system for face perception. Trends in Cognitive Sciences 4, 223-233.

Izuma, K., Saito, D.N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. Neuron *58*, 284-294.

Kanwisher, N., J. McDermott, et al. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17, 4302-4311.

Kluver, H. and P. C. Bucy (1997). Preliminary analysis of functions of the temporal lobes in monkeys. 1939. J Neuropsychiatry Clin Neurosci 9, 606-620.

O'Doherty, J., J. Winston, et al. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia 41, 147-155.

Ongur, D. and J. L. Price (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10, 206-219.

Rolls, E., J. O'Doherty, et al. (2003). Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. Cereb Cortex 13, 308-317.

Saver, J. L. and A. R. Damasio (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. Neuropsychologia 29, 1241-1249.

Saxe, R. (2006). Uniquely human social cognition. Curr Opin Neurobiol 16, 235-239.

Saxe, R. and N. Kanwisher (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage 19, 1835-1842.

Saxe, R., D. K. Xiao, et al. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42, 1435-1446.

Schultz, J., H. Imamizu, et al. (2004). Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. J Cogn Neurosci 16, 1695-1705.

Shamay-Tsoory, S. G., R. Tomer, et al. (2003). Characterization of empathy deficits following prefrontal brain damage: the role of the right ventromedial prefrontal cortex. J Cogn Neurosci 15, 324-337.

Singer, T., B. Seymour, et al. (2004). Empathy for pain involves the affective but not sensory components of pain. Science 303, 1157-1162.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. Neuron 58, 273-283.

CHAPTER THREE

Neural substrates that respond to social rewards

Chapter 3: **Neural substrates that respond to social rewards**

3.1 Introduction

In Chapter 2 we selected a basic social reward and found that behaviorally it met the definition of reward. We now turn our attention to the neural properties of social rewards in the context of learning. We build off of the vast body of work that has looked at learning in the brain as our starting point.

The brain needs to compute several distinct signals in order for an organism to learn how to make sound decisions among alternatives. First, at the time of choice, values need to be assigned to the different stimuli associated with each choice option (which we refer to as stimulus values; SV); these are subsequently compared in order to choose the option with the highest value (Kable & Glimcher, 2009; A. Rangel, Camerer, & Montague, 2008; A. Rangel & Hare, 2010; Rushworth, Mars, & Summerfield, 2009; Wallis, 2007). Stimulus value signals have been found in ventral and medial sectors of the prefrontal cortex (vmPFC) in several human fMRI (Chib, Rangel, Shimojo, & O'Doherty, 2009; FitzGerald, Seymour, & Dolan, 2009; Todd Hare, Camerer, & Rangel, 2009; T. A. Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Kable & Glimcher, 2007; Levy, Snell, Nelson, Rustichini, & Glimcher, 2010; Litt, Plassmann, Shiv, & Rangel, 2009; Plassmann, O'Doherty, & Rangel, 2007; Plassmann, O'Doherty, & Rangel, 2010; Tom, Fox, Trepel, & Poldrack, 2007) and nonhuman primate electrophysiological studies (Kennerley, Dahmubed, Lara, & Wallis, 2009; Kennerley & Wallis, 2009; C. Padoa-Schioppa, 2009; Camillo Padoa-Schioppa & Assad, 2006; C. Padoa-Schioppa & Assad, 2008; Wallis & Miller, 2003) during choices involving non-social rewards, as

well as during social decisions such as donations to charities (T. A. Hare, Camerer, Knoepfle, & Rangel, 2010).

Having made a choice, the brain needs to compute the reward value associated with the outcomes generated by the choice. These signals are often called reward magnitude or experienced utility (R). Several human fMRI studies have found that activity in medial regions of orbitofrontal cortex correlates with behavioral measures of experienced utility for a wide variety of social and non-social reward modalities (Blood & Zatorre, 2001; de Araujo, Rolls, Kringelbach, McGlone, & Phillips, 2003; Kringelbach, 2005; McClure, Berns, & Montague, 2003; Plassmann, O'Doherty, Shiv, & Rangel, 2008; Small et al., 2003; Small, Zatorre, Dagher, Evans, & Jones-Gotman, 2001; Smith et al., 2010).

A third critical component is the combination of the previous two signals into a prediction-error signal (PE) that is used to update stimulus values (W. Schultz, Dayan, & Montague, 1997). The key involvement of the ventral striatum in this third component is borne out by a sizable and rapidly growing body of human fMRI studies of reinforcement learning that have used almost exclusively non-social rewards such as monetary payments (Berns, McClure, Pagnoni, & Montague, 2001; Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; T. A. Hare et al., 2008; J. O'Doherty et al., 2004; J. P. O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Pagnoni, Zink, Montague, & Berns, 2002; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Seymour, Daw, Dayan, Singer, & Dolan, 2007; Yacubian et al., 2006).

Although the findings summarized above have been replicated across species, techniques, and experimental designs, the vast majority of these studies have used only

non-social rewards such as juice, food, or money, and only a handful have directly compared social and non-social rewards. This raises a fundamental question: Do the same brain regions that implement reward-learning computations for non-social rewards also implement social? Or might the areas that encode SV, PE, and R be different for social rewards, analogous to the specialized perceptual processing of social stimuli (N. Kanwisher & Yovel, 2006)? While a very few other studies have recently approached this issue (K. Izuma, Saito, & Sadato, 2008; Smith et al., 2010; Zink et al., 2008), no study to date has investigated the question using identical tasks across the same subjects, and in a task that allows comparison of the encoding of the three types of basic reward signals defined above. We undertook such an investigation using model-based fMRI.

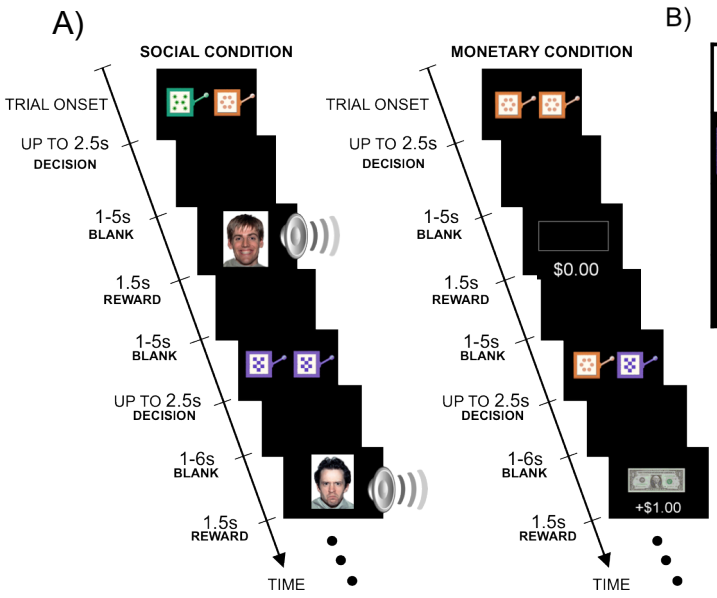
3.2 Methods

3.2.1 Participants

Twenty-seven female participants from the Caltech community participated in the study (mean age = 22.4 years; range 18-28). Five were excluded from further analysis: four due to excessive head movement, one due to failure to understand task instructions. All participants were fully right-handed, had normal or corrected-to-normal vision, had no history of psychiatric or neurological disease, and were not taking medications that might have interfered with BOLD-fMRI. All gave informed consent under a protocol approved by the Caltech IRB.

3.2.2 Task

Participants played two structurally identical versions of an instrumental learning task, one with monetary rewards, the second with social rewards (Figure 3.1A). A trial began with the display of two visually distinctive slot machines, each associated with one of three outcome distributions: mean-positive, mean-negative, or mean-neutral (Figure 3.1B) .



	POSITIVE	NEUTRAL	NEGATIVE
	0%	20%	80%
	80%	20%	0%
	1/3	1/3	1/3

Figure 3.1 Task and behavioral results. A) Timeline of the monetary and social reward trials. Choice trials paired a neutral slot machine with a valenced slot machine. Trials were identical except for the nature of the outcomes: Monetary trials had a gain/loss of +$1, $0, or -$1, whereas social trials revealed happy, neutral, or angry faces accompanied by sound effects of similar emotional valence. The experiment also included no-choice trials (in which a pair of identical slot machines were shown, neutral, negative, or positive) to help separate the learning and stimulus value signals. Specific slot machines were randomly assigned to specific reward outcomes at the start of the experiment for each subject, and distinct between monetary and social condition blocks. B) Distribution of outcomes for each slot machine. First row: negative machine. Second row: positive machine. Bottom row: neutral machine. The same distribution was used in the monetary and social conditions. Actual appearance of the slot machines was randomly paired with a reward outcome distribution, and distinct between monetary and social condition blocks.

All participants completed one social and one monetary block of 148 trials each; block order was randomized between participants. There were two types of trials in each block. In 100 choice trials the neutral slot machine was shown paired with either the positive or negative slot machine (50/50 probability with randomized order), and participants chose one by pressing a left or right button. We refer to these as free-choice trials. In 48 non-choice trials, two identical copies of one of the three slot machines were shown (1/3, 1/3, 1/3 probability with randomized order), and participants merely pressed either the left or right button in order to advance the trial. We refer to these as forced-choice trials. Up to 2.5 seconds were allowed for choice in both cases, followed by a uniformly blank screen displayed for 1-5 seconds (flat distribution), followed by the reward outcome displayed for 1.5 seconds, followed by an intertrial interval of a uniformly blank screen displayed for 1-6 seconds (flat distribution). Note that participants were not told the reward probabilities associated with each slot machine and had to learn them by trial and error during the task.

The forced trials provide an essential control for a potential important confound in the study. One potential concern is that the presentation of positive and aversive social outcomes might induce in the brain "correct" and "error" feedback signals at outcome during the social trials. This is a problem because this would suggest that the common locus of activity is not due to the activation of a social reward, but to the activation of these error feedback signals. The forced trials provide a control for this concern because when there is no free choice, there can be no error feedback regarding the correctness of the choice.

3.2.3 Stimuli and rewards

The slot machines in both conditions were represented by cartoon images of actual slot machines that varied in color and pattern (Figure 3.1). In the social condition, reward outcomes were color photographs of unfamiliar faces from the NimStim collection (Tottenham et al., 2009) showing either an angry (negative outcome), neutral (neutral outcome), or happy (positive outcome) emotional expression, presented together with emotionally matched words played through headphones (normalized for volume and duration). Examples of positive words are "excellent", "bravo", and "fantastic". Examples of negative words are "stupid", "moron", and "wrong". Examples of neutral words are "desk", "paper", and "stapler". Extensive prior piloting had demonstrated the behavioral efficacy of these stimuli in reward learning.

In the monetary condition, the positive outcome was a gain of one dollar (an image of a dollar bill), the negative condition was a loss of one dollar (image of a dollar bill crossed out), and the neutral condition involved no change in monetary payoff (image of an empty rectangle). Subjects were paid out the sum of their earnings at the end of the experiment.

3.2.4 Computational model

We computed trial- and subject-specific values for each of the three variables described in the Introduction. The stimulus value (SV) for every slot machine was calculated as the 10-trial moving average proportion of times that the machine was chosen when it was shown, a continuous value between 0 and1. Consistent with this coding, reward outcomes (R) were assigned a value of 1 if they were positive; a value of

0.5 if they were neutral, and a value of 0 if they were negative. Prediction errors (PE) at the time of outcome were calculated using a simple Rescorla-Wagner learning rule (Rescorla and Wagner, 1972) as the difference between the value of the reward outcome and the stimulus value of the machine selected for that trial: PEt = Rt – SVt.

Note three things about the value normalizations. First, our approach deviates from the usual practice in neuroscience studies of reinforcement learning (T. A. Hare et al., 2008; Lohrenz, McCabe, Camerer, & Montague, 2007; Pessiglione et al., 2008; Pessiglione et al., 2006; Seymour et al., 2007; Wunderlich, Rangel, & O'Doherty, 2009), in which it is customary to fit the values of the SV signal based on the predictions of the best-fitting learning model. Here we depart from that practice because the revealed preference approach provides more accurate measures of the values computed at the time of choice (as shown in Figure 3.1D). Second, without loss of generality, we normalize the reward outcome signals to 0 for negative outcomes and 1 for positive outcomes. Note that, given the parametric nature of the general linear model specified below, this normalization does not affect the identification of areas that exhibit significant correlation with this variable. Third, we use the standard definition of prediction errors used in the literature.

3.2.5 Image acquisition

T2*-weighted gradient-echo echo-planar (EPI) images with BOLD contrast were collected on a Siemens 3T Trio. To optimize signal in the orbitofrontal cortex (OFC), we acquired slices in an oblique orientation of 30^{o} to the anterior commissure-posterior commissure line (Deichmann et al, 2003) and used an eight-channel phased array

headcoil. Each volume comprised 32 slices. Data was collected in four sessions (~12 min each). The imaging parameters were as follows: TR= 2 s, TE= 30 ms, FOV= 192 mm, 32 slices with 3mm thickness resulting in isotropic 3mm voxels. Whole-brain high-resolution T1-weighted structural scans (1 x 1 x 1 mm) were co-registered with their mean T2*-weighted images and averaged together to permit anatomical localization of the functional activations at the group level.

3.2.6 fMRI pre-processing

The imaging data was analyzed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were corrected for slice acquisition time within each volume, motion-corrected with realignment to the last volume, spatially normalized to the standard Montreal Neurological Institute EPI template, and spatially smoothed using a Gaussian kernel with a full-width at half-maximum of 8mm. Intensity normalization and high-pass temporal filtering (filter width = 128s) were also applied to the data.

3.2.7 fMRI data analysis

The data analysis proceeded in three steps. First, we estimated a general linear model with AR(1). This model was designed to identify regions in which BOLD activity was parametrically related to SV, R, and PE. The model included the following regressors:

R1) An indicator function for the decision screen in free-choice monetary trials.

R2) An indicator function for the decision screen in free-choice monetary trials multiplied by the SV of the two slot machines shown in that trial (summed SV).

R3) An indicator function for the decision screen in free-choice monetary trials multiplied by the reaction time for that trial.

R4)-R6) Analogous indicator functions for decision screen events in free-choice social trials.

R7) An indicator function for the decision screen in forced monetary trials.

R8) An indicator function for the decision screen in forced monetary trials multiplied by the SV of the slot machine displayed.

R9)-R10) Analogous indicator functions for decision screen events in forced social trials.

R11) A delta function for the time of response in the monetary condition.

R12) A delta function for the time of response in the social condition.

R13) An indicator function for the outcome screen in free monetary trials (both choice and non-choice).

R14) An indicator function for the outcome screen in free monetary trials multiplied by the PE for the trial.

R15) An indicator function for the outcome screen in free monetary trials multiplied by the R for the trial.

R16)-R18) Analogous indicator functions for outcome screen events in free social trials (both choice and non-choice).

We orthogonalized the modulators for the main regressors that had more than one modulator (e.g., R2 and R3). The model also included six head-motion regressors, session constants, and missed trials as regressors of no interest. The regressors of interest and missed-trial regressor were convolved with a canonical HRF.

Second, we calculated the following first-level single-subject contrasts: 1) R2 vs. baseline, 2) R5 vs. baseline, 3) R14 vs. baseline, 4) R15 vs. baseline, 5) R17 vs. baseline, and 6) R18 vs. baseline.

Third, we calculated second-level group contrasts using a one-sample t test of the first-level contrast statistics. Finally, we also performed a conjunction analysis between the equivalent contrasts for the monetary and social conditions to identify areas involved in similar computations in both cases. The results are shown in Figure 3.4 and reported in Tables 3.1--3.3. For inference purposes we used an omnibus threshold of $p<.001$ uncorrected with an extent threshold of 15 voxels. However, given the strong priors from the previous literature about the role of the vmPFC in encoding stimulus value and reward outcome signals, as well as the role of the ventral striatum in encoding prediction errors, we also report activity in these two areas if they survive small volume corrections (SVC) at $p<.05$. The mask for the SVC in vmPFC at choice was taken using a sphere of 10mm radius defined around the peak activation coordinates that correlated with stimulus values in Rolls et al. (2008a). The mask for the vmPFC SVC at reward outcome was given by a sphere of 10-mm radius defined around the peak coordinates that correlated with the magnitude of reward outcome in O'Doherty et al. (2002). The mask for the SVC in ventral striatum was taken using a sphere of 10mm radius defined around the peak activation coordinates that correlated with prediction errors in Pessiglione et al. (2006). For display purposes only activity in selected SPMs is reported at $p<.005$ uncorrected with an extent threshold of 5 voxels. Anatomical localizations were performed by overlaying the t maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas (Duvernoy, 1999).

3.3 Results

3.3.1 Behavioral results

Participants reliably learned to select the slot machine associated with the highest probability of a positive-valenced outcome within a few choice trials for both social and non-social rewards (Figure 3.2). The figure also reveals two additional interesting patterns about the learning process. First, participants were somewhat slower at learning to discriminate between social rewards than between monetary rewards. For example, by the tenth exposure, the positive monetary machine was chosen with 92% frequency whereas the social positive machine was chosen with 72% frequency ($p<.001$). Second, participants were slower in learning to avoid the negative slot machines than in learning to choose the positive ones. For example, by the tenth presentation the positive slot machines were chosen 85% of the time, whereas the negative ones were avoided only 68% of the time ($p<.001$). Both differences were not significant on the last third of the learning trials, which suggests that they are related to the speed of learning, and not to the ability to ultimately learn the value of the stimuli.

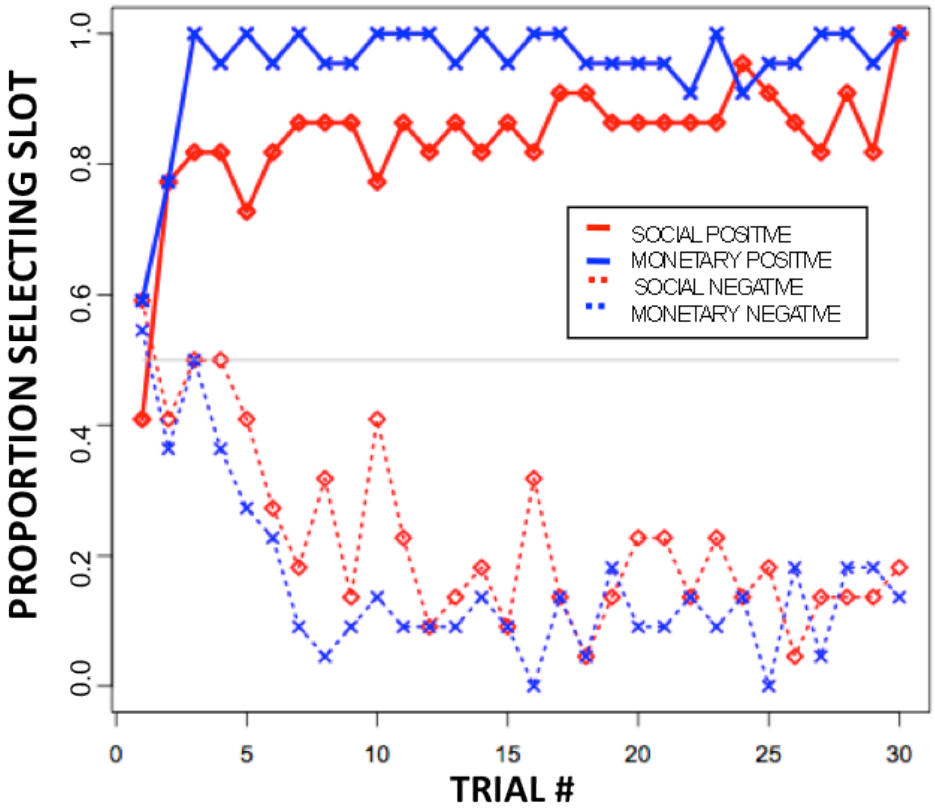


Figure 3.2 Plot of group subject choices across trials. (Only the first 30 are shown.)

Figure 3.3 shows the psychometric choice curves for the social and monetary conditions based on their SV. Note several things about the curves: First, when the values of valenced and neutral slot machines were identical, participants exhibited no choice bias (0.5 on the y-axis corresponds to 0.0 on the x-axis). Second, the choice curves are not significantly different from each other (greatest difference at x=0.25 had p=.32 with Bonferroni correction). Third, the choice curve is asymmetric: whereas participants chose the valenced slot machine over the neutral slot machine with probability close to one when its relative stimulus value was sufficiently positive (far-right side of curve),

subjects chose the neutral slot machine only 80% of the time even when it was the most favorable (far-left side of curve).

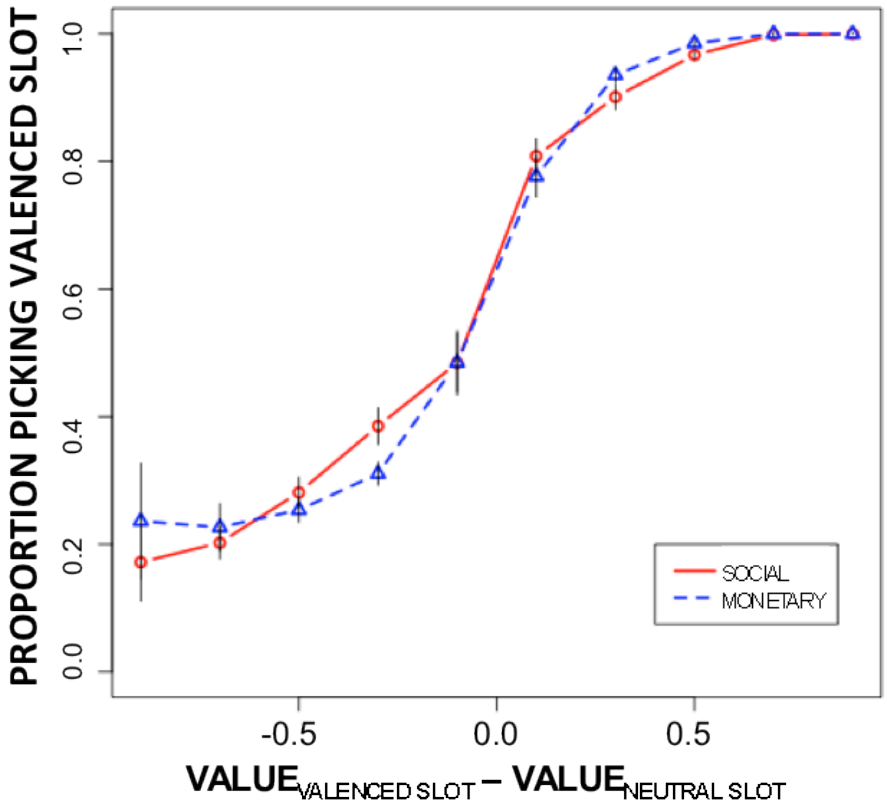


Figure 3.3 Psychometric choice curve for monetary and social conditions. Bars denote standard error measures computed across subjects.

3.3.2 Neural correlates of stimulus values (SV)

We estimated a parametric general linear model of the BOLD signal to identify areas in which activation correlated with SV at the time of choice, and with PE and R at outcome during free-choice trials (see Methods for details). In the free-choice monetary task, activation in the vmPFC correlated with SV of the slot machines. SV signals were additionally found in the mid-cingulum, the superior frontal gyrus, and the angular gyrus (**Table 3.1** and Figure 3.4). In the free-choice social task, activation correlating with SV was also found in a similar region of vmPFC. A conjunction analysis showed that activation in a common area of vmPFC correlated with SV in both social and monetary conditions.

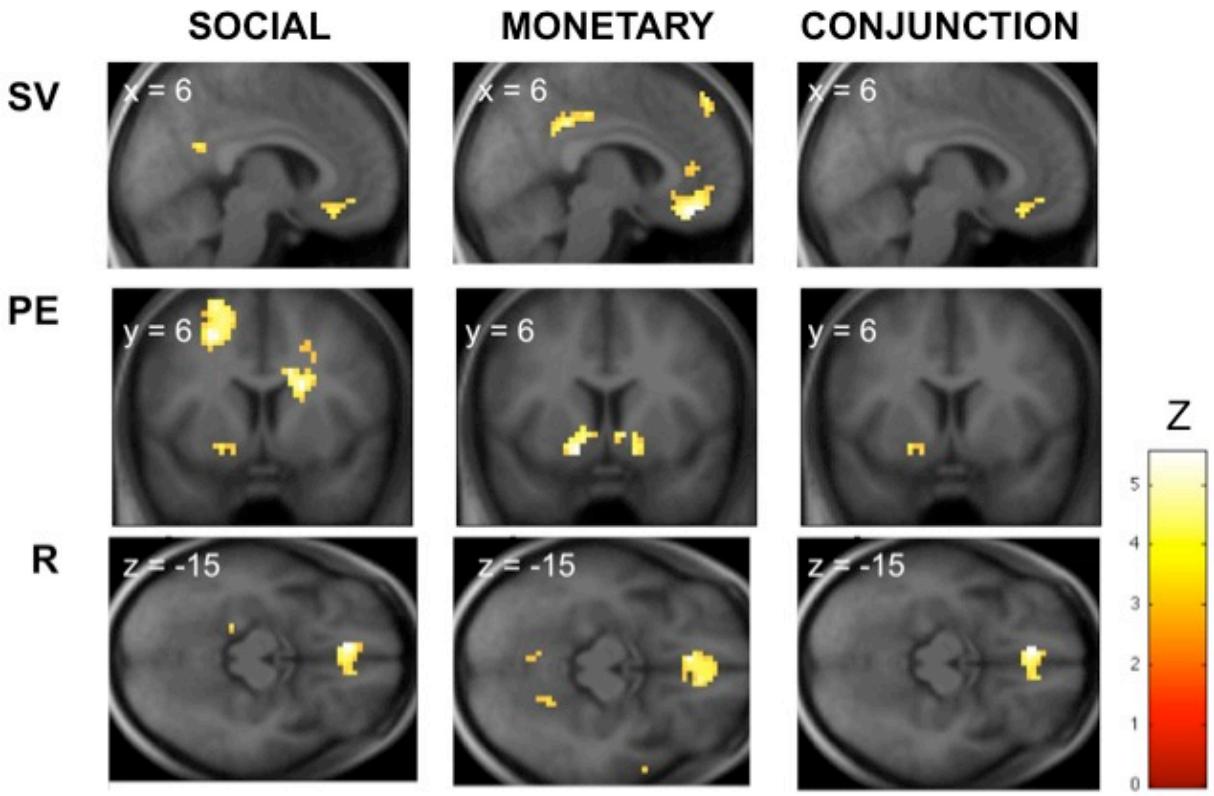


Figure 3.4 Basic neuroimaging results. Top) Activation in the vmPFC correlated with *SV* at the time of free choice in both monetary and social conditions. Middle) Activation in the vStr correlated with *PE* at the time of outcome in both monetary and social free choice conditions (albeit the conjunction did not survive our omnibus threshold). Bottom) Activation in the vmPFC correlated with *R* in both monetary and social free-choice conditions. For illustration purposes only, all images are thresholded at $p<.005$ uncorrected with an extent threshold of 15 voxels, except for the conjunction of *PE* which is $p<0.005$ with an extent threshold of 5 voxels (see Tables 3.1--3.3 for details).

Table 3.1 Regions correlating with stimulus value at cue

Region	# Voxels	Z score	x	y	z
Areas correlating with *SV* in monetary choice trials (R2 vs. baseline)					
Medial Orbitofrontal Cortex	214	4.53†	0	27	-21
Frontal Superior	52	4.19	-18	42	51
Mid Cingulum	46	4.01	0	-30	45
Angular Gyrus	61	3.91	-57	-66	30
Middle Temporal Gyrus	24	3.85	60	-15	-6
Areas correlating with SVs in social choice trials (R5 vs. baseline)					
Medial Orbitofrontal Cortex	40	3.16†	6	27	-15
Areas correlating with SVs in both monetary and social choice trials					
Medial Orbitofrontal Cortex	37	3.16†	6	27	-15

Regions are significant at $p<0.001$ uncorrected and 15 voxels extent threshold.
†Survives $p<0.05$ small volume correction. Coordinates reported in MNI space.

3.3.3 Neural correlates of prediction errors (PE)

In the free-choice monetary task, PE correlated with activation in the caudate and putamen (**Table 3.2**, Figure 3.4). In the free choice social task, PE did not exhibit any correlations at our omnibus threshold ($p<.001$ uncorrected, 15 voxels). However, for completeness we show areas of the striatum that correlate with PE in the social free-choice condition at $p<.005$ uncorrected, as well as the resulting conjunction results using this lower threshold.

Table 3.2 Regions correlating with prediction error at outcome

Region	# Voxels	Z score	x	y	z
Areas correlating with *PE* in monetary choice trials (R13 vs. baseline)					
Putamen	25	4.07†	-15	6	-12
Caudate	22	3.75	9	9	-3
Precuneus	15	3.49	-18	-51	33
Areas correlating with *PE* in social choice trials (R16 vs. baseline)					
-	-	-	-	-	-
Areas correlating with *PE* in both monetary and social choice trials					
-	-	-	-	-	-

Regions are significant at $p<0.001$ uncorrected and 15 voxels extent threshold.
†Survives $p<0.05$ small volume correction. Coordinates reported in MNI space.

3.3.4 Neural correlates of reward magnitude (R)

In the free-choice monetary task, reward outcome correlated with activation in vmPFC, insula, occipital cortex, cingulate gyrus, and superior frontal gyrus (**Table 3.3**, Figure 3.4). In the free-choice social task, reward outcome correlated with activation in vmPFC. A conjunction analysis revealed that activation in a common area of the vmPFC correlated with reward magnitude in the social and non-social conditions.

Table 3.3 Regions correlating with reward at outcome

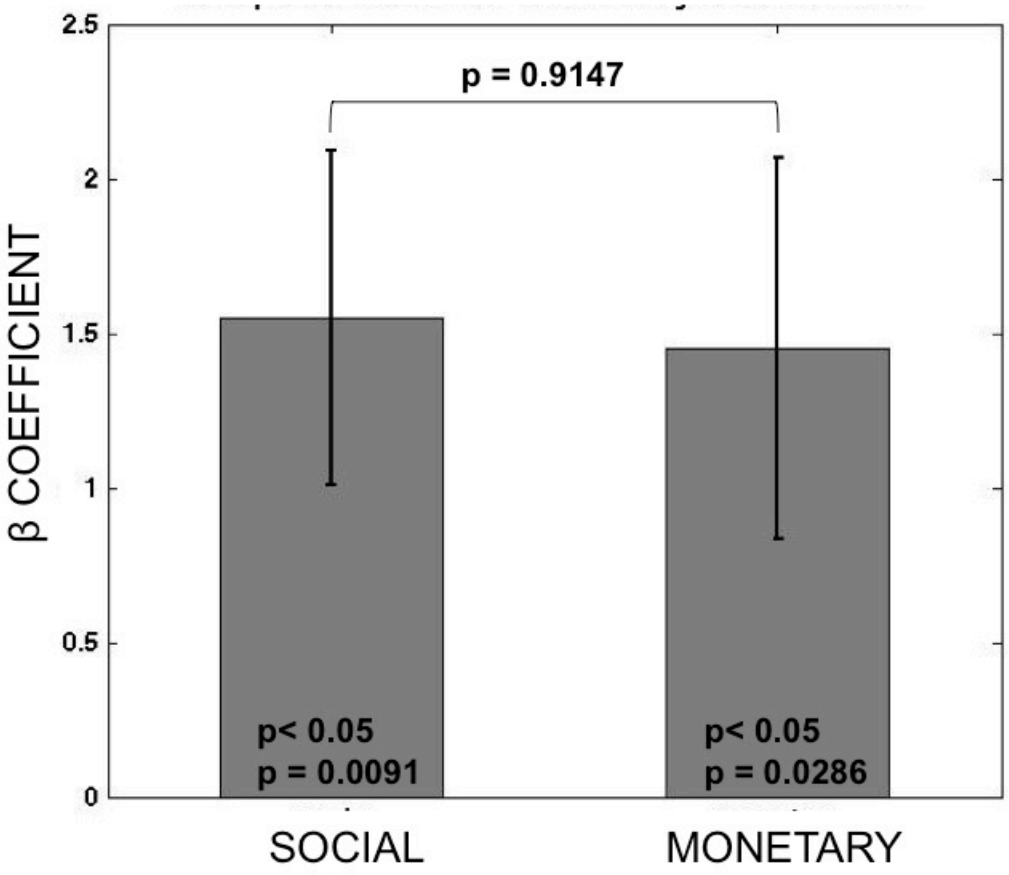
Areas correlating with *R* in monetary choice trials (R14 vs. baseline)					
Region	# Voxels	Z score	x	y	z
Occipital	124	4.74	21	-75	15
Insula	125	4.68	-33	3	12
Inferior Parietal	116	4.43	-51	-36	27
Occipital	59	4.29	-6	87	18
Insula	33	4.23	39	-18	18
Cingulum	52	3.99	-6	9	36
Medial Frontal Gyrus	86	3.96	-15	-6	57
Inferior Parietal	78	3.95	51	-33	30
Medial Orbitofrontal Cortex	136	3.88†	6	33	-12
Superior Frontal Gyrus	26	3.84	-18	27	57
Superior Frontal Gyrus	20	3.66	-30	36	33
Rolandic Operculum	18	3.66	57	0	12
Heschl Gyrus	21	3.63	-39	-24	3
Inferior Parietal	21	3.61	-36	-27	24
Calcarine	15	3.42	-18	-72	9
Areas correlating with *R* in social choice trials (R17 vs. baseline)					
Medial Orbitofrontal Cortex	29	4.16†	-6	36	-15
Areas correlating with *R* in both monetary and social choice trials					
Medial Orbitofrontal Cortex	129	4.16†	-6	36	-15

Regions are significant at $p<0.001$ uncorrected and 15 voxels extent threshold.
†Survives $p<0.05$ small volume correction. Coordinates reported in MNI space.

3.3.5 Ruling out a potential confound

A non-trivial potential confound is that the happy and angry faces might activate "correct" and "error" feedback signals in the brain regarding the adequacy of choice, and that the areas of co-activation might be due to the presence of these error signals, and not the computation of social rewards. In fact, these types of stimuli have previously been used just for that purpose (Cools, Lewis, Clark, Barker, & Robbins, 2007). Fortunately, the forced-choice trials provide a control that allows us to test if the previous results are driven by this potential confound. Figure 3.5 describes the strength of the correlation between outcome reward signals and BOLD activity in the area of vmPFC identified by the conjunction of outcome rewards in both conditions. It shows that the strength of the correlation in the social and monetary trials is of similar magnitude and not statistically different ($p=0.91$, two-sided paired t-test) even in the absence of error feedback. This implies that the signal in the vmPFC during social outcomes cannot be attributed to error feedback, and that the concern about the potential confound in this task was unfounded.

Figure 3.5 ROI analysis of outcome reward signals in vmPFC during forced-choice trials. Average beta plots for activity during reward outcome in forced-choice trials. The functional mask of vmPFC is given by the area that exhibits correlation with reward outcomes in social and monetary free-choice trials at p<.05 SVC. The p-values inside the bars are for t-tests versus zero.



3.4 Discussion

A fundamental open question in behavioral and social neuroscience is whether the brain utilizes a common representation of valuation, and whether this representation includes social rewards in learning how to make sound decisions. Prior evidence suggested that there might be an overlap in how the brain encodes value signals for social and non-social rewards. In the case of stimulus values, a recent paper found that the values of charities at the time of decision making were encoded in areas of the vmPFC that overlap with those that have been found for private rewards (T. A. Hare et al., 2010). In the case of experienced utility for social rewards, several studies found that activity in the orbitofrontal cortex correlates with the perceived attractiveness of faces (Aharon et al., 2001; Cloutier, Heatherton, Whalen, & Kelley, 2008; J. O'Doherty et al., 2003; Smith et al., 2010). Finally, in the case of prediction errors, studies have found that activity in the ventral striatum correlates with prediction error-like signals in a task involving the receipt of anticipated social rewards (Spreckelmeyer et al., 2009) and in tasks involving social reputation and status (K. Izuma et al., 2008; Zink et al., 2008). These latter two studies in particular compared both social and monetary rewards, as we did in the present study, and provided strong initial evidence for the idea that neural representations for these two types of rewards are at least partly overlapping. What has been missing to date is a study that compares social and non-social rewards across tasks whose basic structure and reward probabilities are matched for the two types of rewards, and in which the three basic computations associated with reward learning (SV, PE, and R) are at work.

We addressed this open question by asking subjects to perform an otherwise identical simple probabilistic-learning decision-making task in which stimuli were associated with either monetary or social rewards. We found evidence for common signals in all cases: a common area of vmPFC correlated with SV, a common area of vmPFC correlated with R, and common areas of ventral striatum correlated with PE, albeit in the later case only at a relatively low threshold of $p<.005$ unc. Together with other recent findings (Chib et al., 2009; T. A. Hare et al., 2010; K. Izuma et al., 2008; Zink et al., 2008), our results provide increasing support that overlapping areas of vmPFC and ventral striatum encode value signals for both types of rewards (Montague & Berns, 2002; A Rangel, 2008).

Behaviorally, our subjects were slower to learn the value of social and negative stimuli. Since the type of reinforcement learning models that have been successfully used to account for the behavioral data do not predict such asymmetries (Montague & Berns, 2002; Niv & Montague, 2008; Rescola & Wagner, 1972; Sutton & Barto, 1998), this raises an apparent puzzle. However, there are two potential explanations for this aspect of the findings. First, the reward magnitude of both types of stimuli might not have been perfectly matched in our population (so that, for example, subjects found the $1 outcome more rewarding than the positive social stimuli). Second, individuals stop selecting the negative slot machine after a while, which means that learning stops and subjects might not get sufficient negative reinforcement to learn the full extent of the negative outcomes associated with these machines.

We emphasize that the existence of areas involved in the encoding of reward in social and non-social situations does not mean that the full network involved in processing both types of rewards is identical. For example, it is known that areas involved in theory of mind computations are more likely to become active during social decisions than during choices among non-social rewards (Krach, Paulus, Bodden, & Kircher, 2010; R. Saxe, 2006; R. Saxe & Kanswisher, 2003).

It is important to highlight two limitations of our results. First, given the limited spatial resolution of fMRI we cannot rule out the possibility that there might be neuronal subpopulations within the vmPFC and ventral striatum specialized in valuing certain types of rewards. Future studies using fMRI adaptation designs, or direct electrophysiological recordings within these regions, will have to address this issue before the existence of a common valuation currency can be definitely established. Second, previous experiments suggest that males and females process some types of social rewards differently (Spreckelmeyer et al., 2009), which opens the possibility that there might be a gender difference in the extent to which common circuitry is used in the social and non-social domains to carry out basic reward computations. Unfortunately, we cannot resolve this issue with this dataset since only females participated in the experiment.

References

Aharon, I., Etcoff, N., Ariely, D., Chabris, C.F., O'Connor, E., and Breiter, H.C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron *32*, 537-551.

Berns, G.S., McClure, S.M., Pagnoni, G., and Montague, P.R. (2001). Predictability modulates human brain response to reward. J Neurosci *21*, 2793-2798.

Blood, A.J., and Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci U S A *98*, 11818-11823.

Chib, V.S., Rangel, A., Shimojo, S., and O'Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J Neurosci *29*, 12315-12320.

Cloutier, J., Heatherton, T.F., Whalen, P.J., and Kelley, W.M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. J Cogn Neurosci *20*, 941-951.

Cools, R., Lewis, S.J., Clark, L., Barker, R.A., and Robbins, T.W. (2007). L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. Neuropsychopharmacology *32*, 180-189.

de Araujo, I.E., Rolls, E.T., Kringelbach, M.L., McGlone, F., and Phillips, N. (2003). Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. Eur J Neurosci *18*, 2059-2068.

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

Duvernoy, H.M. (1999). The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply (Berlin, Springer).

FitzGerald, T.H., Seymour, B., and Dolan, R.J. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci *29*, 8388-8395.

Hare, T., Camerer, C., and Rangel, A. (2009). Self-control in decision-making involves modulation of the vMPFC valuation system. Science *324*, 646-648.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci *30*, 583-590.

Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W., and Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci *28*, 5623-5630.

Izuma, K., Saito, D.N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. Neuron *58*, 284-294.

Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. Nat Neurosci *10*, 1625-1633.

Kable, J.W., and Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. Neuron *63*, 733-745.

Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci *361*, 2109-2128.

Kennerley, S.W., Dahmubed, A.F., Lara, A.H., and Wallis, J.D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci *21*, 1162-1178.

Kennerley, S.W., and Wallis, J.D. (2009). Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci *29*, 2061-2073.

Krach, S., Paulus, F.M., Bodden, M., and Kircher, T. (2010). The rewarding nature of social interactions. Front Behav Neurosci *4*, 22.

Kringelbach, M.L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. Nat Rev Neurosci *6*, 691-702.

Levy, I., Snell, J., Nelson, A.J., Rustichini, A., and Glimcher, P.W. (2010). The neural representation of subjective value under risk and ambiguity. Journal of Neurophysiology *(forthcoming)*.

Litt, A., Plassmann, H., Shiv, B., and Rangel, A. (2009). Dissociating goal value and attention signals during simple decision making. Cereb Cortex *(in press)*.

Lohrenz, T., McCabe, K., Camerer, C.F., and Montague, P.R. (2007). Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci U S A *104*, 9493-9498.

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38*, 339-346.

Montague, P.R., and Berns, G.S. (2002). Neural economics and the biological substrates of valuation. Neuron *36*, 265-284.

Niv, Y., and Montague, P.R. (2008). Theoretical and empirical studies of learning. In Neuroeconomics: Decision-Making and the Brain, P.W. Glimcher, E. Fehr, C. Camerer, and R.A. Poldrack, eds. (New York, Elsevier).

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452-454.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D.M., and Dolan, R.J. (2003a). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia *41*, 147-155.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003b). Temporal difference models and reward-related learning in the human brain. Neuron *38*, 329-337.

O'Doherty, J.P., Deichmann, R., Critchley, H., and Dolan, R.J. (2002). Neural Responses During Anticipation of a Primary Taste Reward. Neuron *33*, 815-826.

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. J Neurosci *29*, 14004-14014.

Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature *441*, 223-226.

Padoa-Schioppa, C., and Assad, J.A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat Neurosci *11*, 95-102.

Pagnoni, G., Zink, C.F., Montague, P.R., and Berns, G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. Nat Neurosci *5*, 97-98.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R.J., and Frith, C.D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. Neuron *59*, 561-567.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042-1045.

Plassmann, H., O'Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci *27*, 9984-9988.

Plassmann, H., O'Doherty, J., and Rangel, A. (2010). Aversive goal values are negatively encoded in the medial orbitofrontal cortex at the time of decision making. Journal of Neuroscience *(forthcoming)*.

Plassmann, H., O'Doherty, J., Shiv, B., and Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. Proc Natl Acad Sci U S A *105*, 1050-1054.

Rangel, A. (2008). The computation and comparison of value in goal-directed choice. In Neuroeconomics: Decision Making and the Brain, P.W. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack, eds. (New York, Elsevier).

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci *9*, 545-556.

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. Curr Opin Neurobiol *20*, 262-270.

Rescola, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (New York, N.Y., Appleton Century Crofts), pp. 406-412.

Rolls, E.T., McCabe, C., and Redoute, J. (2008). Expected Value, Reward Outcome, and Temporal Difference Error Representations in a Probabilistic Decision Task. Cereb Cortex *18*, 652-663.

Rushworth, M.F., Mars, R.B., and Summerfield, C. (2009). General mechanisms for making decisions? Curr Opin Neurobiol *19*, 75-83.

Saxe, R. (2006). Uniquely human social cognition. Curr Opin Neurobiol *16*, 235-239.

Saxe, R., and Kanswisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". Neuroimage *19*, 1835-1842.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593-1599.

Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. J Neurosci *27*, 4826-4831.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M.M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron *39*, 701-711.

Small, D.M., Zatorre, R.J., Dagher, A., Evans, A.C., and Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: from pleasure to aversion. Brain *124*, 1720-1733.

Smith, D.V., Hayden, B.Y., Truong, T.K., Song, A.W., Platt, M.L., and Huettel, S.A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. J Neurosci *30*, 2490-2495.

Spreckelmeyer, K.N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., Kircher, T., and Grunder, G. (2009). Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. Soc Cogn Affect Neurosci *4*, 158-165.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (Cambridge, MIT Press).

Tom, S.M., Fox, C.R., Trepel, C., and Poldrack, R.A. (2007). The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science *315*, 515-518.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., and Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res *168*, 242-249.

Wallis, J.D. (2007). Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci *30*, 31-56.

Wallis, J.D., and Miller, E.K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur J Neurosci *18*, 2069-2081.

Wunderlich, K., Rangel, A., and O'Doherty, J.P. (2009). Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci U S A *106*, 17199-17204.

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D.F., and Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J Neurosci *26*, 9530-9537.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. Neuron *58*, 273-283.

CHAPTER FOUR

Social Rewards in Autism

Chapter 4: **Social Rewards in Autism**

4.1 Motivation for testing in people with autism

In Chapter 3, we found that social and monetary rewards processing engage overlapping neural circuitry. Here we look at social rewards in the autism spectrum disorder (ASD) population, a group selectively impaired in social cognition. Autism has been characterized by dysfunction in social cognition along with deficits in communication and language skills and restricted interests and repetitive behaviors (Kanner, 1968; Rogers, Ozonoff, & Maslin-Cole, 1991).

The deficits in social behavior and communication in people with autism make them a particularly attractive population in which to study social rewards. Differences between neurotypicals and a population with impaired social cognition could provide insight into what makes social reward processing unique. Additionally, identified deficits in processing in the ASD population could lead to development of therapeutic interventions for individuals with ASD.

As laid out in Chapter 2, there are several stages of processing in social reward valuation and decision making. For the ASD population, impairments could arise in several different stages --- from motivational, attentional, sensory, to more complex cognitive processing abnormalities. One theory consistent with observations in ASD and our current framework for understanding reward learning in cognitive neuroscience is that a lack of motivation and attention for social stimuli early in development could result in later impairments in perceptual and cognitive processing of social stimuli that might depend on normal social input during development (Dawson et al., 2002; Dawson,

Meltzoff, Osterling, Rinaldi, & Brown, 1998; Grelotti, Gauthier, & Schultz, 2002). For instance, it is known that neural and behavioral specializations for face processing depend in part on expertise with faces traceable to early domain-specific processing, so one plausible scenario could be that an early lack of motivation to orient towards faces results in reduced sensory input about faces and later reduced ability to process faces. In support of a developmental role for such altered preferences, it is known that children with autism fail to orient normally to social stimuli (Dawson et al., 1998). One study found that a person with autism showed activation of the fusiform face area not to real faces, but faces of preferred cartoon characters (Grelotti et al., 2002).

We present two studies with the ASD population in this chapter that further probes our understanding of social reward processing.

The first study addresses whether ASD involves a domain-specific impairment for the valuation of social stimuli. To assess preferences across a range of stimuli, we measured real monetary donations to 50 charities spanning categories pertaining to people, mental health, animals, or the environment. Whereas basic preferences for stimuli can be investigated using measures such as eye tracking in infants, complex real-world preferences in adults are more difficult to assess. We wanted to capture possible impairments at any stage of processing during complex decisions based on relative preferences, and thus chose to measure anonymous charitable donations involving real money. While charitable donations no doubt are based on preferences for the charities, the on-line computation of such preferences likely draws on multiple processes ranging from empathic and altruistic considerations to reward processing. A recent fMRI study (T. A. Hare et al., 2010) demonstrated that charitable donations activate regions within

the ventromedial prefrontal cortex thought to encode a common reward currency (Chib et al., 2009; Lin, Adolphs, & Rangel, 2011), as well as regions in the insula and superior temporal sulcus likely involved in empathy, social attention, and altruistic thinking (Frith, 2007; T. Singer et al., 2004). A model motivated by this and related studies is that social preferences during charitable giving are constructed via inputs to the ventromedial prefrontal cortex from regions concerned with processing information about the benefits to others (T. A. Hare et al., 2010), just as this region of the prefrontal cortex constructs values from sensory representations in more posterior cortices in general (Harris, Adolphs, Camerer, & Rangel, 2011). While several of these putative processing stages are thought to be impaired in people with autism --- both basic social reward processing and more complex evaluations of social stimuli that depend on context, mentalizing, or empathy have been reported to be abnormal in autism --- the primary goal in this study was not to dissect these processing components, but rather to provide an inventory across different types of stimuli, some social and others not.

The second study in this chapter investigates the neural basis of this impairment in basic social reward processing. We re-use the basic experiment paradigm we introduced in Chapter 3 but now test it in the ASD population. Findings that show reduced reward processing in the ASD group compared to neurotypicals would lend support to the social motivation hypothesis which explains the social dysfunction in people with autism by attributing it to a deficit in reward processing and motivation specifically for social stimuli.

Together these two studies shed light on the unique aspects of social reward processing and the social dysfunction found in people with autism.

4.2 Charitable donation task

4.2.1 Introduction

People with autism spectrum disorder (ASD) show behaviors suggesting abnormal preferences for stimuli. For instance, certain sensory stimuli or unfamiliar situations appear to be highly aversive, whereas other stimuli and familiar or repetitive situations appear to be desired; often, idiosyncratic objects can elicit abnormal attention and interest (Klin, Danovitch, Merz & Volkmar, 2007; Sasson, Turner-Brown, Holtzclaw, Lam & Bodfish, 2008). Together with these sometimes exaggerated preferences restricted to a specific set of unusual stimuli, there is a reduction in preferences for other people (Sasson et al., 2008; Dawson et al., 1998). These findings have motivated the hypothesis that ASD involves a domain-specific impairment for the valuation of social stimuli (Frith, 2001; Baron-Cohen, 1997). Nevertheless, the extent to which these impairments in ASD are confined to the domain of social processing remains an open question.

In this study, we addressed this open question by investigating how the preferences of participants with ASD compare to those of matched controls in a real charitable donation task. We chose a large number (N = 25) of charities benefitting people (for example, American Red Cross), but also nine charities that would benefit mental health (for example, Autism Research Institute), ten charities benefitting animals (for example, African Wildlife Foundation), and six charities benefitting the environment (for example, Heal the Bay). Participants were given pictorial and descriptive information about each charity, asked to choose an amount to donate to that charity, and asked to rate

each charity on a number of attributes. The charitable task is an interesting framework with which to address this issue because it involves the valuation of complex stimuli in more naturalistic behavioral settings than those used in previous experiments (Lin et al., 2011; O'Doherty et al., 2003; Chib et al. 2009).

4.2.2 Methods

Subjects. We recruited 16 high-functioning adults with a Diagnostic and Statistical Manual, Fourth Edition diagnosis of autism or Asperger's syndrome (four female) and 16 age- and education-matched controls (three female; see Table 1 for details). All participants with ASD met cutoff scores for autism or Asperger syndrome on the Autism Diagnostic Observation Schedule (ADOS) Module 4 (Lord, et al., 2000), and 13 out of 13 subjects assessed also met criteria on the Autism Diagnostic Interview-Revised (ADI-R) (Lord, Rutter & Le Couteur, 1994 All participants had an intelligence quotient (IQ) in the normal range, as assessed with the Wechsler Adult Intelligence Scale (Wechsler, 1981) and gave informed consent to participate in the studies under a protocol approved by the Institutional Review Board of the California Institute of Technology.

Table 4.1 Summary of demographic and background information about the participants.

	n	**Gender**	**Age**	**Full-scale IQ[a]**	**Education (years)**	**IRI[b] (EC + PT)**
With ASD	16	12 males 4 females	31.4 (12.3) [19-57]	110 (12.7) [93-133]	15.8 (2.1) [9-18]	24 (11.7) [6-42]
Matched controls	16	13 males 3 females	31.1 (12.7) [19-56]	114 (13.6) [94-133]	16.1 (1.4) [13-18]	37 (5.4) [27-43]
	ADI	**ADOS**	**SRS**			
With ASD	45 (10.5) [27-61]	17 (5.7) [11-25]	91 (26) [43-126]			

[a]The full-scale IQ from the Wechsler Adults Intelligence Test [18]; [b]IRI is the sum of the empathic concern and perspective taking sub-scores from the Davis Interpersonal Reactivity Index [19]. ADI: Autism Diagnostic Interview; ADOS: Autism Diagnostic Observation Schedule; ASD: autism spectrum disorder; SRS: Social Responsiveness Scale. Data are presented as the mean with the standard error in parentheses and the range in brackets below

Experimental Tasks. Subjects participated in the following three sessions in fixed order.

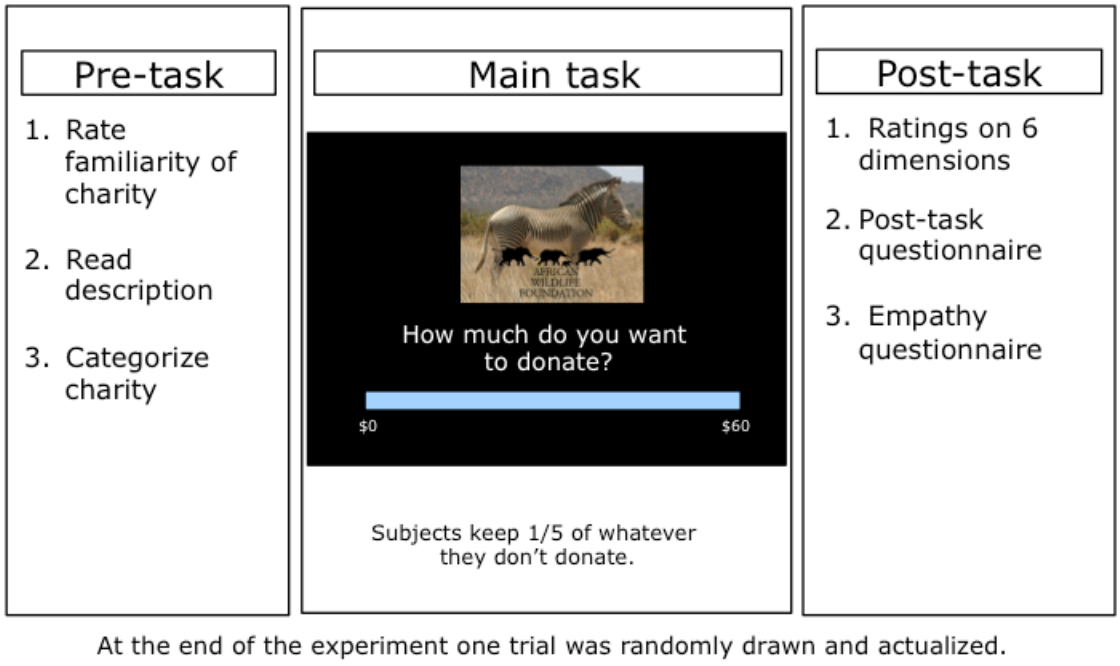
In the first session, they were familiarized with all the charities through a series of simple tasks. First, they were asked to indicate how familiar they were with the charity name. Next, participants were presented with a picture and asked to read a description of the charity's mission. Last, they were asked to place the charity in the best category among the following choices: environment, animal, people, and mental health. While participants were encouraged to provide single assignments, dual categories were allowed in exceptions (for example, charities like Canine Assistants that benefited both animal and people).

Participants' classifications were not used to derive the assignment of charities to categories used in our analyses, but rather as a check to our pre-assigned categorizations. Participant classifications were nearly identical to ours across all categories and none of the results presented below differ significantly if we use participant categorization of the charities. We assigned charity categories by using a filtering method. If the charities included mention of animals, the environment or mental illness, they were classified in their respective category; otherwise they were labeled a people charity. The list of charities used and their categorizations are presented in Additional file 1: Table S1.

In the second session, participants performed the charitable donation task. On every trial, the participant chose how much of $60 they wanted to donate to the charity presented (Figure 1). Participants kept 20 % of whatever amount they chose not to donate. Participants made donation choices for all 50 charities, one at a time, in randomized order. They were told that at the end of the experiment one of their actual

choice trials would be randomly selected and implemented, at which point an actual donation would be made to the selected charity and they would keep any remaining cash. Note that because only one trial was selected to count, the participants could treat each decision as being the only decision made, and did not have to worry about spreading their money across the different charities.

Figure 4.1 Schematic of the donation task. Participants carried out three sessions: first, they were presented with a picture and description of the charity in question, then they decided on their donation (one charity at a time), and finally they provided evaluations of the charity descriptions and pictures through explicit ratings.



In the third and final session, after the donation task, participants rated questions that measured how much the charity would benefit them (e.g., "How much do you think a $500,000 donation to this charity would help you personally?"), close friends and family, other people, and the world. They also rated the impact of the picture and descriptions they had been given for each of the 50 charities in terms of how effective they felt the charities were in promoting donations ("To what extent does the charity picture/description increase your willingness to give?"). This session thus provided us with an inventory of explicit knowledge about and evaluations of the charities. The complete set of questions asked is provided in Appendix 2.

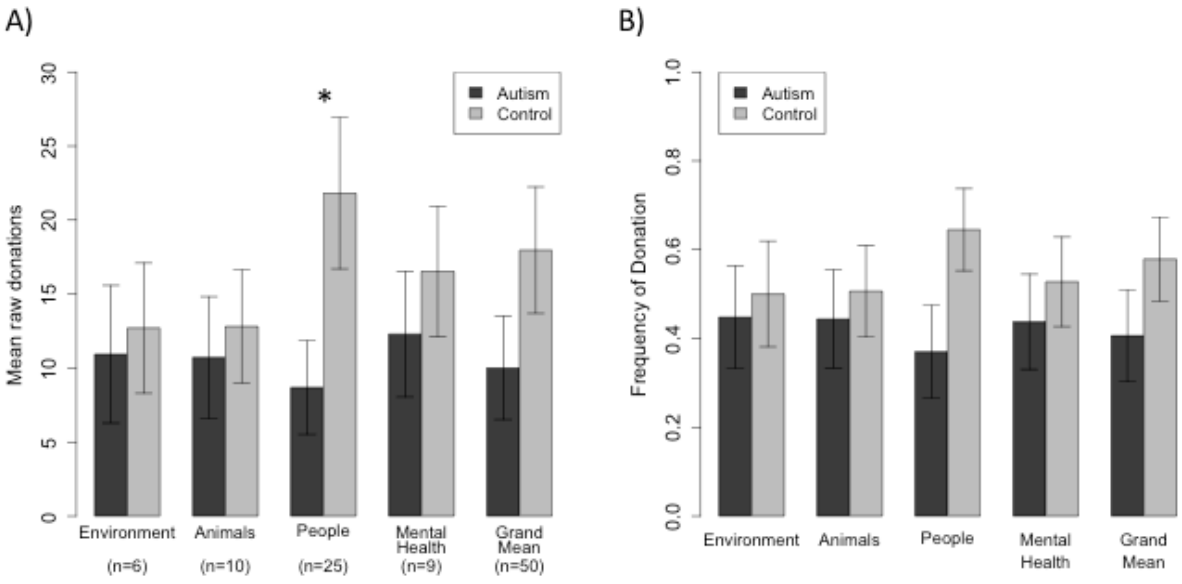
After completing the above sessions, subjects also completed the Interpersonal Reactivity Index (Davis, 1983) personality questionnaire, which measures an individual's dispositional empathy, and a post-task questionnaire that collected demographic background information and free-response questions about their motivations to give.

All significant values reported are two-tailed unless stated otherwise.

4.2.3 Results

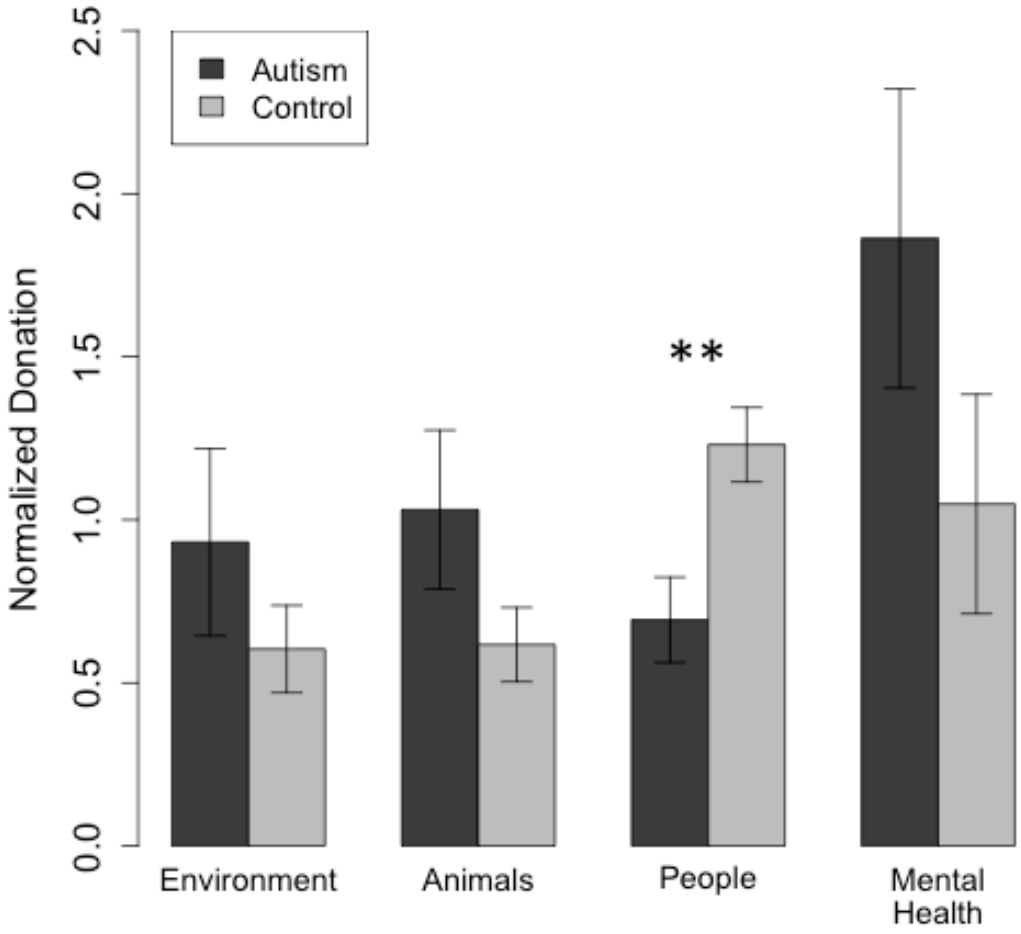
We first tested the hypothesis that the group of participants with ASD would donate less to people charities. Compared to the control group, the group with ASD donated less often to people charities (37 % versus 65 %; $t(30) = -1.97$, $p < 0.03$, one-tailed) and their mean donations to people charities were lower ($8.69 versus $21.82; $t(25) = -2.18$, $p < 0.02$, one-tailed), but so was the frequency of donations and mean donations to all charities on average, although this effect did not reach significance ($10.01 versus $17.97; $t(29) = -1.44$, $p < 0.16$, Figure 2A,B). Even when excluding any zero donations to a charity, mean donations across all charities from the group with ASD were lower, although again this group difference was not significant ($17.04 versus $28.11, $t(23) = 1.89$, $p = 0.07$).

Figure 4.2 Mean and frequency of donations across all four categories (A) Raw donations (mean and standard error of the mean (SEM); not normalized), for the four charity categories, as well as across all charities (Grand Mean). **(B)** Probability of donating to a charity in a particular category, means and SEM. Shown is the probability of making any donation, regardless of its magnitude. *$P < 0.05$



To account better for differences in mean donations between individuals within a group, we normalized each participant's donation by the mean number of dollars he or she donated in the experiment. This revealed a specific abnormality in mean normalized donations specific to the people charities (Figure 3; $t(28) = -3.10$, $p < 0.002$; all other charity categories not significant). A similar result was obtained for median donations per category ($t(24) = -2.34$, $p < 0.02$).

Figure 4.3 Normalized mean donations (mean and SEM), shown for the 4 charity categories. Donation amounts were divided for each participant by that participant's mean donation across all charities. This revealed a disproportionately lower amount donated to people charities than to any other category of charity. **p<0.01

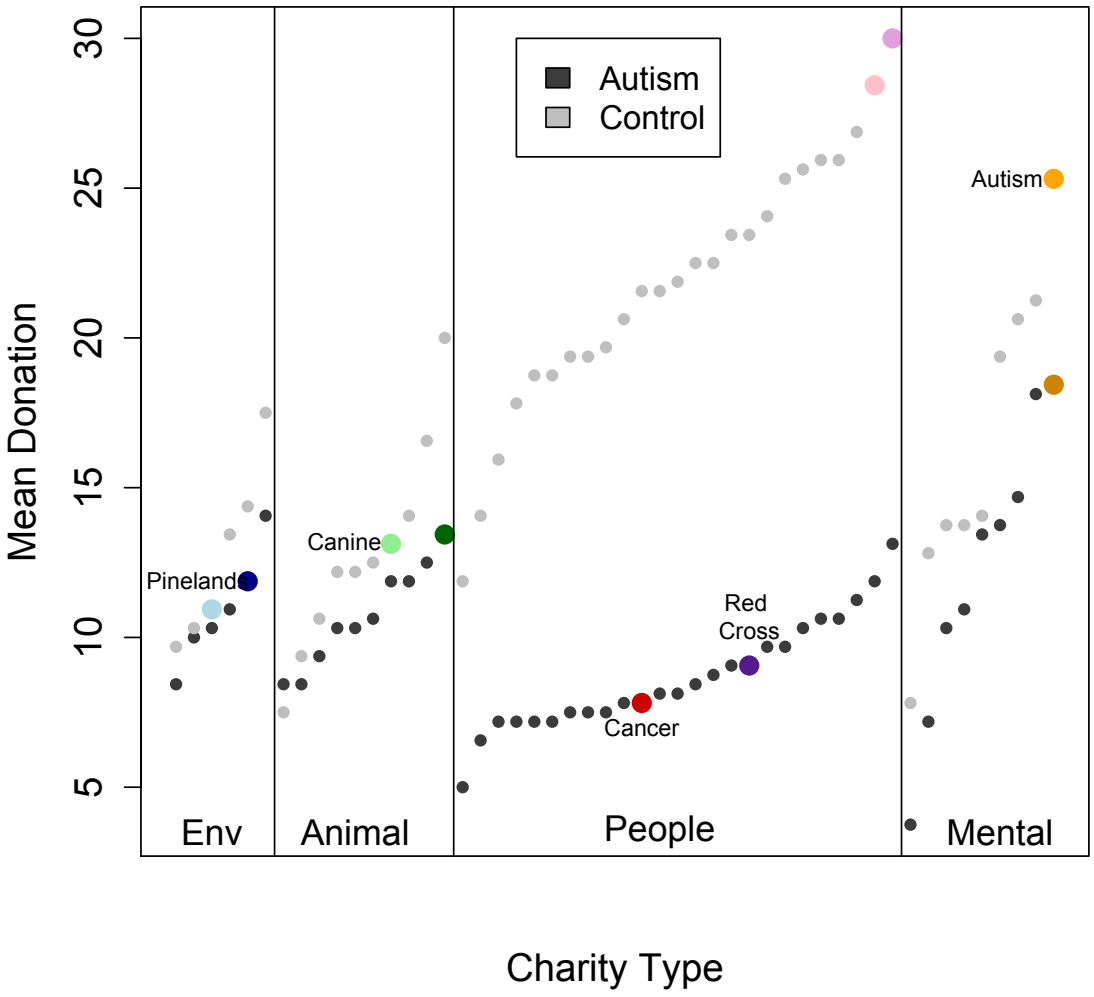


While our hypothesis specifically concerned social preferences, we also carried out a confirmatory mixed analysis of variance (ANOVA) with two levels of group (ASD, control) and two levels of charity category (people, other). This revealed a significant interaction between group and category ($F(1,1) = 8.3094$, $p < 0.005$) and no significant main effects of category or group. Post-hoc t-tests showed that this result was driven by the significant difference between ASD and controls normalized donations to people charities mentioned above. We verified these results with a resampling permutation test. We generated 10,000 random permutation samples and found that fewer than 2 % of resampled differences in mean donation to people charities were higher than what was observed in our data set. In contrast, none of the other charity categories were close to statistical significance (Environment: $p<0.39$, Animal: $p<0.36$, Mental: $p<0.25$; one-tailed).

We next examined individual charities, rank-ordering them by the mean donations within each category separately for each group (Figure 4). This analysis showed two components to the abnormal donations from the group with ASD. First, it confirmed that the group with ASD donated disproportionately less to the people charities. Second, it revealed a lack of discrimination amongst the people charities: whereas both the ASD and control groups showed a similar spread in donations across individual charities within each category, this was notably absent for the group with ASD in the case of the people charities. An exploratory analysis showed that the slope of a linear regression estimated through the people charity donation points was lower for the group with ASD ($m = 0.24$) than control group ($m = 0.58$). A few charities stood out as particularly preferred by the group with ASD. All of these fell into the animal or environment

category. Two of these in particular, Canine Assistants and Pineland Preservation Alliance, were remarkable because more than half of the participants with ASD donated to these (whereas most charities only elicited five or six donations from those with ASD).

Figure 4.4 Mean donations to individual charities, rank-ordered by the donations given by each participant group. Charities indicated by colored data points correspond to those where the ASD group showed particularly large differences in their donations compared with donations to the same charity by the control group. ASD donations are indicated in solid colors and control donations in fainter colors. "Pinelands": Pinelands Preservation Alliance (an environmental charity); "Canine": Canine Assistants (an animal charity); "cancer": National Childhood Cancer Foundation, and "Red Cross": American Red Cross (both people charities); "autism": Autism Research Institute (a mental health charity)



Across all charities, both groups generally gave very similar explicit ratings (Figure 4.5). However, in the people category, the control group gave significantly higher

ratings of impact both to the pictures and the narrative associated with the people charities, a pattern not seen for the descriptions of any of the other categories of (

Figure 4.6). Specifically, we found a significant group difference for the impact of the picture (ASD:2.0 vs. NT: 2.7; $t(27) = 2.72$, $p<0.01$) and narrative (ASD: 2.4 vs. NT: 3.2; $t(23) = -2.59$, $p<0.02$) associated with people charities.

Figure 4.5 Ratings given to the charities. Mean (and SEM) explicit ratings given to the charities, after all donations had been made. See Methods and Appendix 2 for detailed description of the ratings.

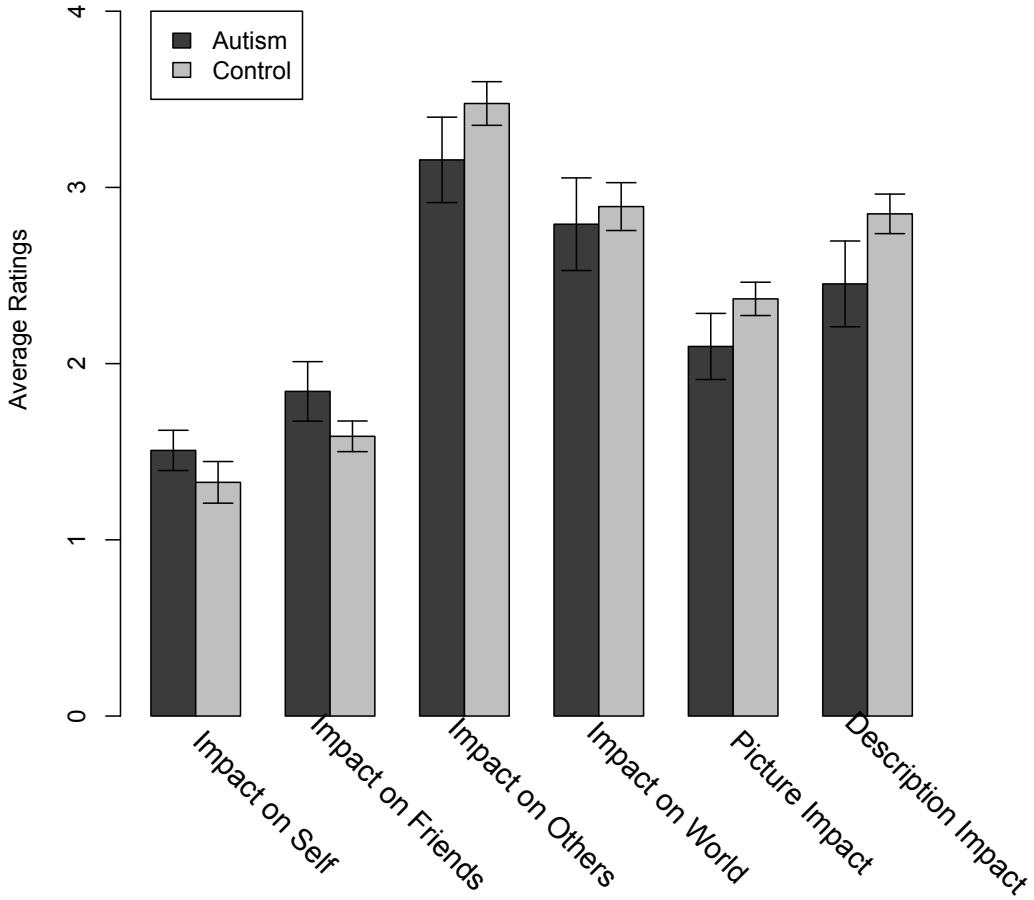
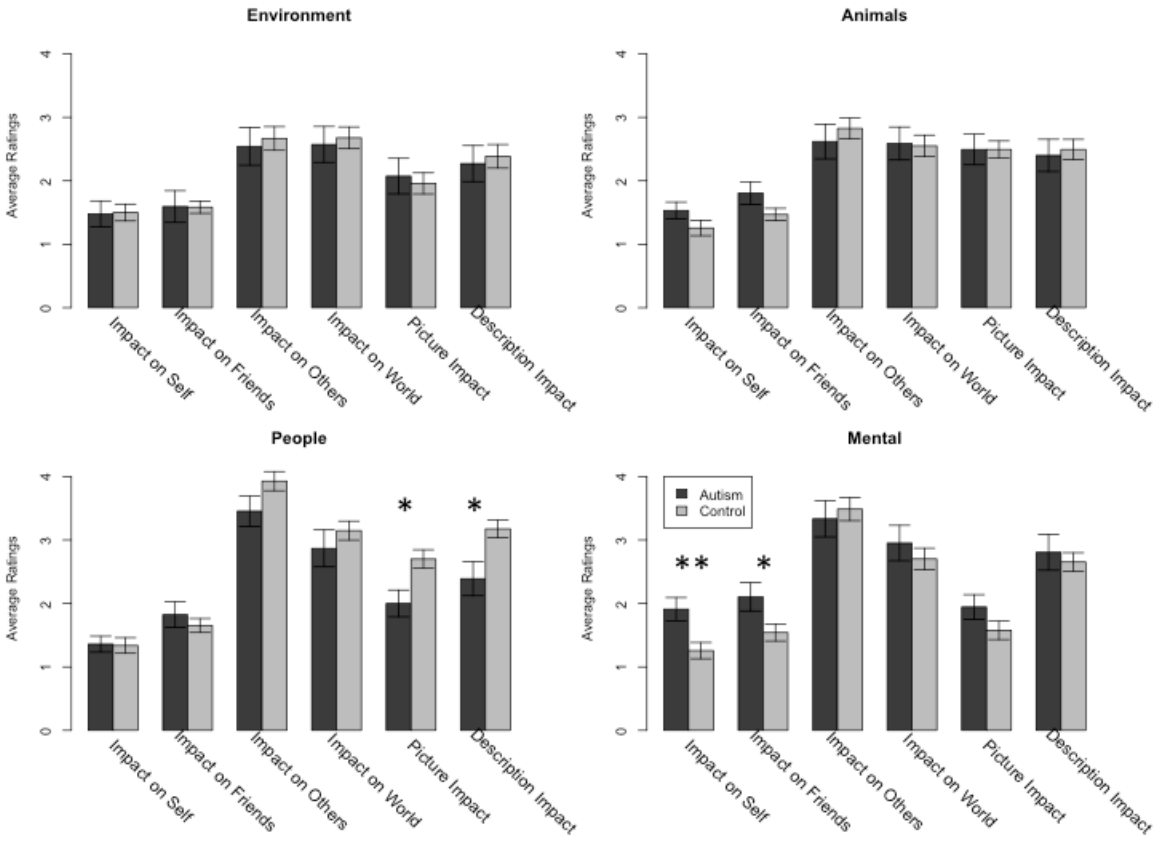
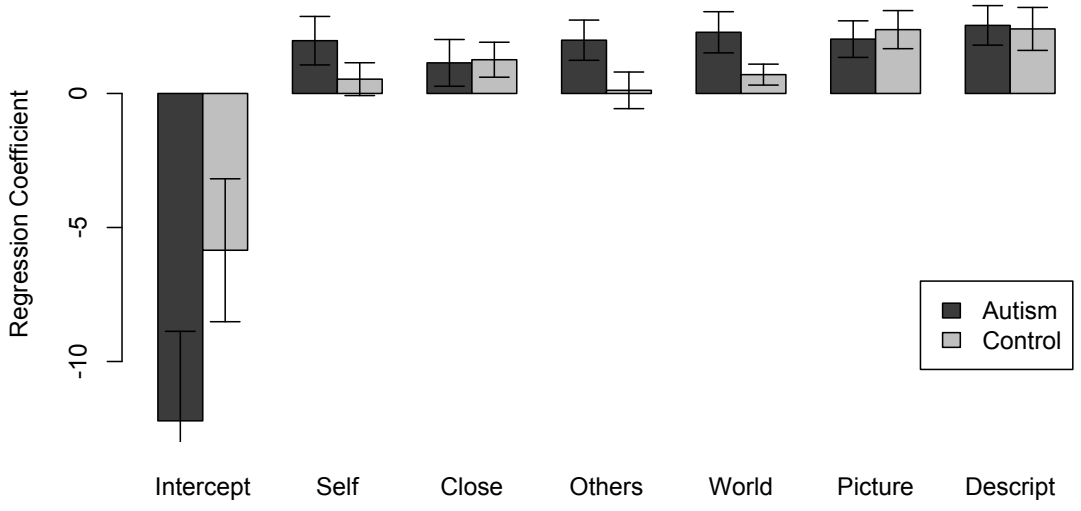


Figure 4.6 Ratings broken down by charity category. The ASD group gave significantly lower ratings to the impact of the picture and description just for the people charities. *p<0.05, **p<0.01



Regressing ratings onto donation on an individual-by-individual basis resulted in no statistically significant differences between the group means of the regression coefficients (Figure 4.7). This suggests that while both explicit ratings of the people charities as well as the donations made to them were abnormally low in the group with ASD, the link between evaluations of the descriptions of the charities and donation behavior was unaltered. In the mental health category, participants with ASD gave significantly higher ratings for impact on self (1.9 vs. 1.3; $t(27) = 2.92$, $p<0.007$) and friends (2.1 vs 1.5; $t(25) = 2.17$, $p<0.04$).

Figure 4.7 Regressions: Group mean regression coefficients. We carried out regressions of subjects' ratings onto their donations individually for each participant. There were no significant differences between groups on any of the regressions.



Finally, we carried out exploratory correlations across participants between their mean donation to people charities and several questionnaire-based and diagnostic measures. We did not find any meaningful correlations between mean donation to people charities and age, IQ, income, or the perspective taking and empathic concern scale of the IRI. However, there was a negative correlation ($r = -0.33$) between the ADOS-B subscale (reciprocal social interactions) and mean donation to people charities.

4.2.4 Discussion

Using a simple charitable donation task, we tested the hypothesis that people with ASD would show reduced social preferences. We found a significant reduction in the frequency and magnitude of donations made to charities benefitting other people compared with those benefitting mental health, animals or the environment. In addition, the group with ASD was less sensitive to specific information that discriminated amongst people charities, donating the same (abnormally low) amount to all of them. Control participants rated the impact of pictures and text descriptions on their donation amount particularly highly for people charities, whereas those with ASD gave significantly lower ratings to their impacts. This suggests that higher donations to people charities may normally be driven by the high social salience that they have, a component that is lacking in people with ASD. Taken together, this pattern of findings supports the hypothesis of abnormal social preferences in ASD and suggests specific reasons for it. The abnormally low ratings of the impact of visual and descriptive information provided for each charity given by the group with ASD argues that socially relevant empathy-evoking information was not incorporated into normal valuation for the charity. Consequently, there was little discrimination among the people charities, and the entire category of charities benefitting people was devalued in terms of the actual donations made. While ratings given by people with ASD for the impact of pictures on donations was low for people charities, we did find the group with ASD rated the impact of pictures as high as the control group for animal charities. This is interesting to note because studies have reported people with autism having an easier time connecting with animals than with people.

Across studies, the specific processes and neural structures that have been found abnormal in reward processing in autism are always a subset of those now well-documented to process the value of stimuli, actions, and outcomes in healthy participants. These include regions such as the ventral striatum, as well as ventral and medial parts of the prefrontal cortex (Rangel et al., 2008; Wallis, 2007; Rangel & Hare, 2010), and there is good evidence by now that these regions process reward value from all different types of stimuli (such as money, juice, or social stimuli), conveyed to these regions through convergent inputs from various sensory association cortices (Lin et al., 2011; Harris et al., 2011; Grabenhorst et al., 2010; Izuma et al., 2008; O'Doherty et al., 2002, 2003; Chib et al., 2009; Janowski et al., 2012). In particular, there is evidence that additional processing is required in order to interpret the value of socially relevant stimuli, originating in part from regions known to process social information, such as cortices in the superior temporal gyrus (Hare et al., 2010).

Impairments in such additional processing of socially relevant stimuli have been reported in high-functioning people with autism. One study found a remarkably selective impairment in combining outcomes with intentions to evaluate moral actions as good or bad in high-functioning people with autism (Moran et al., 2011), suggesting that the ability to incorporate multiple sources of social information is particularly compromised. Another study in the Adolphs lab reported that people with autism do not show the normal modulation of pro-social behavior (donations to a charity) when they are observed by another person, suggesting that they are insensitive to social reputation effects (Izuma et al., 2011). Yet even here there was a concomitant more general impairment: Izuma et al., (2011) found that people with autism were insensitive to social reputation effects on

charitable donations, and they also observed that overall donations were considerably less than in the control group.

In our present study, we found a similar effect: people with ASD donated less on average, across all stimuli, but in addition to this general difference they also showed a disproportionate reduction in donations specifically to charities benefitting other people. One caveat worth mentioning here is while there was no explicit monitoring in our study, as in the Izuma study, we concede that subjects could have been thinking about the analysis at the end of the experiment and how in principle we could trace who gave to what and how much. This could have created an observer effect that would partly explain the lower average donation amount in people with autism compared with controls. One could also argue that social reputation concerns might disproportionately weigh on people charities in controls. It may be that the often-present requirement to integrate multiple sources of complex information in order to synthesize a single reward representation is particularly acute for social stimuli, and accounts for a good part of the basis for the impairment seen in people with ASD when they process social rewards.

An ANOVA comparing non-people (collapsing animal and environment charities) versus people (collapsing people and mental health charities) also showed no significant interaction effects and only a main effect of non-people versus people. This suggests that people with autism treat charities in the mental health category (specifically those benefiting autism) in a special manner, different from their usual donation pattern for other people charities. Indeed the group with ASD gave these charities higher ratings for ‘benefit to self’ and ‘benefit to friends’ than did the control group, as shown in Figure 4.6. One interpretation of this pattern in the results could be that thinking about charities

benefiting people in general requires some empathy. For the control group, this may be one factor driving their donations to the people charities; for the group with ASD, it may be one lacking factor accounting for their low donations to people charities. In the mental health category, however, empathy may not have been required for the participants with ASD to recognize the value, since several of these charities were closely related to their own condition.

The phenotype of ASD shows a complex pattern of impairments, typically diagnosed as falling into three classes that together constitute the criteria for clinical diagnosis: language development, reciprocal social interactions, and repetitive behaviors and restricted interests. Arguably, the present findings may contribute to both of the last two, in that they suggest that people with autism have reduced interests in, or preferences for, charities benefitting people as compared to charities benefitting other categories. Moreover, we found a few charities that elicited unusually high donations from the group with ASD, a finding that should be followed up in future studies to better understand what it is about these particular charities that makes them preferable to people with autism. It is also interesting that we found a negative correlation between the amounts that participants with ASD donated to the people charities and the ADOS-B subscale. This subscale comprises items assessing unusual eye contact, facial expression directed to others, empathy and comments on others' emotions, responsibility, quality of social overtures, quality of social response, and amount of reciprocal social communication. While exploratory, this finding provides preliminary evidence that the abnormal social preferences revealed in our task may relate to abnormal social interactions in people with autism.

References

Baron-Cohen, S. (1997). Mindblindness: An Essay on Autism and Theory of Mind. Cambridge, MA: MIT Press.

Chib, V.S., Rangel, A., Shimojo, S., O'Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. The Journal of Neuroscience 29, 12315-12320.

Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology 44,113-126.

Dawson, G., Carver, L., Meltzoff, A.N., Panagiotides, H., McPartland, J., Webb, B. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. Child Dev 73,700-717.

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. Journal of Autism and Developmental Disorders 28,479-485.

Frith, C.D. (2007). The social brain? Philos Trans R Soc Lond B Biol Sci 362, 671-678.

Frith, U. (2001). Mind blindness and the brain in autism. Neuron 32, 969-979.

Grabenhorst, F., D'Souza, A.A., Parris, B.A., Rolls, E.T., Passingham, R.E. (2010). A common neural scale for the subjective pleasantness of different primary rewards. Neuroimage 51, 1265-1274.

Grelotti, D.J., Gauthier, I., Schultz, R.T. (2002). Social interest and the development of cortical face specialization: what autism teaches us about face processing. Developmental Psychobiology 40, 213-225.

Grelotti, D.J., Klin. A., Gauthier, I., Skudlarski, P., Cohen, D.J., Gore, J.C. (2005). fMRI activation of the fusiform gyrus and amygdala to cartoon characters but not to faces in a boy with autism. Neuropsychologia 43, 373-385.

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, 583-590.

Harris, A., Adolphs, R., Camerer, C.F., Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. PLoS One. 6, e21074.

Izuma, K., Matsumoto, K., Camerer, C., Adolphs, R. (2011) Insensitivity to social reputation in autism. PNAS 108, 17302-17307.

Izuma, K. and Saito, D.N. l. (2008). Processing of social and monetary rewards in the human striatum. Neuron 58, 284-294.

Janowski, V., Camerer, C., Rangel, A. (2012). Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL. Social Cognitive and Affective Neuroscience; doi: 10.1093/scan/nsr086.

Klin, A., Danovitch, J.H., Merz, A.B., Volkmar, F. (2007). Circumscribed interests in higher-functioning individuals with autism spectrum disorder: an exploratory study. Research and practise for persons with severe disabilities 32, 89-100.

Lin, A., Adolphs, R., Rangel, A.. (2011). Social and monetary reward engage overlapping neural substrates. Social and Cognitive Affective Neuroscience; doi:10.1093/scan/nsr006.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders 30, 205-223.

Lord, C., Rutter, M., Le Couteur, A. (1994). Autism Diagnostic Interview- Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders 24, 659-685.

Moran, J.M., Young, L., Saxe, R., Lee, S.M., O'Young, D., Mavros, P., Gabrieli, J. (2011). Impaired theory of mind for moral judgement in high-functioning autism. PNAS 108, 2688-2692.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D.M., Dolan, R.J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia 41, 147–55.

O'Doherty, J.P., Deichmann, R., Critchley, H., Dolan, R.J. (2002). Neural responses during anticipation of a primary taste reward. Neuron 33, 815–26.

Rangel, A., Camerer, C., Montague, P.R.. (2008). A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience 9, 545–56.

Rangel, A., Hare, T. (2010). Neural computations associated with goal-directed choice. Current Opinion in Neurobiology 20, 262-70

Sasson, N., Turner-Brown, L.M., Holtzclaw, T., Lam, K.S., Bodfish, J. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. Autism Research 1, 31-42.

Scott-Van Zeeland, A., Dapretto, M. (2010). Reward processing in autism. Autism Research 3, 53-67.

Singer, T., Seymour, B. (2004). Empathy for pain involves the affective but not sensory components of pain." Science 303, 1157-1162.

Triesch, J., Teuscher, C., Deak, G.O., Carlson, E. (2006). Gaze following: why (not) learn it? Developmental Science 9, 125–147.

Wallis, J.D. (2007). Orbitofrontal cortex and its contribution to decision-making. Annual Review of Neuroscience 30, 31–56.

Wechsler, D.A. (1981). The Wechsler Adult Intelligence Scale -- Revised. New York: The Psychological Corporation.

Appendix 1: Complete list of charities

Charity Name	Category
Achievement Centers for Children	mental
African Wildlife Foundation	animal
American Bird Conservancy	animal
American Kennel Club	animal
American Red Cross	people
Animal Haven	animal
Animal Rescue	animal
Anxiety Disorders Association of America	mental
Autism Research Institute	mental
Blue Card	people
Brain Tumor Society	people
Camphill Village Kimberton Hills	mental
Canine Assistants	animal
CARE	people
Chicago Foundation for Women	people
Child Abuse Prevention Center	people
Direct Relief International	people
Dogs for the Deaf	animal
Fisher Center for Alzheimer's Research Foundation	people
Global Fund for Women	people
Heal the Bay	environmental
Horizons for Homeless Children	people
Infant Crisis Services	people
International Eye Foundation	people
International Rett Syndrome Foundation	mental
Make-A-Wish Foundation International	people
Mercy Medical Airlift	people
National Breast Cancer Foundation	people
National Center for Missing and Exploited Children	people
National Childhood Cancer Foundation	people
National Inclusion Project	mental
National Ovarian Cancer Coalition	people
National Spinal Cord Injury Association	people
National Transplant Assistance Fund	people
Oprah's Angel Network	people
Organization for Autism Research	mental
Parkinson's Disease Foundation	people

Pasadena Humane Society and SPCA	animal
Pine Tree Society	mental
Pinelands Preservation Alliance	environmental
Save the Children	people
Society for the Protection of New Hampshire Forests	environmental
Southeast Alaska Conservation Council	environmental
Spinal Bifida Association	people
The American Chestnut Foundation	environmental
The Children's Clinic	people
The Churchill School and Center	mental
The WILD Foundation	animal
Thoroughbred Retirement Foundation	animal
Upper Raritan Watershed Association	environmental

Appendix 2: Complete list of questions (ratings) asked

- How much do you think a $500,000 donation to this charity would help you personally?
- How much do you think a $500,000 donation to this charity would help others you care about?
- How much do you think a $500,000 donation to this charity would help other people?
- How much do you think a $500,000 donation to this charity would help the world?
- To what extent does the charity picture increase your willingness to give?
- To what extent does the charity description increase your willingness to give?

4.3 Neuroimaging social rewards

4.3.1 Introduction

Underlying the abnormal social behavior of autism may be a difference in neural reward processing of social stimuli specifically or a reflection of a more general deficit in stimulus-reward association. Answers to this question would provide important support for the social motivation hypothesis of autism, and provide a mechanism to explain the patent social dysfunction in everyday life that is a key component of the diagnosis. Are social difficulties derivative to general processing difficulties (perhaps because social stimuli are simply more complex and difficult to process), or is there evidence for a specific impairment in social processing with relative sparing of other domains of cognition? A couple of neuroimaging studies have looked at social vs. non-social reward processing in ASD but findings have not been entirely consistent (Dichter, Richey, Rittenberg, Sabatino, & Bodfish, 2012; Scott-Van Zeeland, Dapretto, Ghahremani, Poldrack, & Bookheimer, 2010).

We re-visit the instrumental reward learning task that contrasted learning with social rewards against learning with monetary reward described in Chapter 3. We showed then that there is overlap in social and non-social reward neural processing in neurotypicals. If there is a deficit in reward processing and motivation specifically for social rewards as the social motivation hypothesis suggest, we would expect to find differences between the two groups at a neural level in learning with social rewards but not monetary rewards.

I tested ten high-functioning people with ASD (7M, 3F) and ten healthy controls who were matched on gender, age, and education. BOLD-fMRI was collected on a 3T scanner while participants had to learn to choose among slot machines associated with differently valued outcomes. In the social version of the task, outcomes were smiling, neutral, or angry faces accompanied by matching sound effects (happy, neutral, or angry voices). In the monetary version, outcomes were variable gain or loss of money. The two tasks were structurally identical except for the type of reward, permitting direct comparisons.

4.3.2 Methods

	n=	Gender	Age	FSIQ	Education (in years)	SRS
ASD	10	7M3F	28 (3.1) [18-45]	113 (4.7) [93-133]	15 (0.7) [10-18]	76 (9.7) [24-113]
Matched Controls	10	7M3F	27 (3.1) [17-44]	114(13.4) [104-123]	15 (0.6) [12-18]	56 (15.6) [41-72]

	ADI-A	ADI-B	ADI-C	ADI-D	ADOS-A	ADOS-B	ADOS-C	ADOS-D
ASD	20 (1.8) [12-28]	17 (1.5) [11-24]	6 (0.9) [2-12]	3 (0.5) [0-5]	4.9 (0.5) [3-7]	10.2 (1.4) [4-17]	1 (0.2) [0-2]	1.4 (0.4) [0-3]

Table 4.2 Summary of demographic and background information about the participants. FSIQ is full-scale IQ from the Wechsler Adults Intelligence Test (Wechsler, 1981).

Participants. Twenty-seven subjects participated in the study (mean age = 22.4 years; range 18-28). Seven ASD subjects were excluded from the analyses: six due to failure to understand task instructions (e.g., slot machine choices were based on favorite colors) and one who rated the social stimuli abnormally. Behavioral analyses reported is based on 20 subjects: 10 subjects with ASD (3 female) and 10 age- and education-matched controls (3 female) (Table 4.2). One ASD subject and his matched control were dropped from the neuroimaging analysis because of excessive head movement. Neuroimaging results reported are based on the 18 remaining subjects. This did not materially change subject demographic composition (Table 4.3). All ASD participants met the Diagnostic and Statistical Manual of Mental Disorders, Revised 4th Edition diagnostic criteria for autism or Asperger's syndrome and met the cutoff scores for autism or Asperger's

syndrome on the Autism Diagnostic Observation Schedule, Module 4 (Lord et al., 2000) and Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994) (Table 4.2). All participants had normal or corrected-to-normal vision, had no history of psychiatric or neurological disease, and were not taking medications that might have interfered with BOLD-fMRI. Participants gave informed consent to participate in this study under a protocol approved by the Caltech IRB.

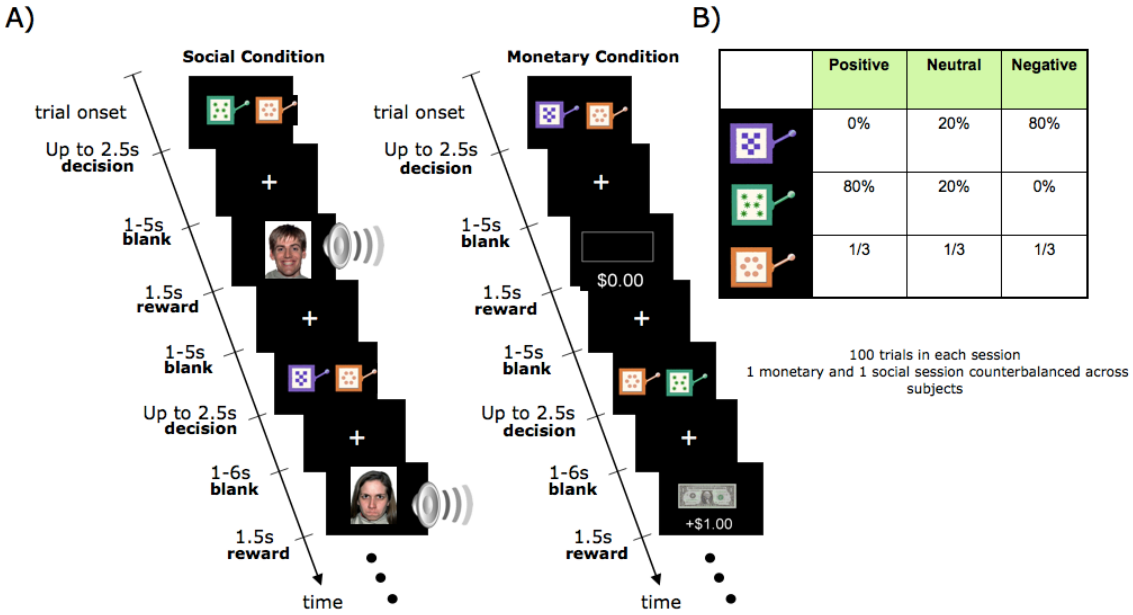
	n=	Gender	Age	FSIQ	Education (in years)	SRS
ASD	9	6M3F	26 (3.1) [18-45]	115 (4.4) [101-133]	15 (0.8) [10-18]	76 (9.7) [24-113]
Matched Controls	9	6M3F	26 (2.9) [17-44]	114(13.4) [104-123]	15 (0.6) [12-18]	56 (11.0) [41-72]

	ADI-A	ADI-B	ADI-C	ADI-D	ADOS-A	ADOS-B	ADOS-C	ADOS-D
ASD	21 (1.8) [12-28]	17 (1.7) [11-24]	6 (1.1) [2-12]	3 (0.6) [0-5]	4.8 (0.5) [3-7]	9.8 (1.5) [4-17]	0.9 (0.1) [0-1]	pa

Table 4.3 Summary of demographic and background information of only participants included in imaging analysis. FSIQ is full-scale IQ from the Wechsler Adults Intelligence Test (Wechsler, 1981).

Task. Participants played two structurally identical versions of an instrumental learning task, one with monetary rewards, the second with social rewards (Figure 4.8A). A trial began with the display of two visually distinctive slot machines, each associated with one of three outcome distributions: mean-positive, mean-negative, and mean-neutral (Figure 4.8B).

Figure 4.8 A) Timeline of the monetary and social reward trials. Choice trials paired a neutral slot machine with a valenced slot machine. Trials were identical except for the nature of the outcomes: Monetary trials had a gain/loss of +$1, $0, or -$1, whereas social trials revealed happy, neutral, or angry faces accompanied by sound effects of similar emotional valence. Specific slot machines were randomly assigned to specific reward outcomes at the start of the experiment for each subject, and distinct between monetary and social condition blocks. **B) Distribution of outcomes for each slot machine.** First row: negative machine. Second row: positive machine. Bottom row: neutral machine. The same distribution was used in the monetary and social conditions. Actual appearance of the slot machines was randomly paired with a reward outcome distribution, and distinct between monetary and social condition blocks.



All participants completed one social and one monetary block of 100 trials each; block order was randomized between participants. At the beginning of each trial, participants were shown a neutral slot machine paired with either the positive or negative slot machine (50/50 probability with randomized order). Participants chose one by pressing a left or right button. Up to 2.5 seconds were allowed for choice, followed by a uniformly blank screen displayed for 1-5 seconds (flat distribution), followed by the reward outcome displayed for 1.5 seconds, followed by an inter-trial interval of a uniformly blank screen displayed for 1-6 seconds (flat distribution). Note that participants were not told the reward probabilities associated with each slot machine and had to learn them by trial and error during the task.

Stimuli and Rewards. The slot machines in both conditions were represented by cartoon images of actual slot machines that varied in color and pattern (Figure 4.8). In the social condition, reward outcomes were color photographs of unfamiliar faces from the NimStim collection (Tottenham et al., 2009) showing either an angry (negative outcome), neutral (neutral outcome), or happy (positive outcome) emotional expression, presented together with emotionally matched words played through headphones (normalized for volume and duration). Examples of positive words are “excellent”, “bravo”, and “fantastic”. Examples of negative words are “stupid”, “moron”, and “wrong”. Examples of neutral words are “desk”, “paper”, and “stapler”. Extensive prior piloting had demonstrated the behavioral efficacy of these stimuli in reward learning.

In the monetary condition, the positive outcome was a gain of one dollar (an image of a dollar bill), the negative condition was a loss of one dollar (image of a dollar bill

crossed out), and the neutral condition involved no change in monetary payoff (image of an empty rectangle). Subjects were paid out the sum of their earnings at the end of the experiment.

Face ratings and other post-task activities. At the end of the experiment, we asked subjects to rate the pleasantness of each of the faces and matching sound effects. We were interested in whether the two groups experienced the stimuli similarly. Lastly, subjects also completed the Interpersonal Reactivity Index (Davis, 1983) personality questionnaire, which measures an individual's dispositional empathy.

Computational model. We computed trial- and subject-specific values for stimulus value (SV), prediction error (PE), and reward (R) consistent with the methods of our previous study (Lin et al., 2011). The stimulus value (SV) for every slot machine was calculated as the 10-trial moving average proportion of times that the machine was chosen when it was shown, a continuous value between 0 and1. Consistent with this coding, reward outcomes (R) were assigned a value of 1 if they were positive; a value of 0.5 if they were neutral, and a value of 0 if they were negative. Prediction errors (PE) at the time of outcome were calculated using a simple Rescorla-Wagner learning rule (Rescorla and Wagner, 1972) as the difference between the value of the reward outcome and the stimulus value of the machine selected for that trial: $PE_t = R_t - SV_t$.

Image acquisition. T2*-weighted gradient-echo echo-planar (EPI) images with BOLD contrast were collected on a Siemens 3T Trio. To optimize signal in the orbitofrontal

cortex (OFC), we acquired slices in an oblique orientation of 30o to the anterior commissure-posterior commissure line (Deichmann, Gottfried, Hutton, & Turner, 2003) and used an eight-channel phased array headcoil. Each volume comprised 32 slices. Data was collected in two sessions, ~15 min each. The imaging parameters were as follows: TR= 2 s, TE= 30 ms, FOV= 192 mm, 32 slices with 3mm thickness resulting in isotropic 3mm voxels. Whole-brain high-resolution T1-weighted structural scans (1 x 1 x 1 mm) were co-registered with their mean T2*-weighted images and averaged together to permit anatomical localization of the functional activations at the group level.

fMRI pre-processing. The imaging data was analyzed using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were corrected for slice acquisition time within each volume, motion-corrected with realignment to the last volume, spatially normalized to the standard Montreal Neurological Institute EPI template, and spatially smoothed using a Gaussian kernel with a full-width at half-maximum of 8mm. Intensity normalization and high-pass temporal filtering (filter width = 128s) were also applied to the data.

fMRI data analysis. The data analysis proceeded in three steps. First, we estimated a general linear model with AR(1). This model was designed to identify regions in which BOLD activity was parametrically related to SV, R, and PE. The model included the following regressors:

R1) An indicator function for the decision screen in free-choice monetary trials.

R2) An indicator function for the decision screen in free choice monetary trials multiplied

by the SV of the two slot machines shown in that trial (summed SV).

R3) An indicator function for the decision screen in free choice monetary trials multiplied by the reaction time for that trial.

R4)-R6) Analogous indicator functions for decision screen events in free choice social trials.

R7) An indicator function for the decision screen in forced monetary trials.

R8) An indicator function for the decision screen in forced monetary trials multiplied by the SV of the slot machine displayed.

R9)-R10) Analogous indicator functions for decision screen events in forced social trials.

R11) A delta function for the time of response in the monetary condition.

R12) A delta function for the time of response in the social condition.

R13) An indicator function for the outcome screen in free monetary trials (both choice and non-choice).

R14) An indicator function for the outcome screen in free monetary trials multiplied by the PE for the trial.

R15) An indicator function for the outcome screen in free monetary trials multiplied by the R for the trial.

R16)-R18) Analogous indicator functions for outcome screen events in free social trials (both choice and non-choice).

We orthogonalized the modulators for the main regressors that had more than one modulator (e.g., R2 and R3). The model also included six head-motion regressors, session constants, and missed trials as regressors of no interest. The regressors of interest and missed trial regressor were convolved with a canonical HRF.

Second, we calculated the following first-level single-subject contrasts: 1) R2 vs. baseline, 2) R5 vs. baseline, 3) R14 vs. baseline, 4) R15 vs. baseline, 5) R17 vs. baseline, and 6) R18 vs. baseline.

Third, we calculated second-level group average beta plots using a novel leave-one-out method to avoid concerns of non-independence selection (Vul, Harris, Winkielman, & Pashler, 2009). For each contrast, we extracted the mean signal from a group of voxels customized for each subject by our novel leave-one-out method. These voxels were selected through the following procedure:

1) Excluding the subject, second-level group contrasts for the rest of the subjects, from both this study and the study in Chapter 3 were computed
2) Voxels surviving a threshold of $p<0.005$ were intersected with an anatomical mask selected based on results from the study in Chapter 3. The anatomical mask for the vmPFC at choice was taken using a sphere of 10-mm radius defined around the peak activation coordinates that correlated with stimulus values in Rolls et al. (E. T. Rolls, McCabe, & Redoute, 2008b). The anatomical mask for the vmPFC at reward outcome was given by a sphere of 10-mm radius defined around the peak coordinates that correlated with the magnitude of reward outcome in O'Doherty et al. (J. P. O'Doherty, R. Deichmann, H. D. Critchley, & R. J. Dolan, 2002). The anatomical mask for the ventral striatum was taken using a sphere of 10-mm radius defined around the peak activation coordinates that correlated with prediction errors in Pessiglione et al. (Pessiglione et al., 2006).
3) The intersection of the second-level group contrast and the anatomical mask defined the ROI we extracted the mean signal for each subject.

4.3.3 Results

Behavioral Results

I compared a group of 10 high-functioning adults with ASD with 10 healthy controls matched on age, sex, and education (Table 4.2).

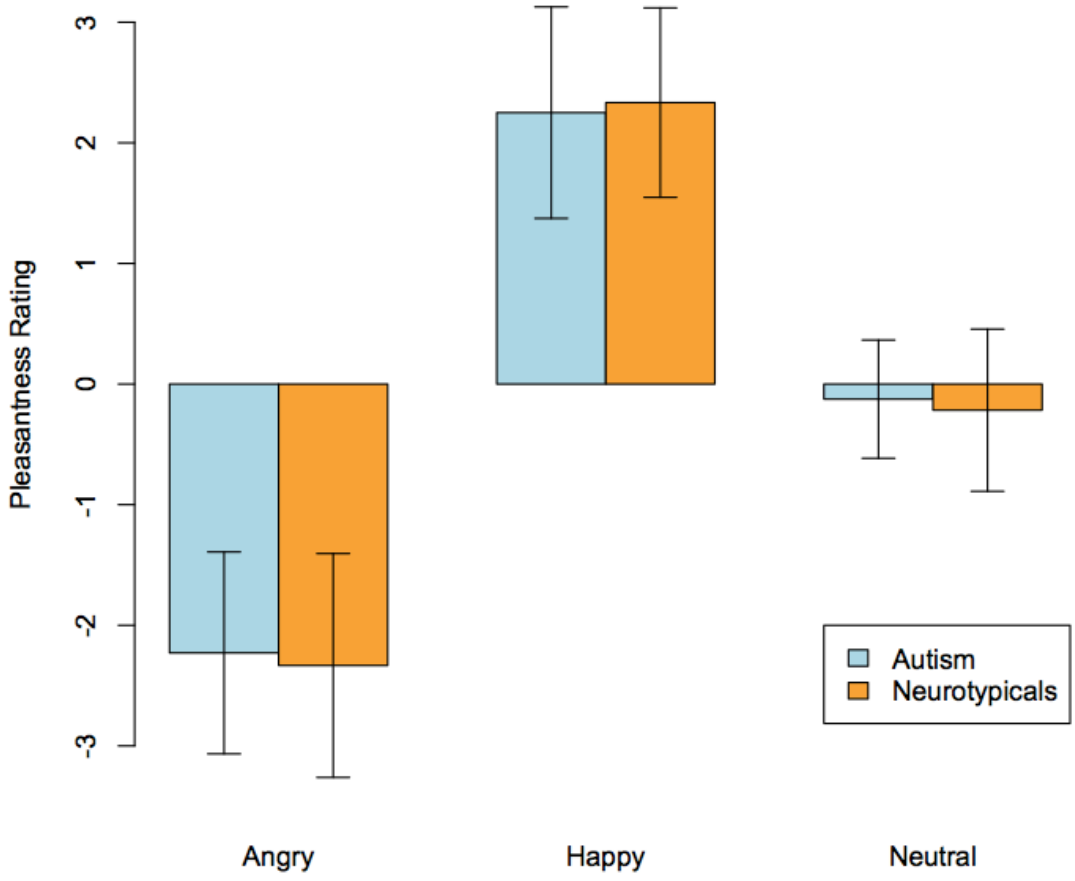
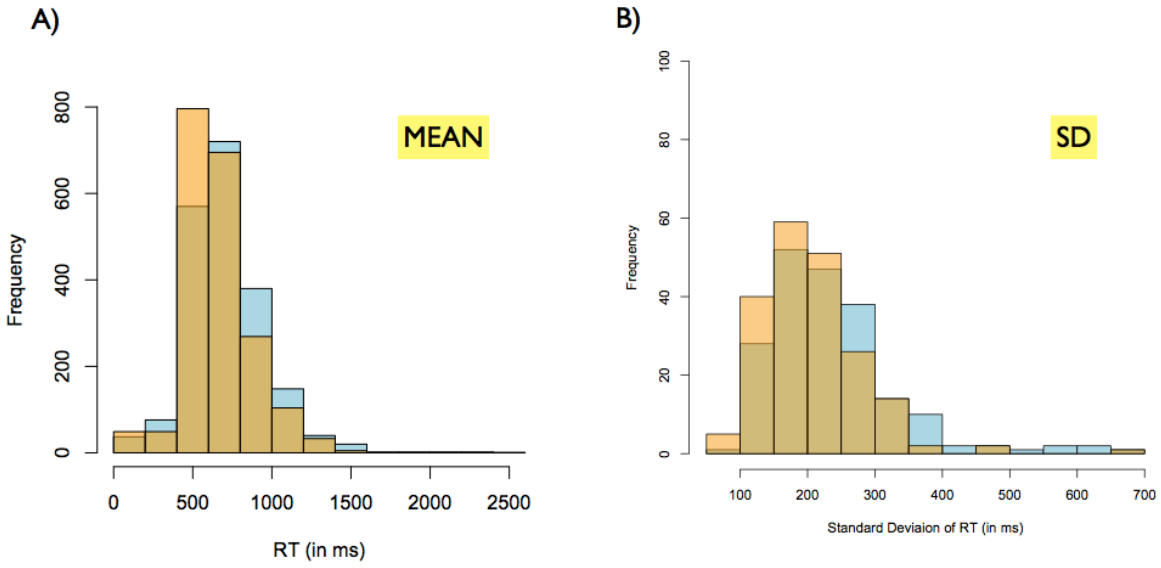


Figure 4.9 Pleasantness ratings of the happy, neutral, and angry social stimuli. There were no significant differences between groups on any of the categories.

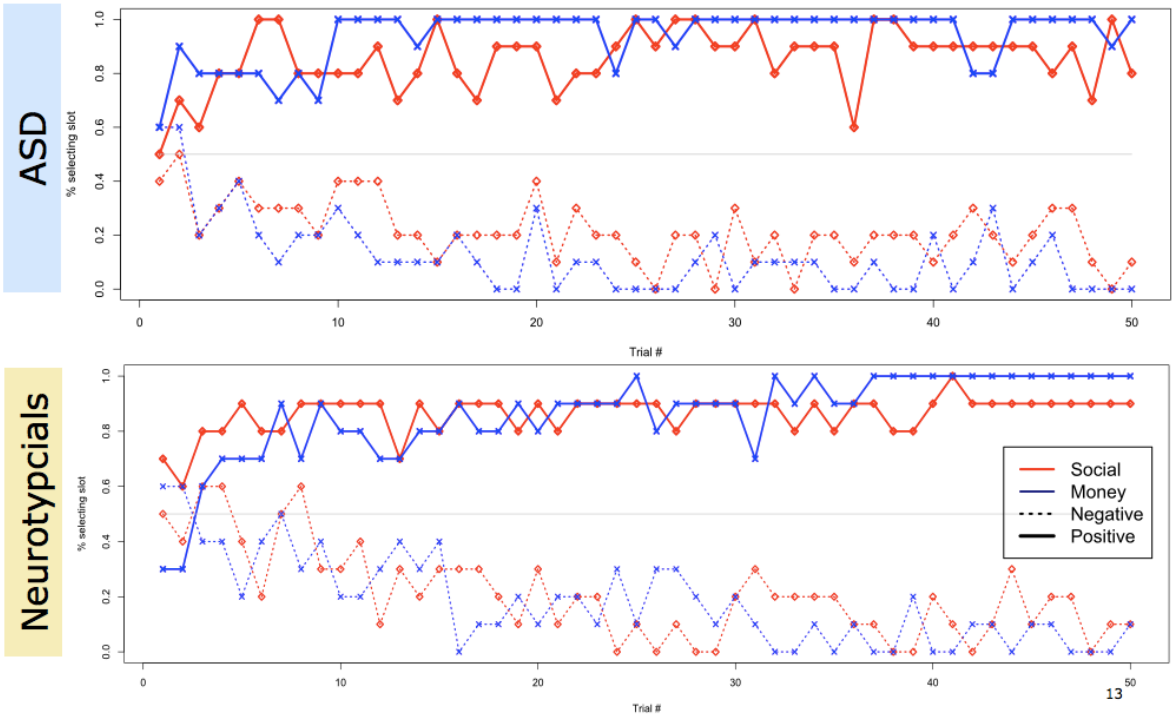
My first check was to confirm the two subject groups had similar subjective experiences of the face stimuli. Figure 4.9 plots the pleasantness ratings for angry, happy, or neutral social stimuli for each group. There were no statistically significant differences in any of the valence categories. Neither were there any meaningful differences in reaction times (Figure 4.10).

Figure 4.10 A) Distribution of mean reaction-times between ASD and NT B) Distribution of SD of reaction-times between ASD and NT



Turning to the choice data from the main task, I found that both groups reliably learned to select the slot machine associated with the highest probability of a positive-valenced outcome for both social and non-social rewards (Figure 4.11).

Figure 4.11 Plot of group subjects choices across trials. Both groups reliably learned to select the slot machine associated with the highest probability of a positive valenced outcome and avoid the slot machine associated with the highest probability of a negative valenced outcome in both monetary and social conditions.



I then plotted the cumulative number of optimal choices trial by trial. What was particularly interesting was that when social and monetary trials were collapsed, both ASD and NT lines lay essentially on top of one another (Figure 4.12). However, when I separated out social and monetary trials, I found a double dissociation: ASD subjects were better than NT on the monetary condition, but NT subjects were better than ASD on the social condition (Figure 4.13). A t-test for differences of the slopes of the best-fit line for cumulative positive trials revealed that slopes for the NT group were significantly higher than those for the ASD group in the social condition (ASD: 0.86 vs NT: 0.97;

$t(10)=-2.35$, $p<0.04$). The same analysis did not result in any significant differences between the groups on the negative trials for the social condition or any of the monetary conditions.

Figure 4.12 Plot of cumulative optimal responses across trials combining monetary and social trials

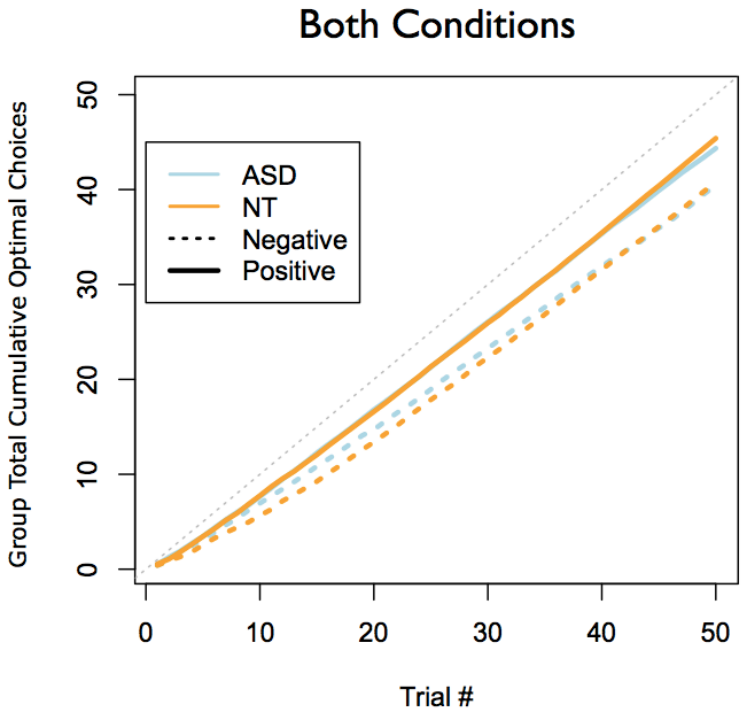


Figure 4.13 A) Plot of cumulative optimal responses across trials in monetary condition. B) Plot of cumulative optimal responses across trials in social condition

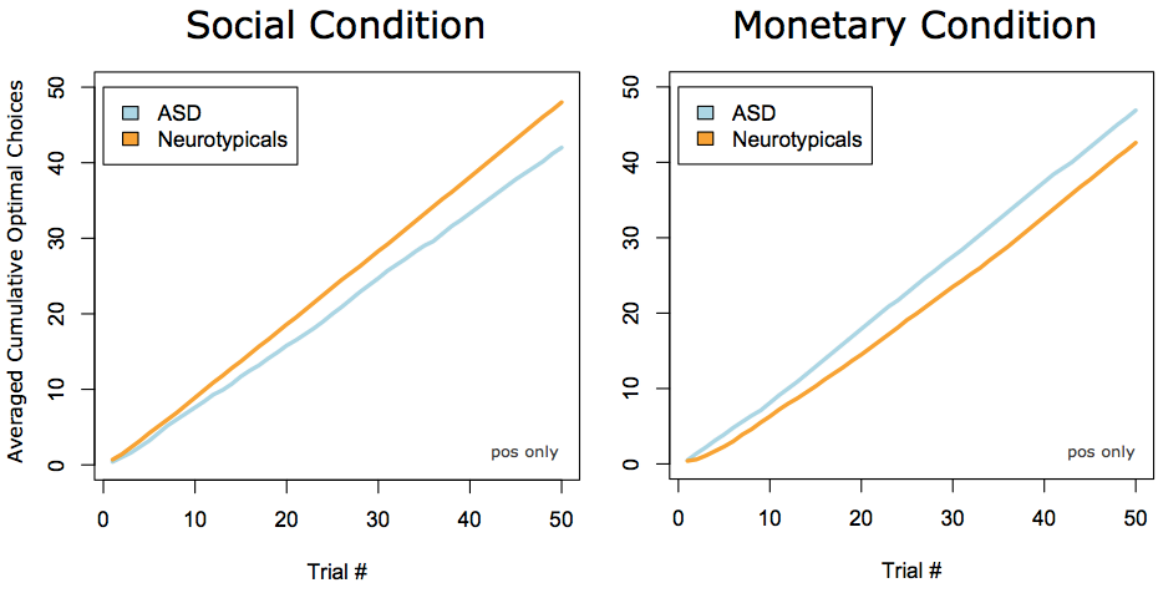
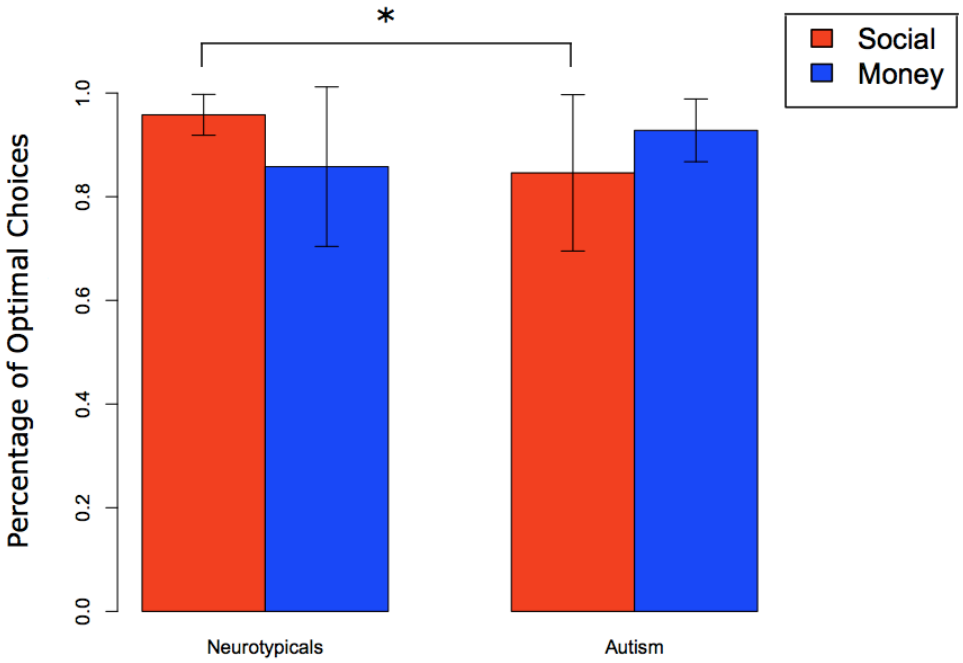


Figure 4.14 Total percentage of optimal slot machine selection (mean and SEM) for positive trials in social and monetary condition



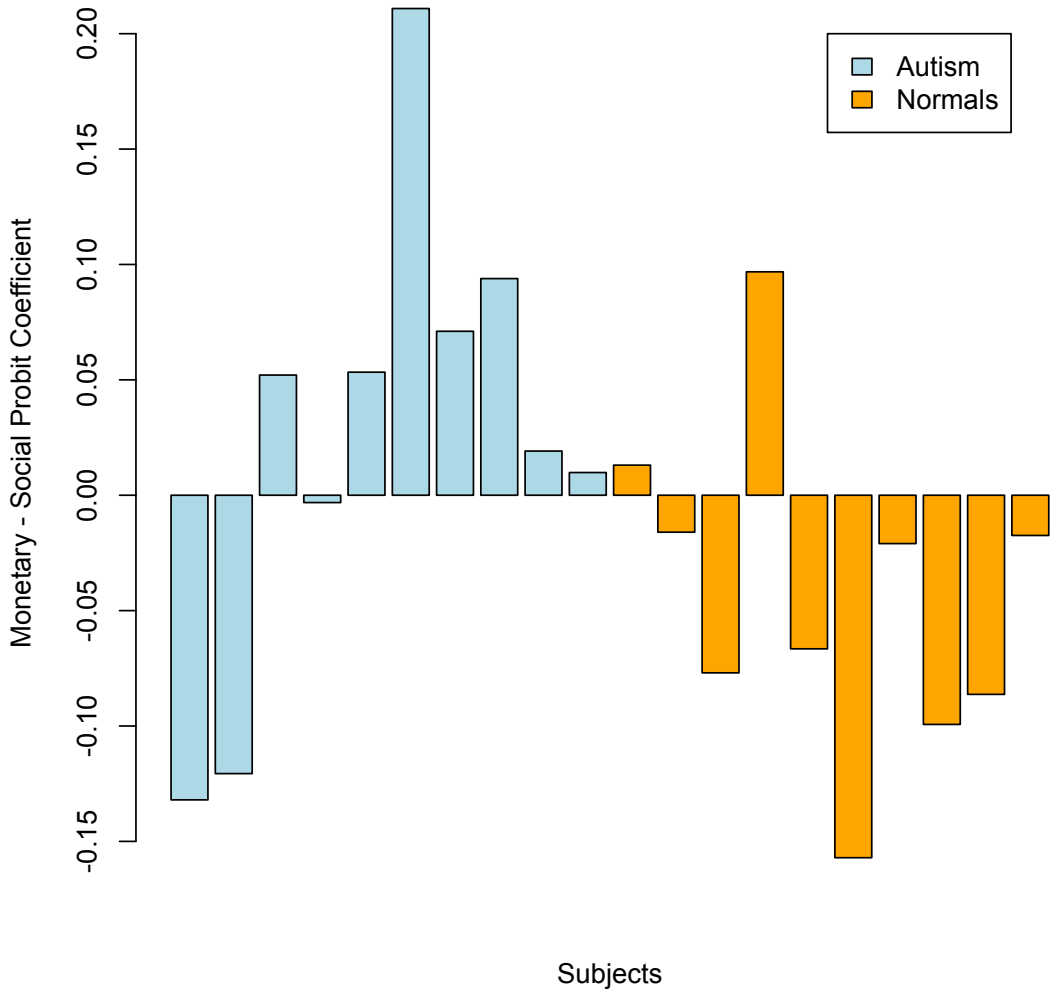
I found a similar result when I looked at total percentage of optimal slot machine selection at the end of the experiment (Figure 4.14). There was a significant difference between ASD and NT on positive trials in the social condition (ASD: 0.85 vs NT: 0.96; $t(10)= -2.27$, $p<0.05$) but not in any of the other conditions. A 2 by 2 ANOVA of subject group (ASD or NT, a between-subject factor) and condition (monetary and social) collapsing positive and negative trials (a within-subject factor) had no significant main effects of category or group (all F's <1, all p-values > .4) but revealed a significant group by condition interaction effect ($F(1,1)=4.20$, $p<0.05$).

Lastly, a qualitative look at individual subject data seemed to show differing rates of learning. To capture this in a quantitative manner, I modeled each subject's choice data with a probit regression. The slope estimate of the probit regression was thought to be a good metric for learning rate. I fit a probit regression through each subject's raw data appended with 10 alternating left and right trials at the beginning. I padded the start to give the model enough learning trials, since some subjects were able to identify the high value slot machine within 1 or 2 trials. Visual inspection confirmed this resulted in the best estimates.

I also checked whether the probit regression differentially modeled one group's data better than the other's by comparing Akaike Information Criteria (AIC) scores between the two subject groups. The answer in this case was no (all p-values >0.05). The probit regression did not fit NT data any better or worse than ASD data. After these checks, I felt confident that the slope estimate from each subject's probit regression was an accurate reflection of learning rate.

Figure 4.15 plots the difference between the probit slope coefficient for positive trials in the monetary and social condition for each subject. This showed an interesting group split. I quantified these findings with a Fischer's exact test on the distribution of greater monetary vs social probit slope coefficient in ASD and control subjects. I found there was a significant contingency between subject group (ASD vs. Controls) and a faster learning rate in the monetary over the social condition ($p = .032$). A greater proportion of participants in the ASD group, than participants in the control group, had learning rates that were faster for the monetary than the social condition.

Figure 4.15 Difference between monetary and social probit regression coefficients (positive trials only). We fit probit regressions to each subject's choices on the positive trials in each condition. We then plotted the difference between the fitted monetary and social coefficient for each subject.



Exploratory neuroimaging results

At the writing of this thesis, neuroimaging analysis for this project was still ongoing. I present very exploratory results here to give a flavor for the data.

Figure 4.16 shows the average beta plots between groups for each of the main contrasts. The first nice thing to notice is the replication of the results from study 1. In all contrasts the average beta plots are non-zero and positive for the NT group (all p-values <0.05).

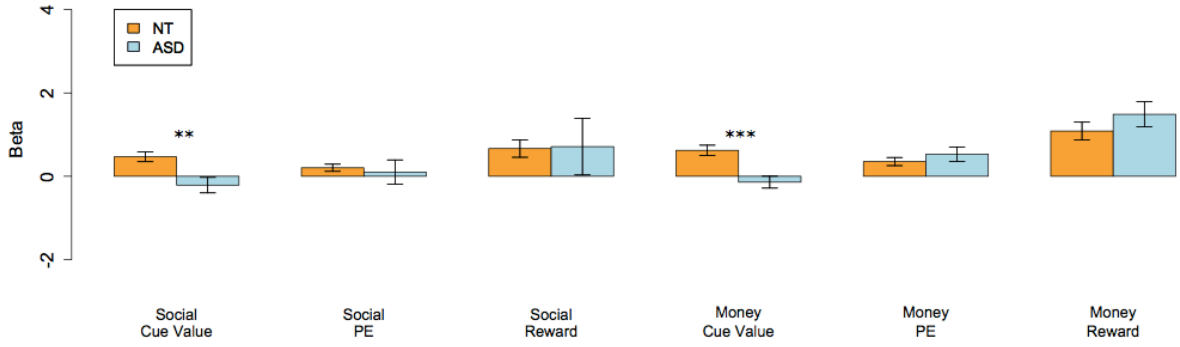


Figure 4.16 Average beta plots for activity during the 1) time of decision modulated by SV, 2) time of outcome modulated by PE, and 3) time of outcome modulated by reward in both social and monetary trials for both group types. The functional masks were given by the intersection of leave-one-out analysis described in the methods and anatomical masks vmPFC for SV and R and VStr for PE.

While ASD results are not as easy to interpret, they do show that activity for monetary PE (0.53, $t(8)=3.08$, $p<0.02$) and rewards (1.48, $t(8)=4.98$, $p<0.001$) at the time of outcome are significantly different from zero. And vmPFC encoding of SV at the time of decision is significantly lower than the NT group in both the social (ASD: -0.213 vs NT: 0.466; $t(15)=-3.09$, $p<0.01$) and monetary (ASD: -0.133 vs NT: 0.623; $t(21)=-3.98$, $p<0.001$) condition. There were no other significant differences between groups in the other contrasts.

Although not significant, I also found a trend of higher levels of activation in the monetary condition over the social condition in the ASD group. There were no significant differences between corresponding contrasts in the social and monetary conditions for the NT group.

Lastly, I found hippocampus activity modulated by SV at the time of decision in social trials in ASD subjects at an uncorrected threshold of $p<0.005$. I also found hippocampus activity at an uncorrected threshold of $p<0.005$ at the time of outcome modulated by PE in monetary trials for ASD subjects. There was no hippocampus activity in any of the contrasts for the NT group.

4.3.4 Discussion

In terms of overall ability behaviorally to discriminate positive from negative slot machines in their choice behavior, reaction-times, and valence ratings, both groups performed remarkably similarly. Both groups learned to choose in favor of the slot machine associated with positive outcomes, and to choose so as to avoid the slot machine with negative outcomes; and both groups learned to do so for either the monetary or the social condition. The fact that both groups showed such similar choice behavior and gave essentially identical valence ratings to the social stimuli provides strong evidence that our ASD group did not have a basic perceptual impairment in recognizing the value of the social stimuli (the emotional faces we used), nor did they have a basic impairment in understanding the task or in showing motivated behavior to obtain rewards. The highly similar overall behaviors and ratings in the two well-matched groups provide a starting point for discovering more specific dissociations, to which we turn next.

When looking in more detail at the cumulative choices made, and at the rate at which participants learned to choose optimally, we found a disproportionate impairment in the ASD group in learning to choose social rewards, compared to monetary rewards. Over time, the ASD group selected significantly fewer of the most rewarding social slot machine compared to the monetary slot machine, and also had a significantly slower initial learning rate for the socially rewarding slot machine, compared to the monetary slot machine. This pattern of findings was particularly compelling because it went in the opposite direction to what we found in the controls. Whereas controls cumulatively made a greater number of optimal choices in the social than the monetary condition, the ASD

group showed the converse pattern. Whereas controls generally learned faster in the social than the monetary condition, the ASD group again showed the converse pattern. These dissociations argue that the impairments in social reward processing found here in the ASD group cannot be attributed simply to an overall greater difficulty on the social than the monetary task. Rather, they appear to reflect a disproportionate impairment showing some domain-specificity for social rewards in people with autism.

One thing to note here is that initially an overt view of the data showed no differences on various task mastery metrics such as total correct and optimal switching between the groups – a finding that has been reported in other behavioral studies with ASD subjects (Kohls et al., 2011). However, once I moved beyond first-order measures, I was able to uncover subtle but robust differences in behavior towards social rewards.

Albeit very preliminary, the neuroimaging data can provide insight into the differences in neural processes that underlay these behavioral differences found between ASD and controls. In Chapter 3, I showed activation in ventromedial prefrontal cortex (vmPFC) coding value at the time of choice, activation in the ventral striatum (vStr) coding the discrepancy between obtained and expected outcome (PE), and activation in vmPFC coding outcome value. Moreover, these regions were activated jointly by monetary and social reward. In this present study, we found controls showed robust activation to both social and monetary rewards at the time of choice and at the time of outcome in a vmPFC ROI, as well as activation to PE in both social and monetary trials at the time of outcome in a VStr ROI. This was a nice replication of the results in Chapter 3. In contrast, while I did find activity for monetary PE and rewards at the time of outcome, I did not find the same activity in social PE and rewards for the ASD group.

This is consistent with the results of the study by Scott-Van Zeeland (2010), who found diminished vmPFC and VStr response during social but not monetary reward learning. They also reported that activity within the VStr predicted social reciprocity within the control group, but not the ASD group. My data offers some evidence for the theory of a social-domain specific deficit for processing social rewards, however the story is incomplete because I did not find vmPFC encoding of decision cues at value in the monetary condition.

What is even more puzzling is the fact that I did not find vmPFC activity encoding SV for either the monetary or the social condition. These results suggest that both monetary and social SV construction is abnormal in ASD. This is not inconsistent with the ideas of Dichter et al. (2012), who suggested that the decreased nucleus accumbens and vmPFC activity that they observed was suggestive of general reward system dysfunction in ASD during anticipation. Collectively, my data seem to suggest that reward experience is specifically impaired for social rewards, but that the construction of value signals at the time of choice is abnormal for general rewards.

Perhaps one compensatory mechanism for people with ASD is to circumvent the construction of value signals at the time of choice in the vmPFC and to rely instead on a hippocampus-mediated memory route. In a whole-brain exploratory analysis, I found hippocampus activation that was modulated by SV at the time of decision in social reward trials that seem to support such a theory. Dichter et al. in their study also noted an unexpected finding of hippocampus hyperactivation in the ASD group during monetary anticipation (Dichter, Richey, et al., 2012).

These findings demonstrate a subtle but specific behavioral insensitivity to social rewards in ASD, consistent with prior hypotheses, and are suggestive of an abnormal representation of reward values at the time of choice and potential specific impairment in processing social rewards at the time of outcome from fMRI. The behavioral results imply that even with abnormal processing, people with autism can employ compensatory behavioral mechanisms that from a first order look make them difficult to distinguish from controls.

These results begin to help us understand why people with autism may avoid other people and may not be motivated normally by social stimuli. Eventually, it will lead to a detailed and quantitative picture of the brain processes that underlie social rewards, and that are responsible for social motivation.

References

Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology 44,113-126.

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430-441.

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., and Bodfish, J. W. (2012). Reward circuitry function in autism during face anticipation and outcomes. J Autism Dev Disord 42, 147-160.

Kohls, G., Peltzer, J., Schulte-Ruther, M., Kamp-Becker, I., Remschmidt, H., Herpertz-Dahlmann, B., & Konrad, K. (2011). Atypical brain responses to reward cues in autism as revealed by event-related potentials. J Autism Dev Disord 41, 1523-1533.

Lin, A., Adolphs, R., Rangel, A.. (2011). Social and monetary reward engage overlapping neural substrates. Social and Cognitive Affective Neuroscience; doi:10.1093/scan/nsr006.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders 30, 205-223.

Lord, C., Rutter, M., Le Couteur, A. (1994). Autism Diagnostic Interview- Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders 24, 659-685.

O'Doherty, J.P., Deichmann, R., Critchley, H., and Dolan, R.J. (2002). Neural Responses During Anticipation of a Primary Taste Reward. Neuron 33, 815-826.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042-1045.

Rescola, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (New York, N.Y., Appleton Century Crofts), pp. 406-412.

Rolls, E.T., McCabe, C., and Redoute, J. (2008). Expected Value, Reward Outcome, and Temporal Difference Error Representations in a Probabilistic Decision Task. Cereb Cortex 18, 652-663.

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., and Bookheimer, S. Y. (2010). Reward processing in autism. Autism Res, 3, 53-67.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., and Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res 168, 242-249.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Voodoo correlations in social neuroscience. Perspectives on Psychological Science, 4.

Wechsler, D.A. (1981). The Wechsler Adult Intelligence Scale -- Revised. New York: The Psychological Corporation.

4.4 Summary and outlook

Several other recent studies have investigated reward processing in people with autism, and have suggested disproportionate impairments in social reward processing, as well as more general impairments in processing rewards across multiple stimulus types. For instance, it was reported that children with autism showed generally impaired implicit reward learning to both money and social stimuli, although the neural response to such stimuli measured with fMRI also showed a disproportionate abnormality for the social stimuli in particular (Scott-Van Zeeland et al., 2010). Another study (Dichter, Felder, et al., 2012) found that the neural response to monetary reward learning was abnormal in people with ASD, but that this abnormality disappeared during processing of interesting objects, possibly corresponding to the restricted-interests aspect of the autism phenotype. These studies are broadly consistent with the results of the two studies in this chapter. In the charitable donation task, people with ASD donated less overall (a domain-general impairment in reward processing); donated disproportionately less to people charities (a domain-specific impairment in social reward processing); and donated a lot to a few idiosyncratic non-social charities (intact or even exaggerated reward processing for a few unusual stimuli). In the neuroimaging study, we found again there was a domain-specific impairment of social reward learning exemplified by the relative ease ASD subjects had learning with monetary rewards over social rewards, while we found the exact opposite learning pattern in neurotypicals. These patterns show that high-functioning people with ASD are not altogether incapable of evaluating stimuli and making reward-based decisions about them--- but that how they evaluate particular categories of stimuli is abnormal.

The ratings data in the charitable donation tasks alludes to stages of processing that may be abnormal in people with ASD. While neurotypicals in the charitable donation task rated the impact of pictures and text descriptions on their donation amounts particularly high for people charities; the ASD group did not. One consistency among the picture and text descriptions of all people charities was the presence and mention of other people. For neurotypicals, it seems, the socially salient images and descriptions of other people, perhaps by recruiting pathways for empathy and theory of mind, contributed to the construction of a higher value signal at the time of donation, which lead ultimately to higher donations to people charities. This suggests in ASD subjects an impairment not with action selection but recruitment of additional social processing regions for the construction of accurate social reward value signals at the time of decision (A. Rangel et al., 2008).

Unfortunately, because of the small sample size in our neuroimaging study we were unable to do a PPI connectivity analysis similar to the one performed by Hare et al. in their charitable giving study. In their study, they found that value computations in vmPFC during charitable decision-making incorporated inputs from the pSTS and anterior insula, both areas involved in social cognition (T. A. Hare et al., 2010). My hypothesis for a PPI connectivity analysis with our neuroimaging data is activity in regions involved in social cognition modulate the inputs to vmPFC in the neurotypical group but not the ASD group.

One highly speculative idea is that the phenotype of autism is a consequence of compensatory mechanisms to cover heterogeneous injuries in the brain. There is compounding evidence that ASD may not be a single disorder with a single cause but the

phenotypic manifestation of many. This may explain why finding ASD-susceptibility genes has been so elusive, and why the few candidate genes that have been identified have been difficult to replicate between studies and populations (Alarcon et al., 2008; Basu, Kollu, & Banerjee-Basu, 2009; Geschwind, 2008). Higher cognition thought is not as crucial to life as lower brainstem functioning. Social cognition is disproportionately affected because, compared to the physical environment in general, the social environment is more complex and difficult to process and model. The processing demands for the social environment may be higher, with requirements to manage the dynamic recursive flow of information from multiple areas like pSTS and anterior insula.

While this theory would not explain characteristics of autism such as restricted repetitive behavior, it would help tie the results of Scott-Van Zeeland et al. (2010) and Dichter et al. (2012) to a unified theory. In both studies, they found generally lower levels of neural reward activity in both the monetary and social conditions in the ASD population compared to matched controls, yet we often observe only a behavioral deficit in social reward processing. Perhaps the social domain-specific deficits we observe in behavior are simply a reflection of the added complexity of social reward processing; for non-social reward processing, compensatory mechanisms such as hippocampus-mediated memory could be sufficient.

Returning to the social motivation hypothesis of autism (Dawson et al., 2002; Dawson et al., 1998; Grelotti et al., 2002), it remains an intriguing question how precisely the pattern of impairments we report here emerges during development. One possibility is that early domain-general impairments in reward processing, in a developmental context, give rise to impairments disproportionate for social stimuli (Triesch, Teuscher,

Deak, & Carlson, 2006). Similarly, early domain-general impairments in integrating complex contextual information may result in impairments particularly acute for social stimuli, simply because these draw more upon integrating multiple sources of information. An important future task will be to map out the abilities, and the concomitant brain responses, of people with ASD to process and evaluate a broad range of stimuli.

References

Alarcon, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., et al. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet, 82*(1), 150-159.

Basu, S. N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. *Nucleic Acids Res, 37*(Database issue), D832-836.

Chib, V. S., Rangel, A., Shimojo, S., and O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci, 29*(39), 12315-12320.

Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., and Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Dev, 73*(3), 700-717.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., and Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *J Autism Dev Disord, 28*(6), 479-485.

Dichter, G. S., Felder, J. N., Green, S. R., Rittenberg, A. M., Sasson, N. J., and Bodfish, J. W. (2012). Reward circuitry function in autism spectrum disorders. *Soc Cogn Affect Neurosci, 7*(2), 160-172.

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., and Bodfish, J. W. (2012). Reward circuitry function in autism during face anticipation and outcomes. *J Autism Dev Disord, 42*(2), 147-160.

Frith, C. D. (2007). The social brain? *Philos Trans R Soc Lond B Biol Sci, 362*(1480), 671-678.

Geschwind, D. H. (2008). Autism: many genes, common pathways? *Cell, 135*(3), 391-395.

Grelotti, D. J., Gauthier, I., and Schultz, R. T. (2002). Social interest and the development of cortical face specialization: what autism teaches us about face processing. *Dev Psychobiol, 40*(3), 213-225.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci, 30*(2), 583-590.

Harris, A., Adolphs, R., Camerer, C., and Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PLoS One, 6*(6), e21074.

Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatr, 35*(4), 100-136.

Lin, A., Adolphs, R., and Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci*.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci, 9*(7), 545-556.

Rogers, S. J., Ozonoff, S., and Maslin-Cole, C. (1991). A comparative study of attachment behavior in young children with autism or other psychiatric disorders. *J Am Acad Child Adolesc Psychiatry, 30*(3), 483-488.

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., and Bookheimer, S. Y. (2010). Reward processing in autism. *Autism Res, 3*(2), 53-67.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.

Triesch, J., Teuscher, C., Deak, G. O., and Carlson, E. (2006). Gaze following: why (not) learn it? *Dev Sci, 9*(2), 125-147.

CHAPTER FIVE

Conclusion and Final Words

Chapter 5: **Conclusions**

At the beginning of this thesis, I laid out some basic questions about social rewards. I answered these through a series of investigations with simple experiments. I summarize the findings and their significance here.

What is a social reward?

A social reward is a social interaction that individuals will seek out or modify their behavior to obtain. With a simple instrumental learning task, I was able to demonstrate that social rewards fit this behaviorist definition of a reward. I found that people will consistently pick cues that are associated with positive-valenced faces and matching sound effects.

What makes social rewards rewarding?

Social rewards are rewarding because they engage key areas of the reward circuitry like the vmPFC and vStr – the same overlapping areas that other types of rewards also engage. My results, as well as other studies (Chib et al. 2009), provide increasing support for the common neural currency theory.

How do they compare to other types of rewards?

Despite the high degree of overlap, there are unique aspects of social reward processing. In people with autism, I was able to demonstrate a domain-specific impairment with social cognition. They were found to have reduced preference and sensitivity towards social rewards. Furthermore, there was suggestive evidence for

deficits in the construction of value signals on which preferences regarding other people are based, despite otherwise intact social knowledge.

My neuroimaging data is also suggestive of deficits in general reward processing, particularly the construction of value signals at the time of decision, but these deficits are perhaps disproportionately worse for social reward processing. This is consistent with the findings of Dichter (2012) and Scott-Van Zeeland (2010), who also found hypoactivation of key reward areas in ASD subjects. Scott-Van Zeeland showed generally impaired implicit reward learning to both money and social stimuli, although the neural response to such stimuli measured with fMRI also showed a disproportionate abnormality for the social stimuli in particular.

Social processing may be differentially affected because the social world is more complex than other stimuli in our external world and requires drawing more upon integrating multiple sources of information. In particular, there is evidence that additional processing is required in order to interpret the value of socially relevant stimuli, originating in part from regions in the brain that social neuroscience has identified as coming disproportionately into play when we think about other people (Hare et al., 2010).

While my research suggests general reward processing is abnormal in ASD, it still lends support to the social motivation theory since it is social rewards that are the most greatly affected. As such, if social cognition deficits in ASD are a result of limited social learning because of decreased social motivational value, one obvious behavioral therapeutic intervention is to find ways to boost the motivational value of social stimuli early in development, perhaps by linking to extrinsic rewards.

Two other interesting research directions that are raised by my findings are a better understanding of how stimulus value signals are constructed at the time of decision, and whether in ASD one compensatory mechanism is to circumvent the construction of value signals at the time of choice in the vmPFC by relying instead on a hippocampus-mediated memory route. A PPI connectivity model with ASD fMRI data would be able to inform us about both questions.

One experimental challenge for these types of studies though is how do we know subjects are actually experiencing stimuli similarly. In my studies, I asked subjects for a self-report on pleasantness ratings of the stimuli to compare subjective experience between the two subject groups. I reported no differences between the two groups and concluded that both groups experience the social stimuli equally. However, it could be the case, that ASD subjects do not actually experience the stimuli similarly but they have been consciously taught to assign positive valence to smiling faces. A better way to assess subjective experience would be through an implicit measure such as reaction time or amount of work willing to perform. One of my colleagues in the lab ran such an experiment with individuals with autism. She showed subjects a vast set of stimuli including some social stimuli. The amount of time a stimulus stayed onscreen was determined by the number of key presses a subject exerted. At the end of the experiment, she also asked subjects to assess all the stimuli on a self-report liking rating. She found high correlation between the self-report and the work task.

And then even if we are able to perfectly match the subjective experience of the stimuli at outcome, a remaining challenge is how do we estimate accurate value signals at the time of choice for each subject. Many studies estimate the decision value by fitting a

computational learning model like Rescorla-Wagner to the data. Again, an intrinsic measure like reaction-time on a trial-by-trial level may be a close proxy but is often times very messy. Considering these challenges in future studies may help clean up the data and allow us to better model how decision values are constructed in the brain at the time of decision.

A related question is whether identical neurons code for both decision stimulus value and experienced reward. In the study in Chapter 3, it appeared that similar areas in vmPFC encoded both stimulus value and reward outcome, however, because we are limited by the spatial resolution of fMRI, we cannot rule out the possibility that there might be neuronal subpopulations within the vmPFC. Future studies using fMRI adaptation designs or direct electrophysiological recordings within these regions will be better able to address this issue. They may be able to elucidate the unusual result we found in the ASD population where the reward at outcome signal seemed to be preserved, while the value signal at the time of decision was absent.

Preserved outcome signal but absent anticipation signal is reminiscent of the findings from Bechara's study on patients with prefrontal damage. Bechara (1997) showed while patients with prefrontal damage had normal skin conductance responses (SCR) to winnings and loses at outcome, they did not generate the same anticipatory SCRs that guide neurotypical behavior before decisions. A natural follow-up to test whether the same case is true in individuals with ASD would be simply to measure electrophysiological responses during the instrumental learning task with both social and monetary rewards in Chapter 3.

Lastly, beyond evaluation of social stimuli in isolation, can social evaluation bleed into neighboring non-social objects? Marketing firms have already honed in our interests in social stimuli; commercials are replete with celebrities and other salient social stimuli. An area of personal interest to me is studying how our valuation of objects is influenced by social rewards. Jamil Zaki has already begun to explore this area. A study of his looks at how social influence modifies stimulus value construction in the vmPFC with a norm compliance task. First, he asked subjects to rate the attractiveness of a series of faces and then he showed them ostensibly how a previous group of participants had rated the faces. Then he had subjects re-rate the faces. He found changes in subject's ratings assigned to faces that were inconsistent with the group's ratings, accompanied with changes in value signal in the vMPFC as a result of social influence (Zaki et al. 2011). While it leaves as an open question how these signals are integrated, it provides the first evidence that leakage of social rewards does happen.

If the valuation of objects can be highly influenced by social interactions, we may continue to see higher and higher IPOs for social media companies like Facebook in the future. As Facebook and Twitter become household names, businesses are quickly realizing the influence of social rewards. Simple Energy, the company I mentioned at the beginning of my thesis, is harnessing the social reward of reputation and comparison to change consumer behavior to use energy more efficiently. What is novel about their business model is it does not require monetarily incentivizing people; they rely completely on the pull of social rewards.

References

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. Science 275, 1293-1295.

Chib, V.S., Rangel, A., Shimojo, S., and O'Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J Neurosci *29*, 12315-12320.

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., and Bodfish, J. W. (2012). Reward circuitry function in autism during face anticipation and outcomes. *J Autism Dev Disord, 42*, 147-160.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci *30*, 583-590.

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., and Bookheimer, S. Y. (2010). Reward processing in autism. Autism Res, 3, 53-67.

Zaki, J., Schirmer, J. and Mitchell, J. (2011). Social Influence modulates the neural computation of value. Psych Sci 22, 894-900.