A Quantitative Neural Biomarker for Rejection Estimation: A Neuroeconomic Approach for Evaluating Theory of Mind

Andrew Valdespino

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Psychology

John A. Richey
Sheryl B. Ball
Brooks King-Casas
Susan W. White

December 10, 2014
Blacksburg, Virginia
Keywords: (social anxiety, social phobia, fMRI, neuroeconomics, theory of mind, rejection)

A Quantitative Neural Biomarker for Rejection Estimation:

A Neuroeconomic Approach for Evaluating Theory of Mind

Andrew Valdespino

ABSTRACT

The clinical presentation of social phobia suggests that alterations in theory of mind (TOM) may play a systematic role in the development and maintenance of the disorder. In the current study we leverage a quantitative neuroeconmic approach to probe for neural and behavior markers of cognitive TOM, as well as rejection estimation, with a particular focus on social phobia. Participants comprised a non-clinical sample that was divided into low (N = 10) and high (N = 7) social anxiety groups based on self-report. Participants completed a one-sided uncertainty ultimatum game designed to probe individual differences in cognitive TOM, as well as rejection estimation. Contrary to predictions, there were no behavioral differences between high and low social anxiety groups in terms of rejection estimation. Although no between-group differences emerged in the traditional TOM network, significant differences were observed in subregions of the striatum during formulation of offers, likely corresponding to estimation of reward expectations. As hypothesized, and consistent with past research, imaging results support the existence of a network regions implicated in TOM, including the medial prefrontal cortex (MPFC) and the temporal parietal junction (TPJ). In addition to these regions, additional areas, including the caudate and insula, were also active during mentalizing components of the task. Collectively, results suggest a novel role for expected-value computations in the development and maintenance of social phobia.

Table of Contents

A Quantitative Neural Biomarker for Rejection Estimation: A Neuroeconomic Approach for Evaluating Theory of Mind

Social exchange between rational partners is constrained by a complex set of rules that establish shared expectations, beliefs and predictability in interaction. To the extent that a social actor has access to at most 50% of the available information in an exchange, the chain of events that emerge in dynamic social interactions can therefore be described as realizations of an ongoing inferential processes whose sole purpose is to generate 'best' responses to antecedents generated by a partner. An efficient method for estimating hidden parameters of social exchange is inferring the mental state of an interaction partner. This generates information that can be stored and used to predict a partner's behavior, a process broadly known as theory of mind (ToM). Recent advances in human neuroimaging and quantitative approaches to human decision making in the form of neuroeconomics provide techniques to reduce complex interactions into a comparatively simplified framework, in which specific parameters of interactions can be isolated and quantified. The current study seeks to leverage these techniques in order to clarify the neural bases of ToM, and the extent to which these relate to social functioning. Specifically, we leverage a behavioral economic model to identify neural systems underlying rejection estimation, and cognitive theory of mind. Further, we characterize the degree to which these systems relate to measures of social functioning, with implications for psychiatric and neurodevelopmental disorders of social cognition, with a specific focus on social phobia.

In a seminal paper, Premack and Woodruff (1978) provided an early conceptualization of ToM that now represents a foundational definition. Conceptually, the authors suggested that an individual uses a coherent set of rules based on best guesses (*theories*) to explain the behaviors of other individuals. The use of the term 'theory' was intentional, since an individuals' theory of mind is thought to function similarly to a traditional scientific theory, by positing a best guess (a mental state), which is then used to predict a phenomenon (behavior). As such, ToM is broadly construed as the ability to attribute independent mental states of others, in order to explain and predict behavior (Fletcher et al., 1995). ToM can be further reduced to a two-pronged process, involving not just inferences about the mental states of another agent, but also understanding that agent's behavior as a function of that mental state. Mental states come in various orders and categories including beliefs, desires, opinions, and emotions. It is therefore conceptually useful to recast ToM as a nexus of processes involving vertical (*depth* of reasoning, such as first, second order beliefs) and horizontal elements (*types* of reasoning, behavioral, cognitive affective). As an example, an individual may successfully infer the mental state of another agent, but be unable to successfully predict the emotion experienced, given that mental state. Furthermore, someone might be able to successfully predict behavior as a function of beliefs, but not as a function of emotions

Empirical study of ToM and related processes has provided insight into the psychopathology a variety of disorders characterized by deficits in interpersonal functioning, including, Autism (Baron-Cogen et al, 1985), Borderline Personality Disorder (Sharp et al., 2011), and Antisocial Personality Disorders (Dolan & Fullam, 2004). One disorder that has received comparatively little attention in the context of ToM work is social anxiety disorder (SAD; also called social phobia). SAD is a debilitating disorder with an early onset and a persistent life course (Davidson, Hughes, George, & Blazer, 1993), and is marked by distorted perceptions about social outcomes and others' beliefs (Hirsch & Clark, 2004). To date, only a handful of behavioral and neuroimaging studies exist assessing the role of ToM in the psychopathology of Social Phobia. In the neuroimaging literature, diminished medial prefrontal cortex (MPFC) activity has been associated with social anxiety during an interactive, multi-round economic exchange game (Sripada et al. 2009). Specifically when comparing conditions for which subjects played against a computer as opposed to a human partner, attenuated activity in the medial prefrontal cortex was found in socially anxious individuals versus controls. Furthermore, during a face perception study, socially anxious individuals exhibited attenuated activity in the posterior cingulate cortex (PCC), a brain region commonly implicated in ToM (Gentili et al., 2009). Together, these studies suggest that theory of mind may play a role in the psychopathology of Social Phobia.

Despite the relative scarcity of ToM work in the field of SAD, the clinical presentation of this disorder provides clues that alterations in ToM may play a systematic and non-trivial role in the etiology and maintenance of the disorder. For example, a hallmark feature of SAD is concern about being negatively evaluated by others (Salemink et al., 2013). Evaluation itself is a thought-centric concept, involving judging or thinking critically about a social actor. Similarly, the experience of humiliation and fear of embarrassment, also require estimation of the thoughts and mental states of others.

Although the broad notion of ToM may be a credible starting point in evaluating how self-other cognition is altered in SAD, the broadness of the ToM construct poses obstacles to direct empirical scrutiny. Thus, in the case of SAD it is useful to narrow the facet of ToM that may be most relevant to the disorder. One possibility is a particular aspect of ToM known as “rejection sensitivity” or “rejection estimation.” Rejection estimation is defined as the degree to which an individual assumes another individual will adopt a rejecting intention toward them. The term ‘estimation’ is used to highlight the ToM-centric nature of the term, since it involves estimating the degree to which someone is adopting a rejecting mental state, as opposed to just being sensitive or feeling hurt by actually being rejected.

Peer rejection experiences have been implicated as an etiological factor in the development of Social Phobia. In particular, distorted perceptions of rejection is thought to be especially characteristic of SAD (Hudson & Rapee, 2000), perhaps evolving into exaggerated rejection sensitivity (Fang et al., 2007). In a more general sense, rejection estimation may also be a fruitful cognitive architecture to study in a clinical context, since it may provide a single dimension along which multiple disorders of interpersonal dysfunction might be placed. Being able to predict whether another agent will be accepting or rejecting of a given behavior is an important social task that is likely essential for normative (naturalistic) social functioning. The ability to accurately estimate acceptance or rejection presumably plays a mechanistic and perhaps deterministic role in guiding behavior, by informing an agent of whether a behavior is permissible in a social context, before the behavior is even generated or emitted. An agent who tends to underestimate the degree to which other agents will reject their behavior may be more likely to violate social norms. One who tends to overestimate rejection may behave in ways that are risk-averse, possibly having negative implications for success in cooperative or perhaps most particularly in competitive relationships. The fiduciary estimation of such a parameter in a quantitative game, could clarify how ToM may be implicated in pathologies of social functioning.

In considering how to reduce the complexity of estimating social acceptance or rejection, the field of neuroeconomics may provide a useful framework to isolate and quantify specific parameters of ToM, such as the ones we outline here. Specifically, neuroeconomics leverages well-established behavioral economic models that offer quantitative depictions of social exchange (for reviews see Kishida, King-Cases & Montague, 2010; King-Casas & Chiu, 2012). Economic models expose quantitative benchmarks for behavior that facilitate comparison of social exchange patterns across individuals and groups. Importantly, these parameters can be correlated with brain activity to isolate where and how certain processes are implemented at a neural level. A general feature of economic exchange games is that players interact via mutual exchange of a commodity (most commonly money; Camerer, 2003). Rejection is a particularly tractable construct to assess in this framework, as a player's estimate of their partner's acceptance threshold can be codified in terms of the amount of money players send to their partners. These approaches allow for the establishment of biomarkers highly specific to a disorder (Chiu et al., 2008).

In the current study, we seek to reduce individual differences in rejection estimation to a single parameter. This parameter will then be correlated with both brain activity, as well as measures of social anxiety and social function. The goal is to isolate a neurobiological biomarker relevant to ToM that varies with anxiety symptomatology. This will be accomplished by leveraging a quantitative behavioral economic model describing subjects' choice behavior in an interactive economic exchange (Rapoport & Sundali, 1996). We will use a two-player game, that is a variation of the "Ultimatum Game" (Guth, 1995). In the traditional Ultimatum Game, one player is designated as the proposer, and a second player the responder. In the traditional Ultimatum Game, the job of the proposer is to take a sum of money, referred to as the 'pie', and to propose a monetary split of the pie between themselves and a second player. The responder can either accept or reject the proposed split. If the responder accepts, both players receive their share of the split. If the responder rejects, both players receive nothing. After the responder accepts or rejects, the round is over. The ultimatum game has traditionally been used to probe social and cognitive processes such as fairness preferences than ToM (Camerer, 2003).

In the version of the game used in the current study, one participant (a 'proposer') proposes a monetary split with a second player (a 'responder'). We systematically introduce uncertainty to the responder, by censoring information about the total amount being split. This is an adaptation known as the one-sided uncertainty Ultimatum Game. We will use this game to provide a quantitative framework within which to estimate individual differences in cognitive ToM and rejection estimation (Rapoport & Sundali, 1996). The traditional Ultimatum Game is a game of 'complete information.' That is, both the proposer and responder know the value of the pie to be split. Ultimatum games characterized by *uncertainty* (incomplete information) have also been developed (Mitzkewitz & Nagel, 1993). In version of the game we use (Rapoport & Sundali, 1996), the proposer splits an amount of money (a 'pie') each round that is randomly drawn from a range of numbers, referred to here as the 'range'. The proposer knows the range, as well as the value of the pie. The responder, however, only knows the range, and not the pie. The responder only knows what their proposed take of the pie is, and does not know with certainty how much the proposer kept (since the responder is unsure of the amount of the original pie). The ranges from which the pies are drawn are systematically increased and decreased, making the responder more or less certain, respectively the proposer is playing fair. For the proposer, the goal of the game is to play a series of single-round games, each with a different responder, and maximize overall earnings.

This task is relevant to ToM in the following ways. First, the proposer must adopt the responder's perspective (known as cognitive ToM), in order to avoid proposing a split that is unnecessarily low and likely to be rejected or unnecessarily high and economically disadvantageous. To do this, the proposer must estimate how confident the responder will be that a *fair* offer is being proposed. This estimate is encapsulated by the rejection estimation parameter designated as $\hat{\alpha}$. Note that cognitive ToM and rejection estimation are actually distinct processes. That is, individuals could theoretically have an equally perfect sense of what it is like to be in the uncertain responder's shoes (cognitive theory of mind) yet still vary in whether they think a responder has a rejecting intention (rejection estimation).

The quantitative model assumes the proposer assigns an empty parameter to the responder ($\hat{\alpha}$) corresponding conceptually to the minimum anticipated monetary value that will be accepted (Rapoport & Sundali, 1996). The parameter ranges from 0 to 1. For example, a proposer who assigns a parameter value of 0.8 to a responder, assumes that responder must be at least 80% certain the prosper is being fair, otherwise they will reject. A 'fair offer' is defined quantitatively as an equal or better split, favoring the responder. Since the proposer's goal in the task is to maximize their earnings, they should not offer the responder too much or too little. The rejection parameter is not measured as a ToM *ability* per se, but more as a *tendency*. That is, there is not an a priori optimal choice solution the proposer infers. This is not to say certain split choices are not better than others. Indeed, proposers assigning parameters of 0.9 are very likely underestimating the confidence tolerability of responder (many responders are still willing accept an offer, even if they are less than 90% sure an offer is fair). This rejection estimation parameter is intended to probe individual differences in the estimation of this parameter, without making explicit whether the subjects' parameters are misplaced.

The primary goal the current study is to probe individual differences in subjects' a priori rejection estimation. Accordingly, we also seek to mitigate potential confounding effects, such as other-regarding preferences, learning rates, and reputation-building on the estimation of the rejection sensitivity parameter. ‘Other-regarding preferences’ refers broadly to the concern over the welfare of others (Burkart, Fehr, Efferson, & van Schaik, 2007; Fehr & Fischbacher, 2003). For example, if a proposer offers a responder $10, this could either be because the proposer estimates a responder will reject a lower offer (rejection estimation), or because the proposer feels it is unfair to offer anything less (other-regarding preferences). To mitigate the interference of other-regarding preferences on the estimation of the rejection estimation parameter, a behavioral version of the dictator game will be used to probe the degree to which proposer behavior is driven by other regarding preferences. Briefly, participants play as the dictator, the role of whom is to make a monetary split between themselves and a second player. The second player cannot accept or reject the split, and must take what they are given. Assessing the size of the split offered by the dictator to the second player provides an isolated probe for fairness preferences. To reduce the impact of individual differences in learning rates associated with ToM and reputation-building effects proposers will be informed that they are randomly paired with a different responder, such that they never know the identity of each responder. This manipulation functions to prevent proposers from learning individual responder behavior, and further functions to prevent proposers from thinking they can influence or build a reputation with specific responders.

Based on the extant neuroimaging literature regarding the identity and spatial extent of a ToM network, we predicted (1) that at the neural level, there will be a positive correlation between level of responder uncertainty (across the 4 conditions ranging from certain to uncertain) in the anterior rostral prefrontal cortex (arPFC) as well as the temporoparietal junction (TPJ). It is noted that arPFC has been suggested as a localized region of ToM mentalizing within the MPFC (Amodio & Frith, 2006). The rationale for this expectation is that placing the responder under increasing levels of uncertainty systematically increases the perspective-taking demands on the proposer. Specifically, in the certainty condition, since both the proposer and responder know the size of the pie, it is relatively easy for the proposer to assess the responder's estimate of the pie size (since it is identical to their own). When uncertainty is introduced, the perspective of the responder systematically deviates from the proposer's, which prompts the proposer to engage in increased perspective taking.

We further hypothesized (2) that individual differences in the rejection estimation parameter will correlate positively with behavioral measures of social anxiety (Liebowitz Social Anxiety Scale, and Social Phobia Anxiety Inventory). Finally, we hypothesized (3) that individuals high in social anxiety would exhibit a relatively weaker correlation between increasing uncertainty and increased MPFC and TPJ activity.

Methods

Participants

Undergraduate students at Virginia Tech were recruited via Sona (Virginia Tech's online portal through which students can sign up for studies). Participants first completed an online screener to assess for social anxiety symptom severity. Participants completed the Liebowitz Social Anxiety Scale (LSAS) as well as the Social Phobia and Anxiety Inventory (SPAI – 23). A subset of low and high social anxiety participants were identified based on established cutoff scores on the LSAS and SPAI -23 self-report. Specifically an *and* rule was used, in order to identify high social anxiety cases. Those who exceeded the clinical cutoff on both the LSAS (total score >60) and the SPAI-23 (difference score >35; Rytwinski et al., 2009; Schry, Roberson-Nay, & White, 2012; Roberson-Nay et al., 2007). Low anxiety subjects were also characterized by an *and* rule, requiring an LSAS total score <30, and SPAI difference scores

<28. It is noted that a score of less than 28 is used on the SPAI-23 since a Difference score of 28 has been used as a more conservative clinical cutoff for SAD (Schry, Roberson-Nay, & White, 2012; Roberson-Nay et al., 2007). A total of 150 students (meeting low and high social anxiety cutoffs) were contacted via email and asked whether they wanted to take part in the imaging portion of the study. For the subset who expressed interest, the application of the following exclusion criteria resulted in a final sample of N=17: Left-handedness, history of head trauma resulting in more than 10-minutes of unconsciousness, history of seizures, and a recent history (past year) of taking psychotropic medication. The resultant pool of eligible subjects (N=17; 10 female) completed the fMRI task. This total fMRI sample was comprised of 10 low social anxiety (mean LSAS score = 19.4, SD = 6.5) and 7 high social anxiety (mean LSAS score = 65.7, SD = 3.6) participants. There were 5 females in the low social anxiety group, and 5 females in the high social anxiety group. A two-sample Welch's t-test did not reveal a significant difference in age or years of education between the low (mean age = 19.8, SD = 1.7; mean years of education = 14.7, SD = 1.2) and high (mean age = 20.0, SD = 1.4; mean years of education = 14.1, SD = 1.5) social anxiety groups.

Measures

Liebowitz Social Anxiety Scale – LSAS. The Liebowitz Social Anxiety Scale (LSAS) comprises 24 items, depicting different social situations. Each item features scale ratings, from 0 to 3, for both fear (ranging from "no fear" to "severe fear") and avoidance (ranging from "never" to "usually") (Liebowitz, 1987). Chronbach's α for the fear and avoidance subscales comprising the LSAS demonstrated high internal consistency at 0.91 and 0.92 respectively (Baker, Heinrichs, Kim, Hyo-Jin., & Hofmann, 2002).

Social Phobia and Anxiety Inventory (Abbreviated) – SPAI-23. The abbreviated Social Phobia Anxiety Inventory (SPAI-23) is a 23-item self-report questionnaire assessing specific somatic symptoms, cognitions, and behavior that may be elicited across a wide array of situations (Roberson-Nay, Nay, Strong, Beidel, & Turner, 1007). The frequency of items are rated on a 5-point scale, ranging from "never" to "always". Cronbach's α for the two subscales comprising SPAI, the Social Phobia and Agoraphobia subscales, demonstrated high internal consistency at 0.93 and 0.85 respectively (Schry, Roberson-Nay, & White, 2012).

Rejection Sensitivity Questionnaire – RSQ. The Rejection Sensitivity Questionnaire is an 18-item self-report questionnaire assessing degree of anxiety and concern, and rejection expectations, across a variety of situations (Downey & Feldman, 1996). The items are rated on a 6-point scale from "very unconcerned" to "very concerned (for questions addressing anxiety and concern about a situation) and from "very unlikely" to "very likely" (for questions addressing the degree to which rejection is likely). The RSQ shows high internal reliability (Cronbach's α = 0.83) and high test-retest reliability (Downey & Feldman, 1996).

Procedure

Upon arrival to the laboratory, and after completing informed consent, participants were screened for MRI contraindications, such as a history of head-trauma, or metal in the body. Participants then completed self-report measures via paper and pencil. Participants were then presented with written instructions explaining the nature of the fMRI task. Participants completed a 10-item multiple choice quiz after reading the instructions, to ensure that all instructions were understood. Subjects did not continue to the fMRI task unless they answered all questions correctly. If a question was not answered correctly, the staff explained why the answer was incorrect, and the participant was asked to repeat that item on the quiz. Participants then completed a practice version of the fMRI task on a computer outside of the scanner to familiarize themselves with the nature of the task. Subjects were informed prior to beginning the fMRI task, that they would be playing with five real players, who are playing from desktop computers in an adjacent room. Participants were told that they would meet these players at the conclusion of the experiment.

The One-Sided Uncertainty Ultimatum Game

In the traditional ultimatum game, one player (the proposer) is endowed with an amount of money, which they then propose to split with a second player (the responder). The responder can then either accept, or reject, the proposal. If the proposal is accepted, the proposer and responder keep the splits. If the proposal is rejected, both the proposer and responder receive nothing. This marks then end of a single round. The game can be either played as a single-shot version (single round), or over multiple rounds (Camerer, 2003). In the one-sided uncertainty ultimatum game, the responder is unsure of the proposer's endowment. The responder knows the range of numbers from which the original endowment is drawn, and the amount the proposer offers to them, and nothing else. From this perspective of incomplete information, the responder must choose to accept or reject the offer (Rapoport & Sundali, 1996; Slonim & Roth, 1998). In the current study, we use both the traditional, as well as the one-sided uncertainty ultimatum game.

For the fMRI task, each participant completed 4 runs. For each run, participants played as the 'proposer'. For each trial within a run, the proposer is endowed with an amount of money, called the 'pie'. The job of the proposer is to propose a split of the pie between themselves and another player, called the 'responder'. Participants were told that they were playing against five human responders, who were playing from another room at the facility. Participants were also told that they would be randomly paired with one of the five responders on each trial. In reality, participants played against a computer program that mimicked actual responder behavior derived from the behavioral results of former experiments using the same task (Rapoport & Sundali, 1996; Slonim & Roth, 1998).

Each of the 4 runs was either characterized by ‘uncertainty’ (3 runs) or ‘certainty’ (1 run) type. Each run was counter-balanced by subject. For the 3 uncertainty runs, the pie was randomly drawn from one of the following three ranges ($0 - $30, $5 - $25, $10 - $20). Each run corresponded to one of these ranges, such that the range remained fixed for the entire run. The range was displayed prominently to the proposer at the top of the screen for the duration of each run. The pie-values drawn on each trial were drawn from an equal distribution of the range, in intervals of $3, $2, and $1 (for the $0 - $30, $5 - $25, and $10 - $20 run respectively). As an example, for the $5 - $25 range, the pie values ranged from $5 to $25, in increments of $2. Each of the pie values was drawn twice per run. This yielded a total of 22 trials for the uncertainty runs (with the exception of the $0 - $30 range run, which featured 20 trials, since it is not possible to split from $0). The pie value was displayed under the range for each trial. Once the pie value was displayed, the participant cycled through possible splits (in increments of $1) using an fMRI compatible button box. They then indicated via button-press when they were satisfied with their split, at which point the proposal was revealed to the responder. The responder would then indicate whether they accepted or rejected the split, which the responder viewed on the screen in text as ‘Accept’ or ‘Reject’. If the responder accepted the split, the proposer received their share of the split, and the responder received theirs also.

For ‘Uncertainty’ runs, proposers are told that the responder knows only the range of possible values, but never knows the total actual value (pie) to be split. The certainty run differed from uncertainty runs in that the responder always knew the pie the proposers was splitting. The degree of uncertainty systematically varies across the four runs. For the certainty run the responder is provided with the exact value of the pie. By contrast, for the uncertainty run with range \$10 - \$20, the pie could be any value between \$10 and \$20. The next level of uncertainty (increased uncertainty) is characterized by the \$5 - \$25 run. This condition is relatively more uncertain, since the pie is drawn from a wider range. Following this logic, the \$0 - \$30 run is characterized by the most uncertainty.

Each trial in a given run consisted of the following: (1) Each exchange began with a crosshair centered on the screen for 4000ms indicating the beginning of a round, followed by a blank black screen with a jittered duration (ranging from 100ms – 300ms). (2) The range from which each pie was drawn was presented centered at the top of the screen. The range remained fixed for the entirety of the run. The pie to be split on the current trial, varied by trial and was displayed simultaneously. The pie amount was represented by a single numeric value. For each trial, the pie to be split was randomly drawn from an evenly distributed full spread of the range. Each subject was presented with an identical pseudorandom distribution of pies (trials are identical across subjects). The participant was represented as a blue schematic face on the left side of the screen, and the responder as a purple schematic face on the right. The proposer then chose via button press (decision phase) how much they wanted to keep, which was displayed under the blue face. The remaining amount was displayed under the purple face. There was no time-limit placed on the proposer during the decision phase. The participant indicated via a button press when they were ready with their choice, at which point the split was revealed to the responder. A question mark was then placed under the purple face, indicating that the responder was deciding whether to accept or reject (the anticipation phase). The anticipation phase lasted

8000ms. The responder's decision, either to accept or reject, was then displayed in white characters under the purple face for 4000ms (the outcome phase). The next trial began after a blank screen of 4000ms. For the decision, anticipation, and outcome phases, the range, current pie, and faces each remained fixed on the screen. The only change in the visual presentation of the paradigm across these phases was the information featured under the faces. The total length of the experiment was approximately 40 minutes, with runs lasting 10 minutes each. Stimuli were presented using PyGame, an open-source python based stimulus presentation software.

After completion of fMRI task, participants completed a behavioral version of a single-shot dictator game to provide an experimental control for other-regarding preferences. The game was played on a standard laptop computer. Participants played as the dictator. Participants were told that they would be matched with one randomly drawn participant for this game.

For the fMRI task, participants were paid for every trial, at a rate of \$0.15 for every dollar earned. Participants were paid using the same payment schedule for the dictator game. This payment schedule was explained to participants at the beginning of the experiment.

Description of the Quantitative Model[1]

A Proposer makes an offer to split $\$k$. The amount $\$k$ is drawn from the distribution $[a,b]$. Let $\$x$ denote the amount the proposer keeps for himself, and $\$y$ the amount given to the responder, where $x+y=k$. Let $p(y)$ denote the responder's probability y, given their knowledge of the distribution, that $y/k \geq 0.50$. A distribution of $y/k \geq 0.50$ is assumed in the current model to be a generous offer. Intuitively, this is a case where the proposer is offering less for themselves than for the responder. For a uniform distribution:

$$p(y) := \Pr[(y/k) \geq 0.5] = (2y-a)/(b-a)$$

[1] A detailed quantitative description can be found in Rapoport & Sundali (1996).

Steps to the model:

Step 1:

The proposer is assumed to view the ultimatum game as a strategic task. That is, a strategic sender views the game such that if they think they can get more than $k/2$, they will do so.

Step 2:

A proposer viewing the game strategically believes the responder will reject any offer y, if $p(y) \le \alpha$, where $0 < \alpha < 1$ is a fixed constant. Since the proposer does not know the true value of α, they must estimate it, denoted by $\hat{\alpha}$.

Step 3:

Upon estimating $\hat{\alpha}$, the proposer determines y, which denotes the maximum share of the pie they are willing to send to the responder. The actual offer y^* is the solution of the following equation:

$$p(y) = (2y - a)/(b - a) = \hat{\alpha} \tag{1}$$

A constraint is applied such that $y \le \gamma k$, where $0 < \gamma < 1$. From this it follows that the proposer will offer the responder

$$y^* = \min(y', \gamma k) \tag{2}$$

where y' is the solution of Equation 1:

$$y' = [\hat{\alpha}(b - a) + a]/2$$

Step 4:

After receiving the offer y, the responder will estimate the size of the pie by $\hat{k}$, and reject the offer if $y/\hat{k} \le \alpha$, and accept it otherwise.

Model assumptions:

Assumption 1: $\hat{\alpha} \le 0.5$. That is, the proposer believes the responder will not reject any offer if her probability that the offer is generous exceeds 0.5.

Assumption 2: $\hat{\alpha}$ is independent of the parameters of the pie size distribution.

Assumption 3: $\hat{\gamma} \le 0.5$. That is, the proposer will keep at least 50% of the pie for himself.

The function of this assumption is to reduce the number of parameters from two to one.

A numerical example:

Assume a uniform distribution $[a,b]$ where a=\$5, and b=\$25. If $\hat{\alpha} = 0.3$. The proposer will offer the responder \$5.5 if $k > [\hat{\alpha}(b-a)+a]$, and $k/2$, if $k \le [\hat{\alpha}(b-a)+a]$. If $\hat{\alpha} = 0.4$ and $y = 0.5$, the offer by the proposer, γ^* will be \$6.5 if $k > [\hat{\alpha}(b-a)+a]$, and $k/2$ otherwise.

Data Analysis

fMRI Data Acquisition

fMRI scanning was performed on a Siemen's 3.0 Tesla Allegra scanner. Head movement was restricted using foam cushions. An eight-channel head coil was used for parallel imaging. Initial high-resolution T1-weighted scans were acquired using an MP-RAGE sequence (Siemens). These scans were used for coregistration with the functional data. Structural images were aligned in the near axial plane defined by the anterior and posterior commissures. Whole brain functional images consisted of 30 slices parallel to the AC-PC plane using a BOLD-sensitive gradient-echo EPI sequence, at TR of 2000 ms (TE: 30 ms; FOV: 22 cm; isotropic voxel size: 3.44 Å~ 3.44 Å~ 4.0).

fMRI Data Processing

Preprocessing of functional data was accomplished using SPM8 (Wellcome Department of Cognitive Neurology, University College London). The analysis was implemented in NiPype, a python-based framework designed for highly pipelined processing of fMRI data from several neuroimaging packages (http://nipy.org/nipype). Separation of brain tissue from skull was accomplished using FSL's brain extraction tool (BET). Functional data was slice-timing corrected, realigned to the middle image of the functional run, temporal band-pass filtered (0.009 Hz < f < 0.08 Hz) and corrected for motion. Motion corrected data was registered to the T1-weighted image, and normalized to standard (Montreal Neurological Institute; MNI) stereotactic space.

Estimating $\hat{\alpha}$

The rejection estimation parameter $\hat{\alpha}$ was calculated individually for every trial in each of the three uncertainty conditions. Trials for which it was not possible for the subjects to express an $\hat{\alpha}$ of at least 1.0 were eliminated. These trials were eliminated to provide a more realistic appraisal of proposers' rejection estimation parameters. For example, say a proposer's a priori $\hat{\alpha}$ is 0.9. In the context of the quantitative model, this proposer believes that unless there is a 90% chance from the responder's perspective that the offer is fair, the responder will reject. In order to reveal this proposer's hidden parameter, one need include a pie for which it is possible to make an offer reflecting that parameter. If on a given trial, that proposer is drawn $7 (and assume the range is $5 - $25) then even a maximum offer of $7 from the proposer, would still only reflect a parameter of 0.45 (an incorrect estimate of this proposer's $\hat{\alpha}$). More generally, to accurately estimate a proposer's $\hat{\alpha}$ one must not artificially create a ceiling effect, by not allowing them to send offers reflective of their $\hat{\alpha}$ estimate.

Identifying brain regions implicated in rejection estimation

To identify brain regions implicated in rejection estimation, all brain activity during the decision phase was identified in a fixed-effects (subject-level) model. Then, the point-estimate for $\hat{\alpha}$ was regressed against all activity during the decision phase. We focused our analysis on the decision phase (i.e. prior to the proposal) under the assumption that subjects are actively estimating how much the responder will accept or reject during this period.

Probing for Learning Effects

To assess for the presence of learning, subjects' rejection sensitivity parameter for the first versus the second half of runs was compared. Specifically, for each subject, a mean will be taken of the trial-wise rejection sensitivity parameters both for the first and second runs, and compared with the third and fourth runs. These means were compared via two-tailed t-test to assess whether the parameter has changed over time.

Results

Probing for the effect of learning on the rejection estimation parameter. Since the goal was to ascertain participants' a priori rejection estimate parameter, the presence of a learning effect would confound the parameter. To probe for learning effects, a one-way analysis of variance was performed of run number (first/second/third/fourth) on the rejection estimation parameter. There was no significant effect of run number on the rejection estimation parameter [$F(1,16) = 0.326$, $p = .58$], suggesting that the order of the run was not related to the point estimate for alpha.

Relating social anxiety symptomatology with the rejection estimation parameter. Given that rejection sensitivity has been implicated in the development and maintenance of social phobia symptomology, we hypothesized that social phobia symptomatology would correlate positively with the rejection estimation parameter. The specific prediction was that participants high in social anxiety would estimate responders to require higher monetary offers in order to accept. To test the hypothesis, a univariate linear regression was performed between the rejection estimation parameter averaged over all runs as a dependent variable and LSAS score as the independent variable. This correlation was not significant [$b = 0.01$, $t(15) = 0.12$, $p = 0.90$]. The same regression was conducted separately using the SPAI-23 difference score, as well as the RSQ score, as independent variables. In both cases the correlation was not significant [SPAI, $b = 0.01$, $t(15) = 0.08$, $p = 0.94$; RSQ, $b = 0.01$, $t(13) = 0.29$, $p = 0.78$). Unstandardized regression coefficients reported.

Replication of previous one-sided uncertainty ultimatum game findings. Consistent with prior results, proposers offered lower percentages of the pie as responder uncertainty increased (Rapoport & Sundali, 1996). Specifically, a two-way mixed analysis of variance (ANOVA) was conducted for Group (low social anxiety/high social anxiety) × Certainty (certainty/low uncertainty/mid uncertainty/high uncertainty) on the percentage of the pie given by responders. There was a highly significant main effect of Certainty [$F(3,45) = 56.14$, $p < 0.001$], such that proposers offered less as the level of uncertainty increased. Both the main effect of Group [$F(1,15) = 0.003$, $p=0.95$], as well as the interaction between Group and Anxiety [$F(3,45) = 0.88$, $p=0.46$] were not significant (see Figure 1).

Brain regions correlating with degree of responder uncertainty. We hypothesized that the VMPFC, and TPJ would scale positively with increasing responder uncertainty. We conducted a random effects (second level) analysis assessing for brain regions that scaled monotonically with increasing uncertainty. To do this, we performed a linear trend analysis in which we weighted the scan run [certainty, low uncertainty, medium uncertainty, high uncertainty] as [1,2,3,4], respectively. Results of a whole-brain analysis during the decision phase indicted activation bilaterally in the caudate head (Figure 2) and insula (Figure 3) as well as areas including the bilateral VMPFC, RTPJ ($p < .001$, FDR corrected).

Between group differences in brain activity related to mentalizing. We compared the mean activation of brain regions surviving this linear trend analysis across low and high social anxiety groups. There were no significant clusters of activation (using a threshold of $p < .05$, FDR corrected). Since we hypothesized that high social anxiety subjects would exhibit reduced activation in TOM regions during perspective taking, the mean activation during the decision phase (across all certainty conditions) was compared between groups. Specifically we weighted group [low social anxiety, high social anxiety] as [1,-1], to assess for regions that were activated less in low social anxiety versus high anxiety subjects. This analysis yielded decreased activation in bilateral putamen, left precentral gyrus, left superior frontal gyrus, and left supra marginal gyrus ($p < .05$, FDR corrected). High social anxiety was not associated with increased activity during the decision pahse in any areas ($p < .05$, FDR corrected).

Brain regions correlating with the rejection estimation parameter. A central goal of the study was to highlight brain regions involved in rejection estimation, in addition to mentalizing more generally. To probe regions implicated in rejection estimation, all brain activity during the decision phase was identified in a fixed-effects (subject-level) model. Then, the point-estimate for $\hat{\alpha}$ was regressed against all activity during the decision phase. A set of regions were negatively correlated with the rejection estimation parameter during the decision phase ($p < .001$, FDR corrected) including bilateral frontal pole, bilateral anterior insula, bilateral caudate, and bilateral paracingulate gyrus. The left post central gyrus was the only region that positively correlated with the rejection estimation parameter during the decision phase ($p < .001$, FDR corrected). At a threshold of $p < .01$ (FDR corrected) additional regions exhibited positive correlations with the rejection estimation parameter, including bilateral anterior paracingulate gyrus, bilateral subcallosal cortex, bilateral anterior superior temporal gyrus, and bilateral posterior parahippocampal gyrus.

Discussion

This study leveraged a neuroeconomic approach to assess neural and behavioral correlates of theory of mind, with a focus on rejection estimation. We also sought to clarify whether specific behavioral or neural biomarkers related systematically to social anxiety symptomatology. To assess for brain regions implicated in TOM, we used a behavioral economic approach to identify which brain regions were shown to parametrically track with rejection estimation, as well as perspective taking under varying levels of responder uncertainty. In the current study, the ventral MPFC and the TPJ were implicated in perspective taking under conditions of uncertainty. The current result that the VMPFC parametrically tracks with increasing uncertainty is potentially consistent with prior quantitative approaches implicating the VMPFC in complex strategic thinking. Specifically, Coricelli and Nagel (2009) found the VMPFC to mediate activity when people interacted with high-level, but not low-level thinkers, suggesting the VMPFC is involved with more sophisticated mentalizing. Furthermore, the VMPFC has been shown to scale with depth of mentalizing (Coricelli & Nagel, 2009).

Several unexpected results were also observed. A set of additional regions beyond the traditional TOM network, including the caudate and insula, were found to increase with degree of responder uncertainty. Activity in the caudate under conditions of uncertainty is potentially consistent with literature implicating the caudate in expected value computations (Zhu, Mathewson & Hsu, 2012; Montague, King-Casas, & Cohen, 2006; Knutson, Adams, Fong, & Hommer, 2001). Specifically, in the current study we found a highly significant behavioral result, that as the level of uncertainty increases, proposers offer a lower percentage of the pie to proposers (and hence keeping more for themselves). Therefore, one potential interpretation is that the caudate is tracking the increased expected value associated with the relatively higher amounts of money kept by proposers under conditions of increasing uncertainty.

The finding of increased insula activity under increasing levels of uncertainty is further consistent with this interpretation. The insula has been associated with risk taking (Platt and Huettel, 2008). One interpretation is that the insula is tracking increased risk taking associated with the relatively smaller offers made by proposers under conditions of increasing uncertainty. It might be argued though that the lower offers made under conditions of increasing responder are not necessarily more risky, since the responder is less sure of whether the proposer is being fair. The insula has also been implicated in norm violation (King-Casas et al., 2008). When proposers offer less, they do so only responders are unsure of whether proposers are being fair. As such, the element of deception associated with lower proposer offers under conditions of uncertainty may be represented by the insula as a norm violation.

An additional hypothesis of the current study was that the rejection estimation parameter would positively correlate with social phobia symptomatology. This hypothesis was not supported. In clinically focused neuroimaging studies, it is not uncommon for between-group differences to manifest at the neural, but not at the behavior level. As an example, Sripada et al. (2009) found no behavioral differences between social phobia participants and controls in the context of a trust game designed to probe mentalizing. At the neural level though, individuals with social phobia were marked by diminished activity in the MPFC. Similarly, in the context of neuroimaging research on autism, a neural but not behavioral marker was found to distinguish high functioning autism participants from controls in a neuroeconmic trust game (Chiu et al., 2008).

The absence of a positive correlation between social phobia symptomatology and the rejection estimation parameter still deserves careful consideration, since this was a central hypothesis. One interpretation is that individuals both high and low in social phobia symptomology make similar inferences about the rejection thresholds of other individuals. Notably though, responders in the current task provided immediate feedback as to whether they accepted or rejected the proposals. Therefore, while steps in the current study were taken to prevent proposers learning about the behavior of specific responders, high social anxiety subjects may still have learned and adjusted to the overall probabilistic tendency for a responder to accept or reject. This is particularly possible since responder behavior in the ultimatum game is relatively stable across individuals. As an example, in a regular ultimatum game, responders tend to accept almost all offers for which they receive forty percent of the split or greater (Slonim & Roth, 1998). Learning about the overall response tendencies of responders may have diminished the ability to detect participants' a priori rejection estimation parameter.

We further hypothesized that high social anxiety individuals would exhibit relatively decreased activity in the MPFC and TPJ while engaging in mental state reasoning, and further that. However a between-groups analysis indicated that high social anxiety individuals were characterized by reduced activity in bilateral putamen versus low social anxiety subjects during the decision making phase of the task. In a clinical context, a reduction in putamen activity has been found in patients with bipolar disorder while mentalizing. The putamen has been found to be involved in reward-based and belief-based learning in a social context. Where as reward-based learning only takes expected and received rewards into account (ignoring social context) belief-based learning in involves understanding and anticipating the behaviors of others (Zhu, Mathewson & Hsu, 2012). Consequently, the current findings point to a potential inroad regarding altered social learning computations at the neural level in the context of social anxiety.

Results of this study should be evaluated in light of study limitations. Specifically, a relatively low sample-size limits the ability to make between group statistical inferences. In addition, despite the sample being ‘non-clinical’, self-report of high social anxiety subjects suggests, for a subset at least, that a diagnosis of social phobia is possible. As such generalizations to a non-clinical sample may be problematic. In addition, as noted previously in the discussion, it is possible that proposers may have learned about the relatively stable response tendencies across responders, making it potentially problematic to ascertain proposers’ a priori rejection estimation parameters. Future studies might consider eliminating responder feedback completely, with the function of preventing proposers from learning about responder preferences.

In conclusion, this study utilized a modified ultimatum game in a neuroimaging framework, to evaluate regions of the brain that are associated with estimating a probability distribution for a counterpart under conditions of one-sided uncertainty. Uncertainty is ubiquitous in social interaction. Specifically, humans are rarely fully aware of the complex set of personal histories, possessions, and intentions of those in their social group. Therefore, humans must frequently assess how their behaviors will be perceived by agents with incomplete information. Quantitative approaches in neuroeconomics make it possible to represent abstract social processes in a quantifiable and explicit form. The current study sheds light on an array of mental processes implicated in social interaction characterized by uncertainty.

References

Aldridge, F. J., Gibbs, V. M., Schmidhofer, K., s Williams, M. (2012). Investigating the clinical usefulness of the Social Responsiveness Scale (SRS) in a tertiary level, autism spectrum disorder specific assessment clinic. *Journal of autism and developmental disorders*, 42(2), 294–300. doi:10.1007/s10803-011-1242-9

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature reviews. *Neuroscience*, 7(4), 268–77. doi:10.1038/nrn1884

Baker, S. L., Heinrichs, N., Kim, Hyo-Jin., & Hofmann, S. G. (2002). The Liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. *Behaviour Research and Therapy*, *40*(6), 701–715. doi:10.1016/S0005-7967(01)00060-2

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1), 5–17.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21, 37–46.

Beck, A. T., Steer, R.A., & Garbin, M.G. (1988) Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77-100.

Beidel, D., Turner, S., Stanley, M., & Dancu, C. (1989). The Social Phobia and Anxiety Inventory: concurrent and external validity. *Behavior Therapy*, 417–427.

Burkart, J. M., Fehr, E., Efferson, C., & van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: common marmosets provision food altruistically. Proceedings of the National Academy of Sciences of the United States of America, 104(50), 19762–6.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–35. doi:10.1002/hbm.20671

Camerer, C.F. (2003). Behavioral Game Theory: Experiments in Strategic Interaction (Princeton, NJ: Princeton University Press).

Chiu, P. H., Kayali, M. A., Kishida, K. T., Tomlin, D., Klinger, L. G., Klinger, M. R., & Montague, P. R. (2008). Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron*, 57(3), 463–73. doi:10.1016/j.neuron.2007.12.020

Constantino, J. N., & Gruber, C. P. (2005). Social Responsiveness scale (SRS). Los Angeles, CA: Western Psychological Services.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9163–8. doi:10.1073/pnas.0807721106

Dennett, D. (1978). Beliefs about beliefs. Behavioral and Brain Sciences, 1, 568–570.

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine*, 34(6), 1093–1102. doi:10.1017/S0033291704002028

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology*, *70*(6), 1327–1343. doi:10.1037//0022-3514.70.6.1327

Domsalla, M., & Koppe, G. (2013). Cerebral Processing of Social Rejection in Patients with Borderline Personality Disorder. *Social cognitive*.

Fang, A., Asnaani, A., Gutner, C., & Cook, C. (2011). Rejection Sensitivity mediates the relationship between social anxiety and body dysmorphic concerns. *Journal of Anxiety Disorders*, 25(7), 946–949. doi:10.1016/j.janxdis.2011.06.001.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. Nature, 425(6960), 785–91.

Fletcher, P.C., Happe´, F., Frith, U., Backer, S.C., Dolan, R.J. (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition*, 57, 109–28.

Gentili, C., Ricciardi, E., Gobbini, M. I., Santarelli, M. F., Haxby, J. V, Pietrini, P., & Guazzelli, M. (2009). Beyond amygdala: Default Mode Network activity differs between patients with social phobia and healthy controls. *Brain research bulletin*, 79(6), 409–13. doi:10.1016/j.brainresbull.2009.02.002

Güth, W. (1995). On ultimatum bargaining experiments - A personal review. *Journal of Economic Behavior & Organization*, 27(3), 329-344.

Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceedings of the National Academy of Sciences of the United States of America 2008;105:6741–6746. [PubMed: 18427116] Izuma

Hirsch, C. R., & Clark, D. M. (2004). Information-processing bias in social phobia. *Clinical Psychology Review*, 24, 799e825, doi:10.1016/j.cpr.2004.07.005.

Hudson, J. L., & Rapee, R. M. (2000). The Origins of Social Phobia. *Behavior Modification*, 24(1), 102–129. doi:10.1177/0145445500241006

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* (New York, N.Y.), 321(5890), 806–10. doi:10.1126/science.1156902

Kirk, U., Downar, J., & Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Frontiers in Neuroscience*, *5*(April), 49. doi:10.3389/fnins.2011.00049

Kishida, K. T., King-Casas, B., & Montague, P. R. (2010). Neuroeconomic Approaches to Mental Disorders. *Neuron*, 67(4), 543–554. doi:10.1016/j.neuron.2010.07.021

Knight, R. G., Waal-Manning, H. J., & Spears, G. F. (1983). Some norms and reliability data for the State-Trait Anxiety Inventory and the Zung Self-Rating Depression Scale. *British Journal of Clinical Psychology/British Psychological Society*, 22, 245–249.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(16), RC159. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11459880

Liebowitz, M. R. (1987). Social phobia. *Modern Problems Pharmacopsychiatry*, 22, 141–173.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour research and therapy*, 28(6), 487–95.

Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 171–198.

Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience*, *29*, 417–48. doi:10.1146/annurev.neuro.29.051605.112903

Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*. 1, 515–526.

Platt, M. L., & Huettel, S. a. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nature Neuroscience*, *11*(4), 398–403. doi:10.1038/nn2062

Rapoport, A., & Sundali, J. A. (1996). Ultimatums in Two-Person Bargaining with One-Sided Undertainty: Offer Games. *International Journal of Game Theory*, 475–494. doi:1996

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Rilling, J. K., Sanfey, A. G., Aronson, J. a, Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*(4), 1694–703. doi:10.1016/j.neuroimage.2004.04.015

Roberson-Nay, R., Strong, D. R., Nay, W. T., Beidel, D. C., & Turner, S. M. (2007). Development of an abbreviated Social Phobia and Anxiety Inventory (SPAI) using item response theory: the SPAI-23. *Psychological Assessment*, 19(1), 133–45. doi:10.1037/1040-3590.19.1.133

Sanfey, A. G., Rilling, J. K., Aronson, J. a, Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science (New York, N.Y.)*, *300*(5626), 1755–8. doi:10.1126/science.1082976

Schmidt, N. B., Richey, J. A., Buckner, J. D., & Timpano, K. R. (2009). Attention training for generalized social anxiety disorder. *Journal of Abnormal Psychology*, 118(1), 5–14. doi:10.1037/a0013643

Schry, A. R., Roberson-Nay, R., & White, S. W. (2012). Measuring social anxiety in college students: a comprehensive evaluation of the psychometric properties of the SPAI-23. Psychological Assessment, 24(4), 846–54. doi:10.1037/a0027398

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child and Adolescent Psychiatry*, 50(6), 563–573. doi:10.1016/j.jaac.2011.01.017

Slonim, R., & Roth, A. (1998). Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic. *Econometrica*, 60(3), 569-596.

Spielberger, C. D. (1989). State-Trait Anxiety Inventory: Bibliography (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Sripada, C. S., Angstadt, M., Banks, S., Nathan, P. J., Liberzon, I., & Phan, K. L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport*, 20(11), 984–9. doi:10.1097/WNR.0b013e32832d0a67

Ruocco, A. C., Medaglia, J. D., Tinker, J. R., Ayaz, H., Forman, E. M., Newman, C. F. (2010). Medial prefrontal cortex hyperactivation during social exclusion in borderline personality disorder. *Psychiatry research*, 181, 233-236.

Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. a. (1989). An empirically derived inventory to measure social fears and anxiety: The Social Phobia and Anxiety Inventory. *Psychological Assessment*, 1(1), 35–40. doi:10.1037//1040-3590.1.1.35

Watson, D., Clark, L. a, & Tellegen, a. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063–70.

Woodbury-Smith, M., Robinson, J., & Baren-Cohen, S. (2005). Screening adults for Asperger Syndrome using the AQ: Diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, 35, 331-335.

Yoon, K. L., Fitzgerald, D. a, Angstadt, M., McCarron, R. a, & Phan, K. L. (2007). Amygdala reactivity to emotional faces at high and low intensity in generalized social phobia: a 4-Tesla functional MRI study. *Psychiatry research*, 154(1), 93–8. doi:10.1016/j.pscychresns.2006.05.004

Yoshida, W., & Seymour, B. (2010). Neural mechanisms of belief inference during cooperative games. *The Journal of …*, *30*(32), 10744–10751. doi:10.1523/JNEUROSCI.5895-09.2010

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–24. doi:10.1073/pnas.11

Appendix A

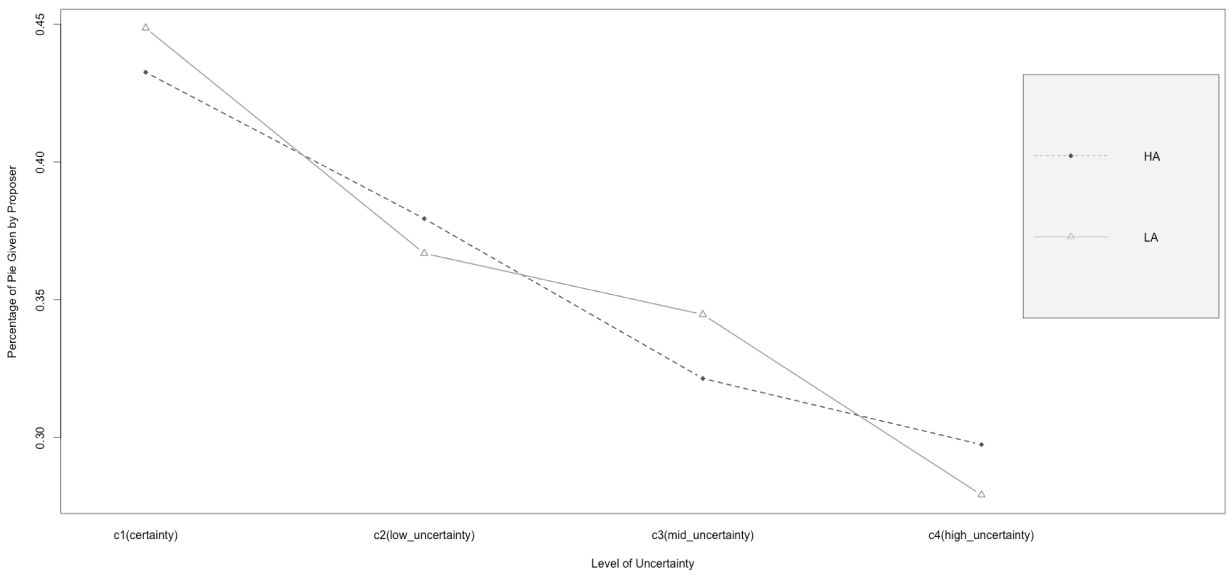

Figure 1: Relationship Between Responder Uncertainty and Proposer Offers. Mean percentage of the pie offered by proposers for ‘certainty’, ‘low uncertainty’, ‘mid uncertainty,’ and ‘high uncertainty’ conditions, for both low (solid line) and high (dashed line) social anxiety participants. The ‘certainty’ condition refers to trials for which the responder knows the size of the pie. For the remaining three conditions, which are characterized by low (range =10), mid (range = 20), and high (range = 30) uncertainty, the responder does not know the size of the pie, but only knows the range the pie is drawn from.

Appendix B

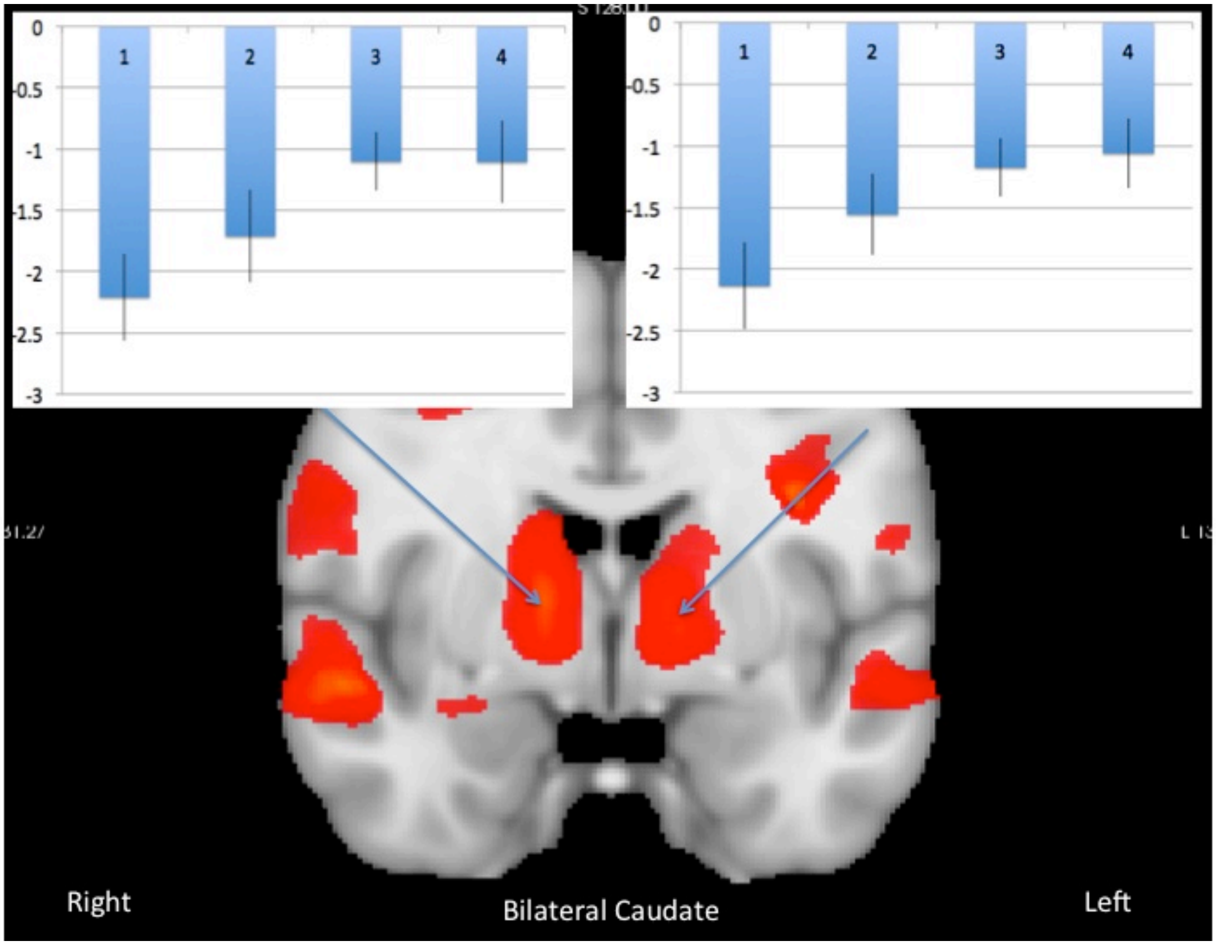

Figure 2. Bilateral Caudate Activation to Increasing Responder Uncertainty. Random effects (second level) analysis assessing for brain regions that scaled monotonically with increasing uncertainty (1 = certainty; 2 = low uncertainty; 3 = mid uncertainty; 4 = high uncertainty) during the decision phase (p < .001, FDR corrected). Error bars indicate 95% confidence intervals.

Appendix C

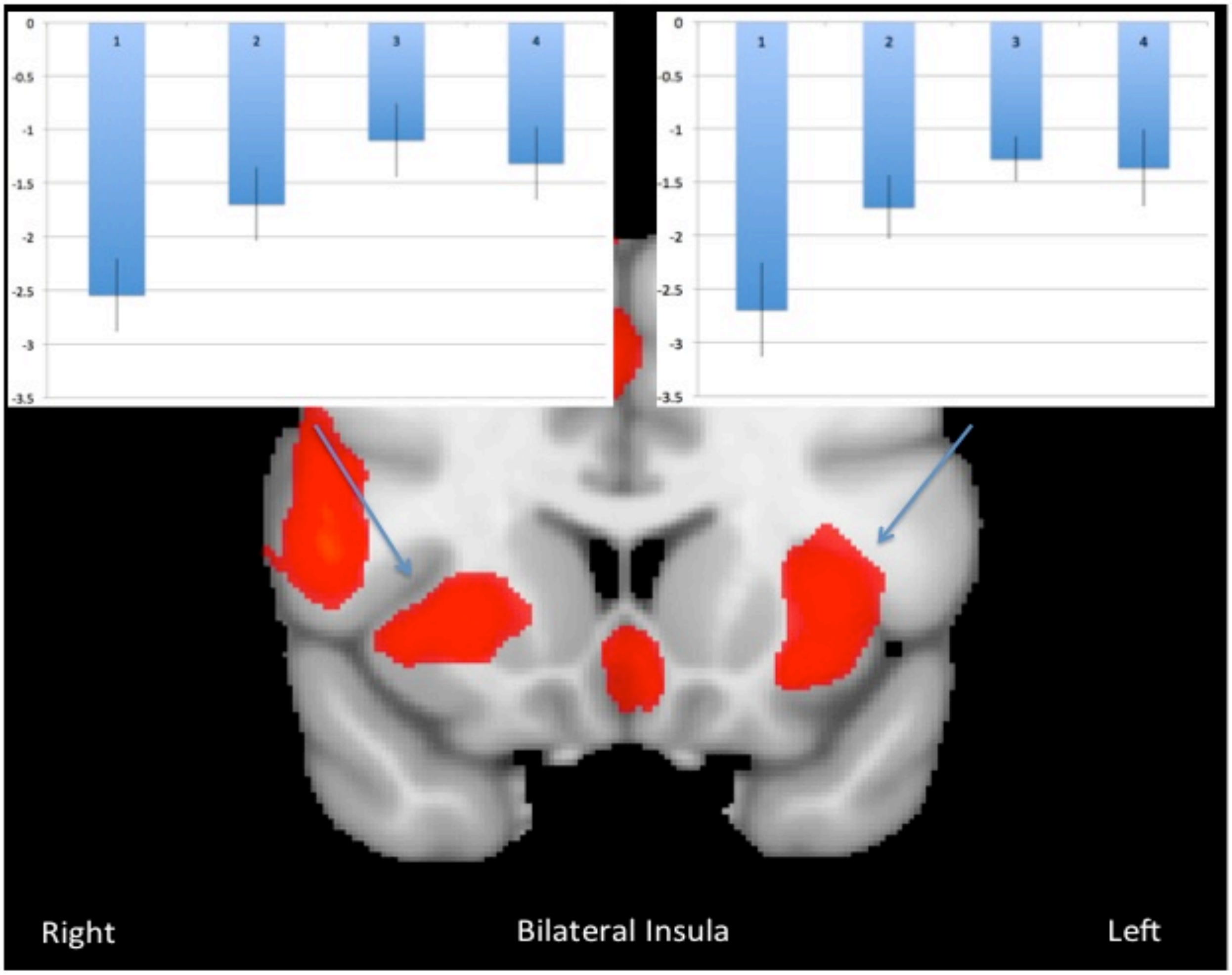

Figure 3. Bilateral Insula Activation to Increasing Responder Uncertainty. Random effects (second level) analysis assessing for brain regions that scaled monotonically with increasing uncertainty (1 = certainty; 2 = low uncertainty; 3 = mid uncertainty; 4 = high uncertainty) during the decision phase (p < .001, FDR corrected). Error bars indicate 95% confidence intervals.