Journal of Economic Behavior & Organization 94 (2013) 234–241

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Economic probes of mental function and the extraction of computational phenotypes

Kenneth T. Kishida[a,*], P. Read Montague[a,b,c]

[a] Human Neuroimaging Laboratory and Computational Psychiatry Unit, Virginia Tech Carilion Research Institute, Roanoke, VA 24016, USA
[b] Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[c] The Wellcome Trust Centre for Neuroimaging, University College London, WCN1 3BG, UK

ARTICLE INFO

Article history:
Received 13 April 2012
Received in revised form 22 July 2013
Accepted 23 July 2013
Available online 16 August 2013

Keywords:
Neuroeconomics
fMRI
Trust game
Mental disorders
Autism
Phenotype

ABSTRACT

Economic games are now routinely used to characterize human cognition across multiple dimensions. These games allow for effective computational modeling of mental function because they typically come equipped with notions of optimal play, which provide quantitatively prescribed target functions that can be tracked throughout an experiment. The combination of these games, computational models, and neuroimaging tools open up the possibility for new ways to characterize normal cognition and associated brain function. We propose that these tools may also be used to characterize *mental dysfunction*, such as that found in a range of psychiatric illnesses. We describe early efforts using a multi-round trust game to probe brain responses associated with healthy social exchange and review how this game has provided a novel and useful characterization of autism spectrum disorder. Lastly, we use the multi-round trust game as an example to discuss how these kinds of games could produce novel bases for representing healthy behavior and brain function and thus provide objectively identifiable subtypes within a broad spectrum of mental function.

1. Introduction

Theoretical approaches to the understanding of human decision-making (Von Neumann et al., 2007) have provided an excellent framework for ongoing empirical investigations, which measure actual human behavior against theoretically optimal actions (Camerer, 2003). The ability to measure brain responses, particularly with the use of functional magnetic resonance imaging (fMRI, Ogawa et al., 1990a, 1990b), associated with these behaviors has lead the development of biological investigations into the relationship between human biology and (ir)rational decision making (Montague and Berns, 2002; Loewenstein et al., 2008). The early revelation that humans do not always act in accord with economic theory and the ability to measure brain responses associated with these decisions are beginning to inform and reshape economic theories about human decision making (Camerer, 2003; Loewenstein et al., 2008). These developments are also giving rise to a new approach, i.e., computational psychiatry, to investigate mental disorders (Montague et al., 2012; Kishida et al., 2010);

* Corresponding author at: Human Neuroimaging Laboratory and Computational Psychiatry Unit, Virginia Tech Carilion Research Institute, 2 Riverside Circle, Roanoke, VA 24016, USA.
E-mail address: kenk@vtc.vt.edu (K.T. Kishida).

0167-2681
http://dx.doi.org/10.1016/j.jebo.2013.07.009

the motivation behind computational psychiatry is to generate objective, computationally framed depictions of unhealthy behavior and brain function associated with the wide range of psychiatric illnesses.

Psychiatric illnesses are brain disorders that 'reveal' their symptoms through aberrant decision-making and personal subjective turmoil. Unfortunately, the causes of these disorders have been extremely elusive; clinical and research efforts have been and continue to be hindered by the challenges associated with determining objectively identifiable symptoms. The use of computational approaches paired with game-theoretic probes and human neuroimaging promises to provide insight into the processes underlying human decision-making. These developments have the potential of generating a whole new perspective on the biological bases of human cognition and decision making by providing a novel entry point for the investigation and discovery of the biological architecture underlying human behavior.

Game theoretic probes provide a powerful framework for studying socially interacting agents where the strategies employed are guided by various concepts of optimal play. These games provide a natural landscape for the application of computational approaches and theoretical frameworks (like computational reinforcement learning theory, Sutton and Barto, 1998) to describe otherwise qualitative features of human experience like familiarity, fairness, or trust. Additionally, these games provide a good experimental setting for exploring features in our social environments that guide our behavior. Computational reinforcement learning theory (Sutton and Barto, 1998) provides a framework for investigating optimal reward harvesting and adaptive behavior and can readily be integrated into multi-round social exchange games (King-Casas et al., 2005). Reinforcement learning models capture notions of optimality in the context of decision-making in novel or changing environments and are flexible to what constitutes an agent, environment, and rewards. Within reinforcement learning approaches is the concept of a policy, which maps states to actions in order to maximize value. The use of these mathematical depictions of human behavior opens the door to new perspectives from which new dimensions of personality (i.e., styles of decision making) and their biological correlates may emerge. These quantitative depictions promise to be useful for characterizing normal and dysfunctional human cognition (Kishida et al., 2010) and can provide a relevant basis for identifying biological substrates important for human cognition at multiple levels of organization including social and individual behavioral, neurobiological and genetic systems.

In this memorial tribute to John Dickhaut, we focus on the use and development of the multi-round trust game (King-Casas et al., 2005, 2008; Weigelt and Camerer, 1988; Tomlin et al., 2006; Chiu et al., 2008), which was related to his and colleagues single round version of the game (Berg et al., 1995). We review early results from the application of the multi-round trust game and human neuroimaging to autism spectrum disorder. Additionally we describe the use of this game to classify behavior expressed in other mental disorders, such as borderline personality disorder (King-Casas et al., 2008) and depression (King-Casas et al., 2008; Koshelev et al., 2010). These early results suggest the ability to characterize human behavior and associated neural processes along new dimensions.

2. The single round trust game (Berg et al., 1995)

Berg et al. (1995) employed a single round game to investigate trust during economic exchange. In the single round trust game two players engage anonymously; there is an "investor" (first-mover) and a "trustee" (responder); the investor is endowed with $10 and decides an amount to share with their partner; the sent amount (i.e., "investment") is tripled on its way to the trustee; the trustee then decides how much, if any, to reciprocate to the investor. In the execution of this game the signals transmitted between the players is restricted to the money sent back and forth. As Berg et al., point out the Nash equilibrium for this game is for no money to initially be sent by the investor since a rational and selfish trustee will keep any money sent their way, thus to maximize ones earnings the selfish investor ought to keep everything. Contrary to this prediction, trust (money sent to the trustee) is observed as is reciprocation (money sent back to the investor) and the authors conclude that trust is likely a "behavioral primitive" (Berg et al., 1995) that maximize long-term genetic fitness over short-term gains through selfish behavior. This interpretation of their results is drawn in contrast to games with repeated interactions where trust can be learned or may show varying degrees of stability. An important point about their conclusion is the notion of a behavioral primitive; by expressing trust in a single interaction the results suggest that people carry around within them a bias toward trust and reciprocity. The authors do not propose in detail where such a bias may be stored; however, from a neurobiological perspective this bias must be engendered in the neural architecture both structurally and functionally and can be measured using the right tools. For example, recent findings suggest a genetic basis for the behavior expressed in this version of the game (Cesarini et al., 2008, 2009).

The single round trust game is also used by Berg et al. (1995) to investigate a "social history" manipulation wherein anonymous and naïve players are provided information about how previous participants played this game; this relatively mild manipulation was observed to increase trust suggesting that learning mechanisms and narratives that modulate expectations are also important in determining the expressed strategies.

Recent investigations employing fMRI have begun to investigate neural responses associated with the behavioral gestures exchanged within a multi-round version of the trust game (Weigelt and Camerer, 1988; King-Casas et al., 2005, 2008; Tomlin et al., 2006; Chiu et al., 2008). Additionally the use of the multi-round trust game and fMRI has been used to investigate neurobehavioral responses in populations characterized by clinically abnormal social behavior including participants diagnosed with autism spectrum disorder (Chiu et al., 2008) and borderline personality disorder (King-Casas et al., 2008). These studies demonstrate early developments in using game theory and computational approaches for understanding mental disorders (Kishida et al., 2010), which are believed to be strongly influenced by genetic predispositions. Below, we

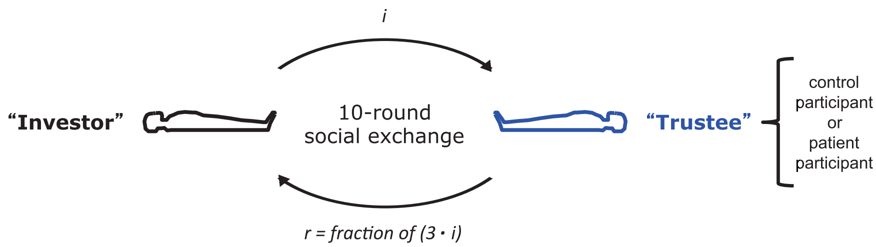

Fig. 1. Multi-round trust game probes social exchange in known psychopathological categories. The multi-round trust game (King-Casas et al., 2005; Tomlin et al., 2006) is a repeated interaction (10-round) version of the single round trust game (Weigelt and Camerer, 1988; Berg et al., 1995). An "investor" is given an initial endowment and is to choose how much to share with his/her partner. This investment "i" is tripled on its way to the "trustee". The trustee then chooses how much of "$3i$" to send back to the investor. The total points each player earns in a single round are placed into a "bank" and the game is repeated for a total of ten rounds. This deviation from the single round version allows the observation and measurement of reputation formation and learning signals embedded in this simple interaction. The multi-round trust game has been used to probe social exchange in a number of "patient" categories classified by DSM-IV criteria including: autism spectrum disorder, borderline personality disorder, and attention deficit hyperactivity disorder.

discuss the multi-round trust game, early neurobehavioral findings in autism spectrum disorder, and the direction this work may take in order to determine the degree to which game theoretic probes may be used to characterize mental illness.

3. Multi-round trust game and computational models of learning

The multi-round trust game (King-Casas et al., 2005; Weigelt and Camerer, 1988; Tomlin et al., 2006) allows the investigation of signals associated with iterated social exchange, including agent detection (Tomlin et al., 2006; Chiu et al., 2008), learning, and the development and expression of expectations (King-Casas et al., 2005). The initial development of the single round trust game (Berg et al., 1995) intended to reduce the effects of knowledge and reputation in order to examine the underlying bias regarding trust and selfish decision making, whereas the multi-round version aims to study these processes while eavesdropping on the underlying neural processes. Like the single round version two players engage anonymously; there is an "investor" (first-mover) and a "trustee" (responder); the first round is implemented in the same manner as the single-round version, however, subjects know that they will engage in a total of ten iterative rounds with the same partner (Fig. 1). This manipulation allows the study of signals sent between participants that know there will be feedback and a chance to respond to that feedback. It also allows the investigation of the modulation and development of internal models about the intentions and beliefs expressed between the two agents. Along these lines King-Casas et al. (2005) measured brain responses during the multi-round trust game using functional magnetic resonance imaging and identified brain responses consistent with reinforcement learning signals previously associated with dopaminergic neural activity.

King-Casas et al. identified these brain responses by taking advantage of the ability to quantify and computationally model the "social gestures" in the context of the game. Expressions of increases or decreases in trust are captured by changes in the values sent to ones' partner from one round to the next (Fig. 2). In early rounds, increases in trust by the trustee (i.e., increases in reciprocity) are preceded by an increase in a response (black trace, top right panel of Fig. 2) in the striatum (Fig. 2, left inset) following revelation of the amount of money sent from ones' partner. This response is consistent with reward-related processing of a social gesture leading to increased reciprocation of trust. On the other hand, subsequent decreases in trust are not preceded by an increase in striatal responses (red trace, top right panel of Fig. 2). Interestingly, in later rounds the striatal response becomes anticipatory and responds to the earliest phase of the trial where a positive signal can be predicted (black trace, bottom left panel of Fig. 2). Here the trustee brain may be predicting a positive signal and when the expectation is met an increase in trust is delivered. These results are consistent with reputation formation and the development of positive expectations of trust between the two partners. These results also suggest something more fundamental; the pattern of activity observed in the striatum matches very closely with learning dynamics previously observed in the dopaminergic system in non-human primates engaged in a simple Pavlovian learning paradigm (Montague et al., 1996; Schultz et al., 1997). The computational depiction of these simple learning signals predicts the observed temporal shift in the response pattern observed in the dopaminergic system and those observed in King-Casas et al. (2005) social exchange with brain imaging study. Further work suggests that the possibility that the dopaminergic system may serve as a common valuation system during learning and decision making in a wide range of valuation scenarios (Pagnoni et al., 2002; McClure et al., 2003; O'Doherty et al., 2003; Seymour et al., 2004; Fliessbach et al., 2007; Lohrenz et al., 2007; Behrens et al., 2008; Klucharev et al., 2009; Kishida et al., 2011; Kishida and Montague, 2012).

Biologically interesting participants can be anonymously interchanged in the trustee role to investigate how these individuals modulate the iterated dynamic exchange. This kind of manipulation has been carried out to investigate a range of psychopathologies including participants diagnosed with autism spectrum disorder, attention deficit hyperactivity disorder, borderline personality disorder, and major depression (Chiu et al., 2008; King-Casas et al., 2008; Koshelev et al., 2010).

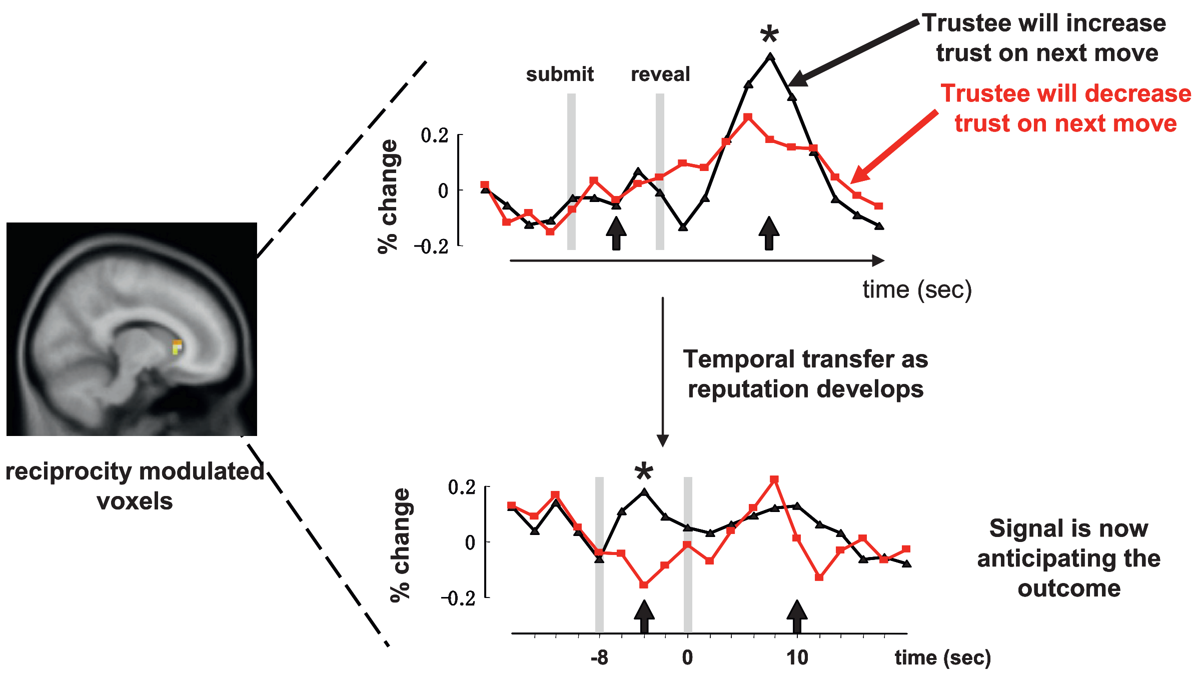

Fig. 2. Hyperscanning during two-person trust game reveals the development of signals for reputation formation (figure adapted from King-Casas et al., 2005). Left: Brain responses in the trustees' brain to "benevolent" investor behavior. Statistical parametric map showing significant activation in the bilateral head of the caudate nucleus in the trustees' brain for "better than expected" behavioral gestures from the investor (n = 125 gestures). Right: Neural correlates of reputation building. Blood-oxygen-level-dependent (BOLD) responses from the regions defined in the image on the left; time series of the BOLD response is time locked to the "investment" revelation, but separated according to the trustees' next decision (black: future increase in trust; red: future decrease in trust). In early rounds (top rows) a significant increase in the BOLD response in the caudate follows investment revelations that lead to the trustee increasing their trust in the next round (black trace). This signal undergoes a temporal transfer in later rounds (bottom rows) to just prior to investment revelation, which suggests that the trustee brain is anticipating trustworthy investments from the investor before they are revealed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

4. Autism spectrum disorder and borderline personality disorder in the multi-round trust game

These are still early days in the investigation of psychopathologies using computational approaches in game theoretic settings and neuroimaging (Kishida et al., 2010). However, there are already promising developments where different categories of psychopathology are showing differentiating strategies in game behavior and associated brain responses. Recent successes of the application of game theory to mental disorders include investigations into autism spectrum disorder (Fig. 3 adapted from Chiu et al., 2008) and borderline personality disorder (King-Casas et al., 2008).

Chiu and colleagues used the multi-round trust game and hyperscanning (Montague et al., 2002) to investigate social exchange in individuals diagnosed with autism spectrum disorder (ASD). ASD is characterized by deficits in social exchange and communication and reduced ability to infer the intentions of others. In Chiu et al., participants diagnosed with ASD were assigned to the trustee role and compared to age-matched participants also in the trustee role. These participants were relatively high functioning (as assessed by an estimate of their IQ) and repaid their investor quite similarly to age-matched controls in the multi-round trust game (Fig. 3A from Chiu et al., 2008). A challenging feature of the trust game is that participants must either possess or develop an accurate model of their partner in order to maximize their returns. Chiu et al. showed that a previously described agent-specific response in the cingulate cortex (Tomlin et al., 2006) was diminished in the ASD cohort (Fig. 3B from Chiu et al., 2008). Specifically, a spatial pattern of activity dubbed the "cingulate self response", which was observed in contrast to the "cingulate other response" (Tomlin et al., 2006), was shown to be diminished proportional to the participants symptom severity (Fig. 3B from Chiu et al., 2008). Further work suggests that the cingulate self response pattern, which was only observed during real social exchange (versus simulated game play), is associated with perspective-taking (Chiu et al., 2008).

The relatively cooperative behavior typically observed in the trust game suggests that players share norms about fairness in these kinds of exchanges and reciprocation of trust appears to be normal behavior. This normative observation suggests that some psychopathologies may be more or less sensitive to signals and calculations of fairness and equitable distributions. These signals may be derived from initially shared expectations of cooperation thus resulting in expectations of high investment. Deviations from these expectations may result in social exchange dynamics that lead to the break down in trust; problems can also arise when partners do not share the same model of what pro-cooperative signals look like. King-Casas et al. (2008) studied multi-round trust game behavior and the associated neural responses in individuals diagnosed with borderline personality disorder (BPD). Individuals diagnosed with BPD demonstrate pervasive instability of interpersonal

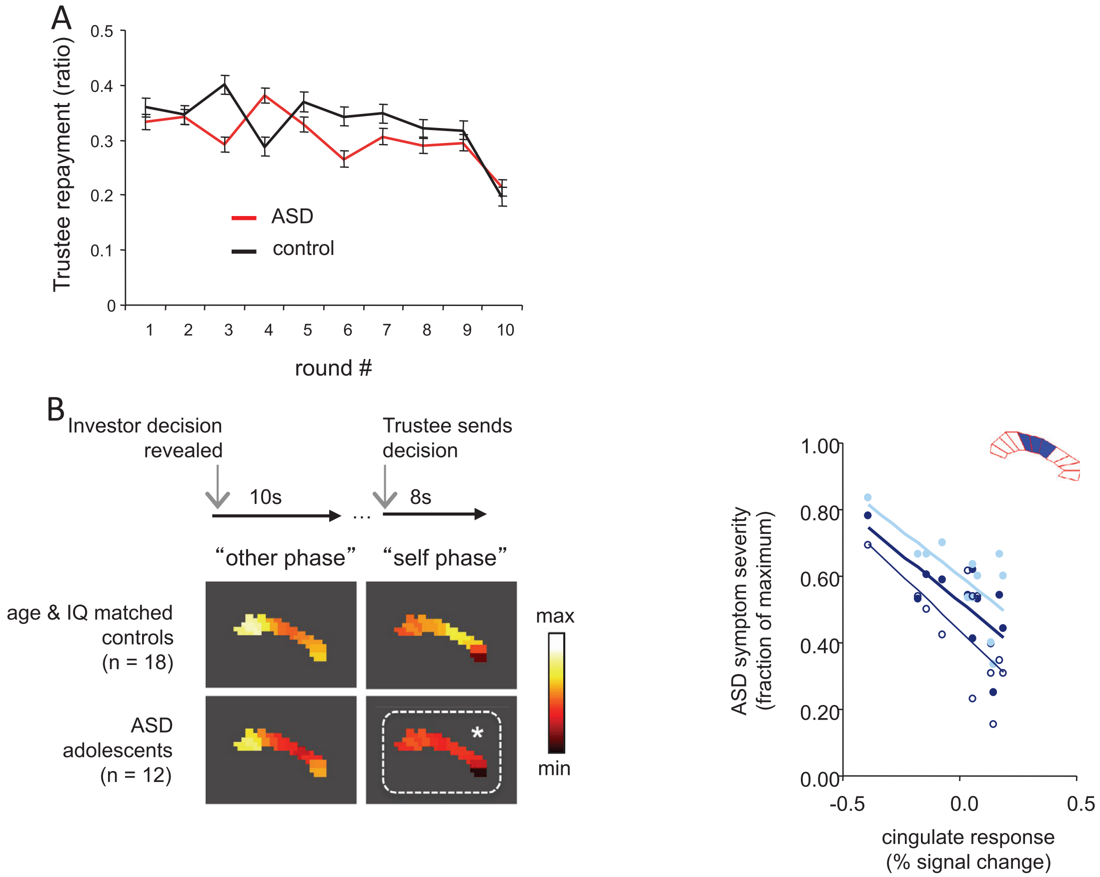

Fig. 3. Multi-round trust game reveals diminished cingulate response in participants diagnosed with autism spectrum disorders (adapted from Chiu et al., 2008). (A) Average trustee repayment ratio round-by-round. The repayment ratios are not significantly different round-by-round in ASD participants compared to controls. (B) Diminished cingulate response pattern during "self phase" of the iterated multi-round trust game. Left: heat maps showing spatial pattern of activity indicative of self- and other-responses during the multi-round trust game (Tomlin et al., 2006), where the cingulate self-response is revealed to be specifically diminished in individuals diagnosed with ASD (see response labeled with white asterisk). Right: the magnitude of signal change in the middle portions of the cingulate cortex during the self-response phase of the task show significant correlation with the assessment of ASD symptom severity (Chiu et al., 2008) (open circles: ADI communication subscale, $r = -0.69$, $p = -0.012$; light blue filled circles: ADI social subscale, $r = -0.70$, $p = 0.011$; dark blue filled circles: ADI total score, $r = -0.73$, $p = 0.007$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

relationships, self-image, and affect, which begins early in adult life (American Psychiatric Association, 2000). In the King-Casas et al. study, the control population (trustee role) showed brain responses in the insula cortex that correlated with diminishing investments from the investor participants. This is consistent with the insula detecting norm violations in a range of experimental paradigms (Montague and Lohrenz, 2007). Interestingly participants diagnosed with BPD showed no parametric relationship between the size of the offer and their insula response, rather their insula was consistently activated by all offer sizes (King-Casas et al., 2008). This neural strategy was associated with an unwillingness to cooperate over the multi-round interactions. Additionally, investors (unknowingly) playing with a BPD trustee were nearly half as likely to send gestures consistent with coaxing behavior (King-Casas et al., 2008). Coaxing is typically seen in pairs of healthy participants in response to low offers and is considered to be an attempt to draw the partner into a cooperative mode (King-Casas et al., 2008); the data suggest that the relatively low bandwidth signaling afforded by the trust game setup provided enough information to the investor to alter the participants' behavior in a meaningful way.

5. Application of computational approaches to patient behavior in simple exchange games

The value of computational approaches extends beyond the basic science underlying choice behavior in humans. Framing experimental paradigms in mathematical theory may provide access to parameters and new concepts that are not directly available to our conscious psyche. The fruits of these maneuvers are beginning to express themselves as new insight into long standing issues in psychiatric populations.

The multi-round trust game (King-Casas et al., 2005; Tomlin et al., 2006) has been employed to investigate neural and behavioral responses in a number of psychiatric populations (see Fig. 4A, top) including autism spectrum disorder (ASD) (Chiu

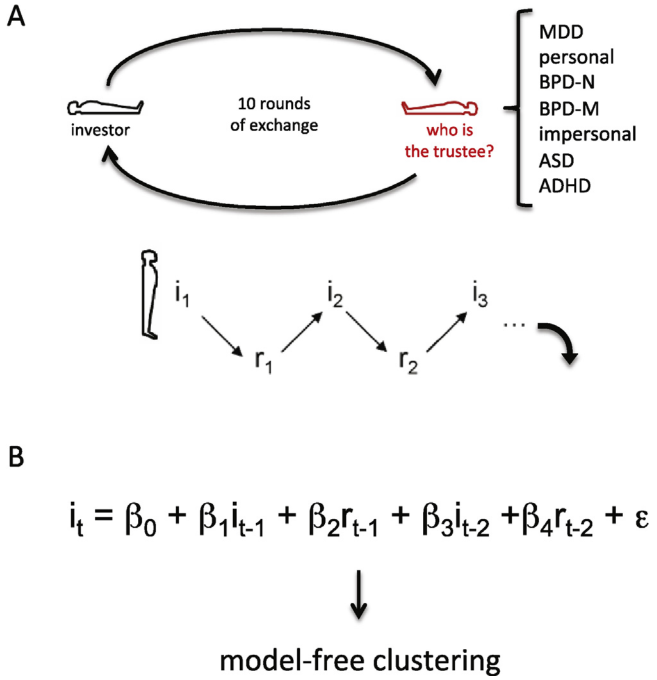

Fig. 4. Classification of trustee "type" from investors' behavior in two-party exchange. (A and B) Depiction of model free clustering approach using multi-round trust game data. The data used in this approach was collected in previous studies (King-Casas et al., 2005, 2008; Tomlin et al., 2006; Chiu et al., 2008; Koshelev et al., 2010). (A) The multi round trust game is played between a healthy investor (black player, left) and a "target" trustee (red player, right). The "target" trustee was one of the following "types": major depressive disorder (MDD), personal (Tomlin et al., 2006), borderline personality disorder-non-medicated (BPD-N, King-Casas et al., 2008), borderline personality disorder-medicated (BPD-M, King-Casas et al., 2008), impersonal (King-Casas et al., 2005), autism spectrum disorder (ASD, Chiu et al., 2008), and attention deficit hyperactivity disorder (ADHD, Tomlin et al., 2006). The approach described in detail in Koshelev et al. (2010) examines the investor behavior as a polynomial of past rounds of investments and returns (see panel B). $i_1, i_2, \ldots, i_t$ are the investments made by the investor during round t. Likewise, $r_1, r_2, \ldots, r_t$ are the repayments made by the trustee during round t. (B) Classification of the investor-trustee dyad is performed by predicting the investors' decision at round t using a polynomial where the order of the polynomial, the number of past rounds, and the number of clusters discovered are left as free parameters to be discovered. The diagnostic categories for the trustee "type" listed in panel A are blinded in this classification procedure. Only the behavior (investments and repayments over rounds) in the multi-round trust game is used. The result of this classification determined that a 1st-order polynomial, 2 rounds back, and 4 clusters were optimal (figure adapted from Koshelev et al., 2010). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

et al., 2008), major depressive disorder (MDD), borderline personality disorder (BPD) (King-Casas et al., 2008), and attention deficit hyperactivity disorder (ADHD). In each of the studies listed above, the investor was a healthy adult volunteer while the role of the trustee was filled with an anonymous patient who met DSM-IV diagnostic criteria. The effect of maintaining anonymity of participants (Fig. 4A: impersonal) or allowing the participants to meet prior to playing (Fig. 4A: personal) has also been investigated (King-Casas et al., 2005, 2008; Tomlin et al., 2006). One powerful applications of this reduced form of social exchange is highlighted by recent work where the trustee type was blindly classified based on parameters discovered in data that predict investors' decisions while engaged with a patient (Fig. 4) (Koshelev et al., 2010). In this work, data from all participant pairs (i.e., MDD, personal, BPD, impersonal, ASD, and ADHD) were pooled together and clustered using a model-free approach (Fig. 4).

The approach used by Koshelev and colleagues, is computationally involved, but the idea behind it is rather straightforward. Objectively measured behavioral signals (iterated investments and repayments) can be used to classify dyad "types", for instance healthy pairs versus pairs consisting of a healthy investor and a patient. Furthermore there may be subtleties in the interaction that will allow the differentiation of patient populations (e.g., participants diagnosed with ASD may play differently than participants diagnosed with BPD). Finally, and a distinguishing feature of the approach, the differences can be read out by looking at differences in how the *healthy investor responds* to the various patient populations; this highlights the biosensor approach described in their report. Briefly, the investor's decision at round "t" (i_t) is predicted using a polynomial, which incorporates previous investment (i_{t-n}) and repayment (r_{t-n}) decisions by both partners (Fig. 4B). The number of "rounds back" ("n"), the order of the polynomial, and the number of clusters to be identified were all left as free parameters to be discovered in their approach. This approach identified four clusters where patient subgroups were either over- or under-represented (Koshelev et al., 2010).

Computational approaches have also begun to investigate choice behavior in social games using participants diagnosed with autism spectrum disorder (Chiu et al., 2008; Yoshida et al., 2010). Chiu et al. (2008) used the multi-round trust game to investigate neurobehavioral responses elicited by participants diagnosed with ASD playing in the trustee role. The data demonstrated that responses in the middle cingulate, that were elicited while subjects sent a signal to their partner, were diminished in individuals diagnosed with ASD. These responses did not vary with the magnitude of points earned or sent (Tomlin et al., 2006), but were absent in healthy trustees playing the game in the absence of a social agent in the investor role (i.e., playing a computer) (Tomlin et al., 2006; Chiu et al., 2008). Yoshida and colleagues have investigated social behavior using computational approaches by developing a "computational game theory of mind" approach (Yoshida et al., 2008) where they use Bayesian approaches to model behavior in a two party coordination game (Yoshida et al., 2008, 2010). Parameters inferred from the behavioral data measured in this game also demonstrate the ability to classify healthy participants from those diagnosed with ASD as players attempt to maximize their reward in a game that requires cooperation for maximal payoff (Yoshida et al., 2010).

6. Can game-theoretic probes shed new light for phenotyping humans?

Game theory has provided a powerful framework for considering how an idealized agent ought to behave in highly structured social exchange games (Von Neumann et al., 2007). Indeed, the framing of experiments by game theory has already provided much insight into principles that characterize human decision making, but more interesting are the cases where human behavior deviates from economic theory (Camerer, 2003). Additionally, game theoretic approaches have provided interesting insights into the evolution of non-human organisms and the strategies they employ (Smith, 1982; Smith and Harper, 2003). Human neuroimaging results suggest that the presumed valuation machinery in human brains does not only respond to monetary gains and losses, which are an important guidance signals in economic theory; the data suggest that primary rewards and social status (in the absence of monetary gains and losses) can also engage the same neural machinery. This points to a biologically guided valuation system where money is likely a proxy or a cue for something more fundamental. This idea is consistent with the approach taken by evolutionary biologists where genetic fitness is the guiding principle; strategies (i.e., phenotypes) are selected that maximize genetic fitness. Certainly humans are not exempt from the pressures of natural selection. Our genome and the biological processes it dictates have resulted from generations of successful strategies defined by increased success in reproducing and surviving.

We have presented a brief review of quantitative neural and behavioral results that identify specific deviations from typical behavior in the context of the multi-round trust game. The diminished cingulate response in ASD subjects is specific to this population (BPD patients do not show a deviation in this response). Our early results in this domain suggest that games and the expression of strategies in human populations can be fruitfully explored and may lead to the characterization of normative strategies in human behavior and associated brain responses. These normative descriptions can be exposed in experiments that sample strategic decision-making in large samples of human populations. Measuring the distribution of any quantitative trait within these games will begin to characterize what would be considered normal/healthy human cognition within these dimensions and would set the stage for identifying subpopulations of aberrant decision-making phenotypes. Qualitatively, the DSM criteria for mental disorders achieve this, but without the quantitative rigor or objective threshold criteria that game theoretic approaches promise and that genetic discoveries likely require. Using a naturally quantitative framework to characterize choice behavior in patients with mental disorders will allow computational tools – relatively new to the investigation of mental health – to generate a powerful and novel perspective on an old problem. The reduction of human personality and subjective experience as expressed through our choices in the context of strategic games is an exciting and relatively unexplored direction for investigating the biological basis for human psychopathology.

Acknowledgments

This work was funded by the Wellcome Trust Principal Research Fellowship (PRM), The Kane Family Foundation (PRM) and the National Institutes of Health, R01-NS045790 (PRM), R01-DA11723 (PRM). The authors have no conflict of interest, financial or otherwise, related directly or indirectly to the work discussed herein.

References

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, 4th ed., text revision. American Psychiatric Association, Washington, DC.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2008. Associative learning of social value. Nature 456, 245.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.

Camerer, C., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Russell Sage Foundation, New York.

Cesarini, D., Dawes, C.T., Fowler, J.H., Johannesson, M., Lichtenstein, P., Wallace, B., 2008. Heritability of cooperative behavior in the trust game. Proc. Natl. Acad. Sci. U.S.A. 105, 3721–3726.

Cesarini, D., Dawes, C.T., Johannesson, M., Lichtenstein, P., Wallace, B., 2009. Experimental game theory and behavior genetics. Ann. N. Y. Acad. Sci. 1167, 66–75.

Chiu, P.H., Kayali, M.A., Kishida, K.T., Tomlin, D., Klinger, L.G., Klinger, M.R., et al., 2008. Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. Neuron 57, 463–473.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C.E., et al., 2007. Social comparison affects reward-related brain activity in the human ventral striatum. Science 318, 1305–1308.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2005. Getting to know you: reputation and trust in a two-person economic exchange. Science 308, 78–83.

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 321, 806.

Kishida, K.T., Montague, P.R., 2012. Imaging models of valuation during social interaction in humans. Biol. Psychiatry 72 (2), 93–100, http://dx.doi.org/10.1016/j.biopsych.2012.02.037.
Kishida, K.T., King-Casas, B., Montague, P.R., 2010. Neuroeconomic approaches to mental disorders. Neuron 67, 543–554.
Kishida, K.T., Sandberg, S.G., Lohrenz, T., Comair, Y.G., Sáez, I., Phillips, P.E.M., et al., 2011. Sub-second dopamine detection in human striatum. PloS ONE 6, e23291.
Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A., Fernandez, G., 2009. Reinforcement learning signal predicts social conformity. Neuron 61, 140–151.
Koshelev, M., Lohrenz, T., Vannucci, M., Montague, P.R., 2010. Biosensor approach to psychopathology classification. PLoS Comput. Biol. 6, e1000966.
Loewenstein, G., Rick, S., Cohen, J.D., 2008. Neuroeconomics. Annu. Rev. Psychol. 59, 647–672.
Lohrenz, T., McCabe, K., Camerer, C.F., 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., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron 38, 339–346.
Montague, P.R., Berns, G.S., 2002. Neural economics and the biological substrates of valuation. Neuron 36, 265–284.
Montague, P.R., Lohrenz, T., 2007. To detect and correct: norm violations and their enforcement. Neuron 56, 14–18.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.
Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., et al., 2002. Hyperscanning: simultaneous fMRI during linked social interactions. Neuroimage 16, 1159–1164.
Montague, P.R., Dolan, R.J., Friston, K.J., Dayan, P., 2011. Computational psychiatry. Trends Cogn Sci. 16 (1), 72–80.
O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38, 329–337.
Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990a. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U.S.A. 87, 9868–9872.
Ogawa, S., Lee, T.M., Nayak, A.S., Glynn, P., 1990b. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn. Reson. Med. 14, 68–78.
Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S., 2002. Activity in human ventral striatum locked to errors of reward prediction. Nat. Neurosci. 5, 97–98.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599.
Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., et al., 2004. Temporal difference models describe higher-order learning in humans. Nature 429, 664–667.
Smith, J.M., 1982. Evolution and the Theory of Games. Cambridge University Press.
Smith, J.M., Harper, D., 2003. Animal Signals. Oxford University Press, USA.
Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction. Cambridge University Press.
Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F., Quartz, S.R., et al., 2006. Agent-specific responses in the cingulate cortex during economic exchanges. Science 312, 1047–1050.
Von Neumann, J., Morgenstern, O., Rubinstein, A., Kuhn, H.W., 2007. Theory of Games and Economic Behavior, 16th ed. Princeton University Press.
Weigelt, K., Camerer, C., 1988. Reputation and corporate strategy: a review of recent theory and applications. Strategic Manage. J. 9, 443–454.
Yoshida, W., Dolan, R.J., Friston, K.J., 2008. Game theory of mind. PLoS Comput. Biol. 4, e1000254.
Yoshida, W., Dziobek, I., Kliemann, D., Heekeren, H.R., Friston, K.J., Dolan, R.J., 2010. Cooperation and heterogeneity of the autistic mind. J. Neurosci. 30, 8815.