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/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
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