NEURAL SIGNATURES OF TRUST IN RECIPROCITY

by

Sergey V. Chernyak
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Neuroscience

Committee:

_____	Dr. Frank Krueger, Dissertation Director
_____	Dr. Kevin A. McCabe, Committee Member
_____	Dr. William G. Kennedy, Committee Member
_____	Dr. Gopikrishna Deshpande, Committee Member
_____	Dr. S. David Wu, Director, Krasnow Institute for Advanced Study
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs and Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Spring Semester 2016 George Mason University Fairfax, VA

Neural Signatures of Trust in Reciprocity

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Sergey V. Chernyak
Master of Arts
George Mason University, 2011

Director: Frank Krueger, Professor
Molecular Neuroscience Department

Spring Semester 2016
George Mason University
Fairfax, VA

DEDICATION

This work is dedicated to my precious wife Victoria, whose loving sacrifices and confidence converted me into a believer in unconditional trust and the American dream and to my beloved parents and grandmother, whose eager push for higher education has always motivated me to realize my full potential.

ACKNOWLEDGEMENTS

I would like to thank the many supporters who have made this dissertation possible, but first and foremost my dear wife, Victoria, who, among many loving sacrifices she made, had the courage to put our family budget and her own career ambitions on the line by strongly insisting that I swap my unexciting employment for the thrill of becoming a Ph.D. student. I am grateful to my advisor Dr. Frank Krueger for opening this window of opportunity for me one day and for supporting and guiding me in my research in the years that followed. I'd like to thank my committee members Drs. Gopikrishna Deshpande, Kevin McCabe and William Kennedy, for taking time from their busy schedules to share valuable expertise and advice in support of this dissertation. Last but not least, I'd like to thank my beloved parents for their unconditional love and support and for being the paragons of good education, thorough knowledge and refined culture for all these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Activation Likelihood Estimation ..... ALE
Cingulate Cortex (anterior and posterior) ..... ACC, PCC
Anterior Insular Cortex ..... AI
Blood Oxygenation Level Dependent ..... BOLD
Brodmann Area ..... BA
Caudate Nucleus (body and head) ..... BCd, HCd
Cerebellum ..... CB
Effective Connectivity Analysis ..... ECA
Extrastriate Visual Area V2
False Discovery Rate ..... FDR
Frontal Operculum ..... fO
Functional Magnetic Resonance Imaging ..... fMRI
Globus Pallidus ..... GP
General Linear Model ..... GLM
Hemodynamic Response ..... HR
Hippocampus ..... Hip
Hypothalamus ..... Hyp
Investment Game ..... IG
Lateral Intraparietal Area ..... Area LIP
Montreal Neurological Institute ..... MNI
Multivariate Autoregressive Model ..... MVAR
Multivariate Granger Causality ..... MVGC
Nucleus Accumbens ..... NAc
Orbitofrontal Cortex (lateral) ..... OFC (lOFC)
Parietal Lobule (inferior and superior) ..... IPL, SPL
Prediction Error ..... PE
Precuneus ..... PCu
Prefrontal Cortex (medial, rostrolateral) ..... PFC (mPFC, rlPFC)
Prefrontal Cortex (ventrolateral, dorsolateral) ..... vlPFC, dlPFC
Prefrontal Cortex (ventromedial, dorsomedial) ..... vmPFC, dmPFC
Premotor Cortex ..... PMC
Primary Visual Cortex V1
Pulvinar (Thalamus) ..... Pul
Putamen ..... Pu
Region of Interest ..... ROI
Supplementary Motor Area ..... SMA

ABSTRACT

NEURAL SIGNATURES OF TRUST IN RECIPROCITY

Sergey V. Chernyak, Ph.D.

George Mason University, 2016

Dissertation Director: Dr. Frank Krueger

Trust facilitates conditions for safe sharing of valued resources – a social setting vital to success in a wide range of socio-technological networks. With an increasing reliance of economic initiatives on trust-assured interactions, the need to inquire into the mental processes of trust has emerged. This led to a proliferation of functional magnetic resonance imaging (fMRI) studies focusing on domain-specific measures. However, inherent metric deficits of fMRI have resulted in discrete outcomes highlighting further need for constructing a comprehensive neurocognitive model of trust. Here, a domain-general methodology aims at overcoming the negative tendencies in prior fMRI studies by applying a series of coordinate-based "Activation Likelihood Estimation" (ALE) meta-analyses of the fMRI data and a data-driven Multivariate Granger Causality (MVGC) connectivity analysis of hyperscan-fMRI data – an approach not undertaken in trust studies prior to this dissertation. To determine the effects on behavior of cross-study variability in brain activation during a trust-inducing investment game (IG) task, the meta-analysis aims at revealing the extent of neurocognitive differentiation during trust, learning to trust and

reciprocity. One-shot IG, implicating unconditional trust, is compared to multi-round IG implicating the conditional trust. In the MVGC study, the neurocognitive differences in the effective connectivity of interpersonal ("brain-to-brain") trust are discerned. Meta-analysis revealed a strong differential response between unconditional trust (ambiguity, insula) and conditional trust (reward, ventral striatum). Learning to trust engaged a goal-guided (rostrolateral PFC) transition between decision-making (dorsal striatum, action-valuation) and feedback processing (ventral striatum, reward reinforcement). Reciprocating trust was linked to insula-mediated norm-compliance tendency to avoid breaking trust. For the effective connectivity analysis, a steady increase in trust and reciprocity engaged a mentalizing network as evidenced in the observed dorsal PFC connectivity with the parietal cortex. Within-trustor, dorsomedial (dmPFC) bidirectional connectivity with posterior cingulate cortex was key to guiding trust-valued choices (hypothalamus). Within-trustee, the key motive of norm-compliance – trustworthiness (lateral orbitofrontal cortex) was mediated by dorsomedial and dorsolateral PFC in Stage 1 and by precuneus in Stage 2. For the brain-to-brain exchange, the trustor's dmPFC was most active, but the trustee's dmPFC was virtually absent indicating dissociable patterns of other-regarding preferences for the trustor and trustee. Collectively, this dissertation lends evidence consistent with the putative socio-cognitive, economic-utility and reinforcement-learning models of trust and opens new perspectives by applying an effective domain-general data-driven dynamic approach.

CHAPTER 1. General Introduction

Section 1.1 Scope

The overall objective of this dissertation is to examine how variability in human mental and brain activity affects trust behavior. The specific aim is to resolve a long-standing issue in today's trust fMRI research. FMRI is a good method for measuring segregated neural components of a complex cognitive function but the segregation is an obstacle to an abstraction and generalization of data. At the time that the results of this research are being published, a several dozen of studies have engaged in delineating neurocognitive substrates of trust. Although since the late 1980's this field of study has matured at a great pace and has gone a long way from early theoretical assumptions [1] to the ability to quantify and reliably measure the fundamentally unobservable properties of trust, the need for consolidation of the discrete findings has recently emerged. In meeting this objective, this dissertation relied on strong and diverse background in cognitive psychology, neurobiology and neuroeconomics for analyzing the problem domain and searching for solutions. The interdisciplinary nature of the science at hand is reflected in the flow of the upcoming chapter, which is broken into three components. Section 1.1 is a review of what is known to date about the cognitive, neurobiological and economic origins of trust. Section 1.2 examines the standard research methods that are used for measuring trust. Section 1.3 is focused on the proposal of novel operational instruments for solving some of the outstanding problems the field is currently facing.

Section 1.2 Background

Subsection 1.2.1 Socio-Cognitive Perspective

For demarcating the boundaries of the problem domain to be investigated in this dissertation the socio-cognitive theory of trust has been especially helpful [2]. The socio-cognitive model of trust is based on the notion of a relationship between a human cognitive agent and his environment. The agent relies on the environment to satisfy a wide range of personal needs and aims to maximize gains but also minimize losses. In doing so, he tries to liberate himself from less relevant, less pressing tasks and delegate ("trust") them to some other party ("trustee"). Yet, he is faced with a possibility that the trustee might lack motivation, awareness or skill to accomplish what's needed. In face of the dilemma, he is confronted with a choice of either accepting the risk in favor of a potentially positive outcome in the future or staying on the safe side and withholding the delegation.

A cursory glance at the scenario above highlights a number of important properties of the trust construct to be examined here. The key notion is that trust is a goal-driven, reputation-guided process of delegating valued resources to others. A rationale for delegating can be economy, convenience, lack of skill or a need to realign resources and time for other activities. In general, when someone is trusted, it is expected that the trustee is willing, competent and equipped to accomplish the entrusted tasks. Establishing a trust-based partnership facilitates favorable conditions for safe and reliable sharing and even augmenting of valued resources with trusties. Such partnership is predicated on two basic forms of trust – unconditional (dispositional, implicit) and conditional (trust in reci-

procity). Unconditional trust occurs at the onset of the relationship and can be viewed as collusion between a trustor inclination for trusting and a trustee inclination for honoring trust. In contrast, conditional trust emerges from an ongoing exchange with the trustee and cannot be built without experience or/and theory of mind. Yet, knowing the genuine intentions of the trustee is seldom possible, which makes trust an ambiguous and vulnerable endeavor. Therefore, trust can be viewed as a state of mental ambiguity caused by a quest for reciprocity and at the same time, by a desire to avoid betrayal. This is because trust is about rendering control over a valuable resource at a cost of risking betrayal or failure. Choosing to trust comes with a price of accepting the vulnerability for the sake of improved outcomes in the future.

At a more granular level, trust can be viewed as a belief (i.e., a certain degree of faith in the trustee's trustworthiness) or action (decision). In trust, decisions are always consequential of the beliefs and the beliefs are mostly shaped by how much the trustor knows about the trustee. An uninformed or partially informed trustor is faced with having to rely on a merely dispositional or generalized belief in others' good will. Such a belief is inevitably formed from his experience with others, but not with the trustee. Confidence in a particular trustee can only be achieved in an ongoing person-to-person experience with that trustee, in which the trustor learns from the trustee's responses and becomes increasingly knowledgeable in the process. In reference to unconditional vs. conditional trust, the uninformed trustor is more associated with the former, while the informed or fully knowledgeable trustor with the latter. It follows that unconditional trust can be characterized by high level of ambiguity, while the conditional trust facilitates confidence,

progressively assured by the trustee response. Although valence of the response can be negative at times, assurance of trust can only be determined by an overall positive experience with the trustee. If trustor's goals are supported, the confidence in the trustee increases and consolidates the social structure. Betrayal of trust on the other hand, poses a cognitive conflict between the agents [3]. Conditional trust is therefore a product of inferences a trustor makes about the social characteristics of the trustee, such as competence, benevolence and honesty [4]. While modulated by valence of the trustee response, a trusting decision is predicated on inferring opponents' intentions [5]. Mentally modeling the opponent's future actions strengthens the likelihood of reciprocity to trust [6] and offsets the feeling of ambiguity and risk [7]. Having to mentalize defines trust as a relational state, because building trust requires interaction with other agents. Trust tends to emerge from a resource or information exchange whose purpose is to realize a cognitive task [8]. Arriving at a trust decision in an interaction is an incremental process – a cascading sequence of two dissociable phases, "belief formation" and "choice" (decision-making). In the belief phase, trust forms as an essentially evaluative judgment of the trustee's trustworthiness – "a deliberative process of weighing incomplete or ambiguous evidence" derived from the experience with the trustee [9]. In the decision phase, the trustor chooses whether to proceed or to refrain from delegation. Degree, to which the agents share goals and positivity of their attitudes, determines reciprocity between the agents.

Subsection 1.2.2 Neurobiological Perspective

The socio-cognitive perspective discussed above provides an aerial view of trust as a construct and emphasizes beliefs as governing force of the behavior. A related body

of work in neurobiology investigates how neuronal populations disambiguate evidence of competing stimuli (options) in favor of a particular action – a process commonly referred to as decision-making. For neurobiology, the primary concern is how the brain operates to produce decisions and how decision-making takes place in the nervous system, whereby a neuronal signal is being transformed on its way from a sensation to action. The question is how the signal can be traced along its way through the nervous system in its reflection of the stimulus and its production of behavior (decisions). Decisions emerge as a series of sensations induced by the events and stimuli in the outside world. The environmental and social events are neuronally linked in the perception and behavioral responses. The focus is on the notion of stimuli "disambiguation", whereby for a series of sensory inputs, the job of the nervous system is to produce proper behavior by the virtue of gating sensory signals to the "proper" action selection centers in the brain. Thus, neurobiology brings trust research to an important juncture between the studies of "choice" and "valuation" and to the idea that these two fundamental components of decision-making are supported by dissociable brain systems.

When making choice, a decision-maker is confronted with an ambiguous signal. Disambiguation in the brain is achieved by simultaneous step-by-step sampling of stimuli parameters related to every option. At some point in time, one of the options is "favored" more than others and the combined neuronal activation (integration) reaches a threshold for response, committing the decision to the favored option [10,11]. Sampling of stimuli and integration of evidence are supported by dissociable neural substrates. The sequential sampling and logical comparison of stimuli are known to be modulated by lateral pre-

frontal cortex (PFC) [12], while the "ramping-to-threshold" activity [13] is mediated by lateral parietal cortex [14] and caudate [15,16]. The brain mechanism of disambiguating decisions is now well understood. Yet, the model stops short of accounting for the effects of "utility" ("value") – a defining factor in any decision-making process, according to the neuroeconomic perspective discussed in the next section of the chapter.

Subsection 1.2.3 Neuroeconomic Perspective

Neuroeconomics provides an important perspective on trust research by introducing the notions of "utility" and by giving the notions of "ambiguity" and "choice" a deeper meaning. Strong evidence from neuroeconomics has shown, that some kind of reward utility or aversion to loss of reward are essential to any kind of decision-making, including trust [17]. Converging evidence has emerged for the existence of distinct frontal and striatal signaling that collectively generate subjective value for a wide range of rewards [18]. Studies of appetitive stimuli for example, point at a brain network with medial PFC (mPFC) as focus of critical social value inputs from dorsolateral PFC (dlPFC) [19] orbitofrontal cortex (OFC) [20], hypothalamus [21], amygdala [22,23] and the striatum [24]. Studies of the aversive stimuli implicated the anterior insula in the function of negative valuation and avoidance of negative outcomes.

Today it is well known, that areas key to control of choice are localized to intraparietal and motor cortex, while areas key to reward valuation are localized to ventromedial PFC (vmPFC) – the site where idiosyncratic values are placed on a wide range of rewards [25]. However, in spite of the evidence, little is known about the interaction across the frontoparietal (choice), medial prefrontal (valuation) and cingulo-opercular (aversion) neural

circuitry and whether the circuits form a tightly interconnected global network or a set of loosely coupled specialized functional modules. Behavioral measures together with the evidence of intrinsic brain activity shall provide the necessary insight into the brain mechanisms of both the valuation and choice phases of decision-making.

Another important contribution of neuroeconomics was analyzing the impact of expected utility and ambiguity of intertemporal choice on decision-making and recognizing ambiguity as one of the primary challenges decision-makers face in several cognitive domains. For a trustor, the main source of ambiguity is the uncertainty between positive and negative prospects of trusting i.e., between possibility of reward [26] and possibility of betrayal [27], between the immediate gains and long-term goals [28], between perceived probability of reward and subjective value [29] and between self-interest and other-regarding preferences [30]. In the literature, assessing rewards and making choices are viewed as two fundamental constituents of decision-making. However, it's not at all clear how the two communicate with each other to produce a unified behavior. In the upcoming chapters the possibility of a functional link via two other components of decision-making i.e., ambiguity (Chapter 2) and mentalizing (Chapter 3) will be examined in depth.

Section 1.3 Metrics of Trust

Subsection 1.3.1 Game Theory (Investment Game)

The success of neuroeconomic studies of expected utility affected trust research in a fundamental way. It enabled the assumptions about economic utility and ambiguity to be experimentally tested in a probabilistic interactive paradigm like game theory. Game

theory has proved to be very useful when a researcher needed to examine a variety of incentives (monetary, social etc.) and their effect on decision-making in an interpersonal setting [31]. In a game paradigm, players are typically offered choice between a self-serving and other-regarding incentive. A decision of one player may either open an opportunity for the other player or preclude that player from choosing [32]. Each player is made aware of the counterpart's choices but some uncertainty about the opponent's preferences remains. Thus, an opportunity for cooperation is provided, but cooperation is not necessarily enforced. This situation is certain to arouse a social dilemma, whereby one's regard for welfare of the other is on a collision course with a concern for personal gain.

In this dissertation, the assumptions about trust are examined using a popular testable, quantifiable and replicable game paradigm, called "investment game" (IG), a. k. a. "trust game" [33] – a model of human-to-human interaction between two players, "trustor" and "trustee" (see Chapter 2, Methods section for details). Both players are confronted with a difficult choice of either sharing money with their opponents for the sake of a longer-term mutual benefit or defecting on the opponent for the sake of personal gain. The amount of money sent by the trustor is believed to measure trust; the amount of money paid back by the trustee measures trustworthiness. One of the primary advantages of a behavioral paradigm like IG is that it is a simple measure, in which the confounding effects of decision variables on one another are reduced to a minimum [34].

Among the benefits of IG, are also its low (binary) dimensionality of the reward structure and anonymity of interaction. These properties allow for high signal-to-noise ratio (in contrast to a real-life lending situation) and control for the effects of personal char-

acteristics (e.g., age, appearance, emotional expression) and social context (e.g., socio-economic status of the players). In combination with a selection of participants whose intelligence and competence are sufficiently suitable for the task, an IG experiment can be designed to rule out a large number of confounding effects and allow for an improved measure of more subtle aspects of trust behavior. Among the drawbacks of the IG model however are simplification and reductionism in its approach to trust. As a result, for an experiment like IG to acquire external validity, the researchers must go beyond the study of behavior and gain a deeper insight into the relevant hidden mental states. This goal can be accomplished to a certain extent by the use of neuroimaging methods.

Subsection 1.3.2 Neuroimaging

As noted in previous sections, trust could be viewed as a belief (unobservable) or decision (explicit) condition. The unobservable aspects mount a difficult challenge to any attempt of measuring the condition effectively. As a result to date, a significant number of behavioral studies of trust lack crucial comprehension of the trust's hidden properties. Uses of neuroimaging in general and of functional magnetic resonance imaging (fMRI) in particular, were originally viewed as a powerful initiative to meet this challenge. Subsequently, in recent decades, there has been a proliferation of neuroimaging studies of neural origins of trust (for review, see [35,36,37]). However, the neuroimaging of IG produced no unifying schema of brain localizations of the discrete components of trust identified in stand-alone studies. Furthermore, the disparate results exposed a number of limitations of the fMRI method itself. While on the merit an fMRI study can provide localization of activation in stereotaxic coordinates (measured as blood oxygenation level dependent

[BOLD] signal), several limitations of the fMRI methodology stand in the way of generalizing the results. These limitations include for example, the inherent subtraction logic sensitive to variability across experimental conditions [38,39], lack of statistical power inherent in small sample sizes [40], low reliability [41,42], low-level of analytical abstraction [43], inability to produce direct evidence of the hypothesized mental states (latent neuronal activity) and unassertive external validity i.e., inability of a small number of subjects to fully characterize the spatial localization of activations across the entire population [44]. These and other less significant limitations, make apparent the need for a higher-level analysis and comparison of data trends across discrete experimental conditions. Qualitative meta-analyses of trust have provided high-level summaries of the biological basis of trust by integrating data from several studies. However, a detailed quantification of consistent activation data collectively pooled from hundreds of participants and various experimental designs across studies is an important next step in uncovering the underlying processes of interpersonal trust.

Section 1.4 Proposed Methods

Subsection 1.4.1 Scope

In the upcoming chapters a practical method for overcoming the methodological deficits of fMRI is discussed in depth. This remedy is two-fold. In Chapter 2, the functional image of a domain-general neurocognitive network of trust is laid out. In Chapter 3, a more complete schema of the network is established by augmenting the findings of Chapter 2 with the effective connectivity analysis (ECA). Chapter 2 discusses "Activa-

tion Likelihood Estimation" (ALE) – a practical coordinate-based meta-analysis method for scrutinizing discrete neuroimaging findings. ALE methodology was designed to overcome a prevalent negative tendency in fMRI studies, which is to overlook the unity of global phenomenon of a behavior and focus instead on its discrete components. The slender approach of a stand-alone fMRI study results in lack of conclusive evidence on how the "disconnected" neurocognitive modules act in harmony to produce a comprehensive behavior. Previous fMRI-IG findings thus provide no network level information to predict behavior as a whole. To address these issues, a coordinate-based meta-analysis approach to the IG fMRI data is proposed – a new research mission not undertaken in any of the previous fMRI studies of IG.

Identifying functional correlates of trust is an important first step in the analysis of the IG fMRI data. However, as a high-level cognitive function trust is more likely to be sub-served by a network of associated rather than segregated functional regions [45]. For an improved knowledge of the cognitive function, a researcher would have to establish causal links among neuronal populations comprising the underlying networks. One approach to characterization of these causal influences can be expressed in terms of effective connectivity, or directed influence that one neuronal population can exert over another. Chapter 3 discusses "Multivariate Granger Causality" (MVGC) – a practical data-driven method for the analysis of effective connectivity. Among the advantages of MVGC is that connectivity can be established for pairs of neuronal nodes regardless of whether they are located within an individual's brain or in different individuals' brains. Chapter 3 takes advantage of this powerful feature of MVGC to examine trust as a rela-

tional, interpersonal construct. An improved understanding of trust-based interaction is predicated on the knowledge of causal neurocognitive influences between the brains of interacting persons.

However to date, no previous study of trust in IG has addressed the outstanding issues of either the within-brain or between-brain connectivity. In Chapter 3, the MVGC method is employed to examine the research questions arising for these two now undertaken missions. Yet, for the between-brain analysis to proceed, the study had to find ways to monitor and measure the activity in the brains on both sides of the relationship concurrently. This goal was achieved through the use of hyperscan-fMRI – a method crucial to the between-brain analysis in that it provides the opportunity to simultaneously image two interacting brains [46]. The outcomes are based on an existing dataset collected in one of the earlier hyperscan-fMRI studies of IG [47]. The combination of hyperscan-fMRI and MVGC methods allowed for a discovery of the effects of neuronal activity in both within a brain and in one of the brains on another. The combination of meta-analysis (Chapter 2) and ECA (Chapter 3) provided a comprehensive network schema consisting of a neural structure of the network and patterns of causal influence of the network's structural elements on each other.

Subsection 1.4.2 Meta-Analysis (ALE)

The focus of Chapter 2 is practical application of a consolidating technique, specifically ALE, developed for the purpose of integrating isolated fMRI findings [48]. ALE method readily brings many advantages unachievable in a standard fMRI study. They include the capacity for quantifying consistency in neural responses across numerous ex-

periments and the ability to localize brain coordinates of interest to a given task [38,40,49]. ALE integrates data from a large number of standalone studies and effectively resolves the persistent fMRI issue of small sample size, which tends to undermine statistical power of the findings [41,42]. An adaptation of ALE strategies to trust research provides the researchers a tool for building global neural networks based on the cross-study consistency of activations. However, despite its merit of being a coordinate-based technique for localizing a position in space, ALE lacks the capacity for temporal dimension, which makes relying on this method for understanding the network dynamics rather limited.

Subsection 1.4.3 Effective Connectivity Analysis

Chapter 3 discusses a series of practical approaches to overcoming the limitations of both the ALE and fMRI methods. Although constructing a dynamic schema of a neurocognitive network is predicated on bringing together the global structure of the network, the knowledge of directed interactions among the network segments is required. ECA comes handy as a measure of causal directionality ("causality") between the signals in two or more neuronal populations. It is derived from a statistical prediction of change in the time course of a signal, based on its relation history with a signal in a different region. However, the fMRI measure of these latent neuronal signals is "noisy" as it results from convolution between the brain's neuronal and hemodynamic responses (HR). This obstacle necessitates an extra effort of "denoising" (deconvolving) the latent neuronal signal before it can be fit it into an ECA model [50].

The ECA method applied in Chapter 3 is a practical implementation of the more general MVGC mapping tailored to fMRI needs [51,52]. MVGC is employed to determine

the strength and direction of causal influences among the neurocognitive systems of trust "within" and "between" brains. This approach takes the investigation of trust networks a step further than the existing fMRI-IG literature. Although, the earlier studies of trust have dramatically improved the knowledge of the discrete components of trust, none of the studies show how these functional regions exchange information to influence the behavior. This gap in knowledge is due in part to high costs of the supporting fMRI and hyperscan technologies [46] and is the reason there have been only two hyperscan studies on the subject [47,53]. And only one of these rare studies provides a connectivity perspective. Consequently for trust, both the actual connectivity of neuronal populations within-brain and the "virtual" connectivity, or the relation between activities in different brains, remain to be examined. The proposed two-study solution provides distinct contribution to trust research by attempting to close the gap in knowledge.

CHAPTER 2. Domain-General Neural Network of Trust

Section 2.1 Introduction

Subsection 2.1.1 Scope

Chapter 1 has provided a lead-in from the theoretical assumptions about trust into the practical solutions of measuring it. One of the main concerns of trust to be measured in the upcoming Chapter 2 is whether a trustee is reliable in fulfilling certain goals. Implicit in trust is a disposition to delegate a valued resource (e.g., power, property, ability, or information) to someone (individual, group or institution), who is expected to either augment the value of the resource or generate a new positive value (reward) [54-56]. According to some qualitative models, trust is essentially an active goal-driven strategy for maximizing or optimizing rewards [2,35]. Rewards originate in a cognitive offset between a certain goal and value of the desired target stimulus, which creates the dichotomy of delegation. While the goal may not necessarily be impossible to achieve, the trustor is always better off if it's delegated to someone who is more competent and more willing to accomplish the task. Despite the accompanying risk of betrayal, delegation persists if the outcome positivity merits the risk [57].

The aim in Chapter 2 is to provide a practical concept of a domain-general neurocognitive network of trust – a mission not yet undertaken in previous studies. The main objective is to consolidate, based on the most consistent results, the discrete findings from a series of fMRI-IG studies. The priority is to gain insight into the patterns of con-

sistent activation for trust and its cognitive functions i.e., reward processing, aversion, inference making and choice. ALE has been chosen for this study on its merit of being a coordinate-based technique. In this capacity, ALE is a hands-on approach to overcoming known fMRI methodological limitations such as sensitivity to noise, small sampling size consequential to high costs, reduced spatial resolution, lag in the measured signal and what is more relevant to this study, the indirect nature of the fMRI signal as it relates to the latent neuronal signal. The meta-analysis presented in this chapter is thus expected to consolidate the findings and successfully manage the metric deficits at hand.

Subsection 2.1.2 Two Forms of Trust – Two forms of IG

Behavioral claims about trust can be quantifiably measured through the use of economic games [58]; among which IG (see Methods for details) has been particularly useful for studying trust. This sequential game is based on taking turns and each player is given one move at a time. The paradigm can be structured in two distinct formats, a one-time ("one-shot") exchange [33] and a repeated ("multi-round") interaction [59]. Cooperation can occur in either of the variants but under the condition that both players act to their mutual benefit and base their decisions on shared expectations of common gains and good will [60,61]. The two forms of IG are distinct in terms of the trust-warranting properties ("trustworthiness") implicit in their respective designs such as ambiguity about partner's intentions, the amount of knowledge afforded the players and the degree, to which the players can account for the partner's intentions ("depth of thought") [62]. The one-shot design is intended to lend no information on the trustworthiness, thus leading to unconditional trust. In contrast, in the repeated game the players are provided feedback

on the outcomes of their partner's choices. Nevertheless, risk of betrayal is implicit in both forms of IG, though the games vary greatly in their degree of ambiguity and subsequent aversion. In the one-shot IG for example, the insufficient evidence of positivity of intentions results in trusting "beyond the evidence" [2] and therefore, beyond aversion. The multi-round IG exchange on the other hand, is a well-informed regularity leading to trust with confidence and less aversion.

Subsection 2.1.3 Behavioral Evidence

In both games the players are put in a dilemma of intertemporal choice between a greater but postponed gain and an immediate but smaller reward. This is a "sticking point" for a standard "rational agency" economic solution based on backward induction, which had predicted that the reward delay precludes a "rational" (self-interested) agent from cooperating, i.e. from either trusting or repaying trust [58]. Intriguingly, these extrapolations however intuitive to a classical economist, do not stand up to empirical scrutiny [63]. According to a meta-analysis of hundreds of behavioral studies of IG, the majority of trustors demonstrate significant incentive to trust by sending more than half of their endowment amount to the trustees; who in turn, demonstrate significant incentive to reciprocate by responding with even split [64]. These results can only be explained by the notion that trust is driven and motivated by something more than a mere self-interest. Unconditional trust might be governed by a generalized belief in the partner's good will. Conditional trust on the other hand, is presumably based on reputation (trustworthiness) derived from the trustee response history [65] and information on the trustee personal characteristics obtained prior to making trust decisions [66,67]. The knowledge of the response history

affords the trsutor a firm basis of reputation for estimating trustworthiness, sustaining strategic maneuvering throughout the exchange [58,66] and maintaining an up-to-date mental model of the opponent's intentions that are affecting trustor's choice [62].

Subsection 2.1.4 Neuroimaging Evidence

The differences in cognitive properties between the two forms of trust, conditional and unconditional, are quite complex and demonstrate the richness of the socio-cognitive model of trust. Yet, the model does not provide the necessary insight into the hidden mental properties of the construct. Neuroeconomists have been working to address this gap by combining the IG paradigm with fMRI in an attempt to gain the evidence. Initial work for this mission was undertaken in the pioneering study of McCabe, et al. [60]. The results of this and a couple dozen subsequent fMRI studies of IG were, for the most part, consistent with the earlier evidence generated by neurobiology and neuroeconomics of decision-making. Social reward dependency has been linked with mPFC and the associated subcortical circuitry involving basal ganglia and midbrain structures [6,68,69]. The parietal area LIP was linked with choice [70]. Mentalizing was linked with the default-mode network, specifically with mPFC and posterior cingulate cortex (PCC) [47,71]. Within the network, mPFC was implicated in attributing others' dispositional traits [72-74] and had a role in inhibiting self-serving impulses in favor of postponed mutual gains [60]. PCC was involved in mediating autobiographical memory [75] and self-referential processing [76,77]. Aversion to ambiguity was linked to the anterior insula, anterior cingulate cortex (ACC) and amygdala [27,78]. One of the core hypothesized functions of the anterior insula was to mark salient emotions of aversion [79,80], whereas the ACC was presumably engaged in de-

tecting conflicts and promoting behavioral action adjustments [81,82]. Despite the apparent success of combining fMRI and IG experiment paradigms, the disparity of the resulting findings highlighted the obvious drawbacks of the fMRI data collection and data analysis methods utilized in those studies. Chapter 2 is providing the overview of these restrictions and will introduce a practical solution for overcoming the deficits.

Subsection 2.1.5 Research Inquiry

The meta-analysis presented in this chapter is among the first to investigate consistent activation patterns in the selected fMRI-IG studies. The objective is to statistically dissociate the neuropsychological factors of unconditional and conditional trust and of repayment of trust (reciprocity). The statistical significance of the results will be based on the outcomes of a series of single-dataset and image-contrast ALE studies. In light of the objectives, the analysis will test significance of the following research inquiries:

Inquiry 1: Are the two distinct forms of trust, unconditional (one-shot IG) and conditional (multi-round IG), predicted by dissociable neural activity?

Inquiry 2: Are the two distinct phases of learning trust, decision and outcome, predicted by dissociable neural activity?

Inquiry 3: Are the two distinct forms of cooperative behavior, trust and reciprocity, predicted by dissociable neural activity?

Section 2.2 Methods

Subsection 2.2.1 Behavioral Paradigm

In a standard IG, two anonymous players receive an endowment and are assigned to either the role of "trustor" or "trustee". The trustor may choose to "trust" (by sharing a portion of the endowment with the trustee) or to "distrust" (by sharing nothing). If "trusted" (shared), the money is multiplied (usually tripled) by the experimenter and passed on to the trustee. The trustee may then choose to either reciprocate trust by passing back a portion of the money he receives to the trustor or to defect by passing back nothing. At the end of the exchange both players are given feedback on the outcomes of each other's transfers. Presumably, the amount of money passed by the trustor captures trust and the amount of money returned by the trustee captures reciprocity. Trustor's final payoff equals the initial endowment minus the transfer to the trustee, plus the back transfer from the trustee. Trustee's final payoff equals the initial endowment plus the tripled transfer of the investor, minus the back transfer to the investor.

Subsection 2.2.2 Study Selection

To identify studies pertinent to the analysis, a systematic database search on PubMed, ISI Web of Science and Google Scholar was performed during the period from January to June 2015. The combinations of relevant search keys included "trust", "trust game", "investment game", "trustor", "investor", "fMRI", "neuroimaging", "trustee", "trustworthiness" and "reciprocity". In addition, several other sources were examined, in-

cluding the BrainMap database (http://brainmap.org), the work cited in review papers and direct searches on the frequently occurring author names.

Subsection 2.2.3 Inclusion Criteria

To be considered for inclusion, a study had to satisfy the following criteria: (1) Participants had no history of neurological/psychiatric disorder or medication use[1]. (2) Participants played as either a trustor or trustee in a decision (trust, reciprocity) or outcome phase of either a one-shot or multi-round IG. (3) FMRI was the neuroimaging modality of choice. (4) The fMRI results were derived from a combination of a whole-brain and general linear model (GLM) data analyses based on either a binary contrast or continuous parametric analysis of the data[2]. (5) Brain activations were reported in a standardized stereotaxic space, MNI or Talairach [83].

Subsection 2.2.4 ALE Methodology

ALE determines cross-publication consistency of foci collected from a corpus of functional neuroimaging studies [85,86]. In ALE, foci are interpreted as spatial probability distributions and estimates of the distributions vary with spatial uncertainty caused by the between-subject and between-template variability of the neuroimaging data. The probabilities are assessed against a null-distribution of random spatial association among studies, allowing for random-effects inference. The histogram of ALE scores, obtained from several thousand random iterations in a permutation test (5,000 permutations) is then

[1] Such restrictions would often lead to an inclusion of only a subset of the results pertaining exclusively to a healthy control group.
[2] Results derived from a region of interest (ROI) or small-volume correction analysis would violate assumptions of the underlying algorithm for ALE.

used to assign P values to the observed foci. ALE maps (i.e., images) are created by computing the union of activation probabilities for each voxel.

Subsection 2.2.5 ALE Procedure

The coordinate-based meta-analysis was performed through the use of a revised ALE algorithm [84] built into the freeware activation likelihood estimation program GingerALE (version 2.3, http://www.brainmap.org/ale/). Coordinates published in Talairach space were converted to MNI space via a transform algorithm built into the GingerALE "icbm2tal" program (http://www.brainmap.org/icbm2tal/). Independent, random-effects ALE analyses were performed to determine cross-publication convergence of the activation coordinates (foci) grouped by the published experiment contrasts. The resulting ALE maps were thresholded to correct for multiple comparisons using false discovery rate, $q(\text{FDR}) < 0.05$ [48,87] with a minimum cluster volume of 100 mm^3. ALE single-dataset images were individually created for each of the compared conditions i.e., unconditional and conditional trust, decision and outcome phase and reciprocity. ALE contrast of the images was predicated on the pooled dataset combining the individual datasets. Each of the pairwise comparisons produced two single-dataset images using ALE statistic, one image for the conjunction and two images for the contrast each using z-scores. The results were overlaid onto a normalized brain template "ch2better" – a component of MRIcron visualization software [88] and displayed using Mango brain image navigation software (Jack L. Lancaster, Ph.D. and Michael J. Martinez).

Section 2.3 Results

Subsection 2.3.1 Final Corpus

The initial search-key screening procedures uncovered 55 candidate papers. 30 of them met the inclusion criteria and were included (**Appendix A**). Three patient studies met all but one, the "healthy participant" requirement and were excluded [53,89,90]. However, a set of healthy participant data from a paper reporting both patient and healthy participant outcomes was included in the analysis [91]. The remaining 21 publications were excluded on the grounds of not fitting the rest of the inclusion criteria. Three of those were published as reviews and reported no original results [92-94]. In 3 others, IG was used as either a pre-scanning [96,97] or post-scanning [95] behavioral measure. Yet in 3 more the IG design did not match the requirements of the inclusion criteria [98-100]. In 1 publication, fMRI of IG was performed prior to the decision phase, which again violated the inclusion criteria [101]. In 2 of the papers no whole brain analysis results were reported and in 3 others, the IG paradigm was used for a non-fMRI imaging study. Among the final corpus of 30 articles meeting the inclusion criteria the results for the trustor and trustee were reported in 23 and 11 papers respectively (**Table 1**). In 3 publications, both sides of the task were reported. For the trustor side, 5 articles were one-shot and 18 were multi-round. For the trustee, 4 publications were one-shot and 7 were multi-round. Of the 18 multi-round articles on the trustor, 16 reported decision phase and 8 reported the outcome phase. Six papers reported both phases. Of the 7 multi-round articles on the trustee, 6 reported decision phase, 2 reported outcome phase and 1 article reported both phases.

Subsection 2.3.2 Foci Selection

The final corpus provided a total of 100 contrasts (594 foci observed in 985 participants), 67 of which were reported for the trustor and 17 for the trustee. Specifically for the trustor, there were 44 contrasts (227 foci for 601 subjects) representing the decision phase, 13 (52 foci for 122 subjects) of which were one-shot and 31 (175 foci and 479 subjects) were multi-round (**Table 2**). For the trustee, there were 17 contrasts (203 foci for 186 subjects) representing the decision phase (**Table 4**). And there were 23 contrasts (119 foci for 290 subjects) representing the outcome phase of the trustor (**Table 3**).

Subsection 2.3.3 ALE Results

Inquiry 1 concerned the neural dissociation between unconditional and conditional trust. The results are reported for the ALE pooled analysis of the forms (**Table 5**), single-dataset analysis of each of the forms and their contrast. Consistent peaks were localized 1) for the one-shot IG (**Figure 1, Table 6a**) – to the right ant. insula, frontal operculum, dorsal ACC (dACC) and hippocampus; 2) for the multi-round IG (**Figure 1, Table 6b**) – to the right ventromedial PFC (vmPFC), premotor cortex (PMC), supplemental motor area (SMA), precuneus, inferior temporal cortex (IT), caudate body (BCd), pulvinar, visual cortex (V2), cerebellum; left rostrolateral PFC (rlPFC), dlPFC, PCC, inf. parietal lobule (IPL), ventral striatum, ventral anterior nucleus of thalamus (VA), mid temporal cortex (MT) and bilateral putamen; 3) for the (one-shot > multi-round) comparison (**Figure 2, Table 7a**) – to the right ant. insula; 4) for the (multi-round > one-shot) comparison (**Figure 2, Table 7b**) – to the right dorsal (caudate) and left ventral (NAc) striata.

Table 1 Summary of selected publications with focus on trust, reciprocity and outcome phases of IG

Phase:	Trust (Trustor)						Reciprocity (Trustee)						Outcome (Trustor)		
Game:	One-Shot			Multi-Round			One-Shot			Multi-Round			Multi-Round		
Study	C	F	N	C	F	N	C	F	N	C	F	N	C	F	N
McCabe, et al. [60]	1	1	6				1	1	6						
Kang, et al. [79]	3	19	8												
Lauharatanahirun, et al. [78]	1	9	30										2	9	30
Stanley, et al. [67]	4	15	40												
Aimone, et al. [27]	3	7	30												
	1	1	8												
King-Casas, et al. [59]				1	1	48									
Krueger, et al. [47]				1	4	44									
				1	2	22									
Krueger, et al. [102]				2	5	44				1	2	44			
Baumgartner, et al. [103]				3	13	24									
Sripada, et al. [91]				1	9	26									
Bereczkei, et al. [104]				2	27	12				2	15	12			
Fett, et al. [105]				6	25	45									
Fouragnan [106]				1	8	18									
Wardle, et al. [107]				1	8	29									
Riedl, et al. [69]				1	8	18									
Delgado, et al. [108]				2	24	12							2	12	12
Fareri, et al. [65]				2	17	18							5	50	18
Xiang, et al. [62]				3	5	44							2	5	44
													2	5	49
Fouragnan, et al. [66]				1	4	18							2	4	18
Smith-Collins, et al. [109]				2	14	24							3	12	24
Gromann, et al. [68]				1	3	33							1	3	33
Phan, et al. [6]													1	6	36
Fareri, et al. [71]													3	13	26
van den Bos, et al. [110]							4	10	18						
van den Bos, et al. [111]							2	4	54						
							1	1	15						
Nihonsugi, et al. [112]							2	8	41						
Baumgartner, et al. [23]										1	3	26			
Li, et al. [70]										1	5	52			
										2	22	20			
Chang, et al. [80]										6	123	16			
Bereczkei, et al. [113]										4	33	16			

C, number of reported contrasts; F, number of reported activation foci; N, participant count

Table 2 Summary of selected experiment contrasts for trust (one-shot and multi-round IG)

Study	N	Contrast	F
		One-Shot IG	
McCabe, et al. [60]	6	$[T\ (Hum) > T\ (C)]_{COOP} > [T\ (Hum) > T\ (C)]_{UNCOOP}$	1
Kang, et al. [79]	8	$T_{WARM} > C_{NEUT}$	4
	8	$T_{COLD} > C_{NEUT}$	6
	8	$T_{COLD} > T_{WARM}$	9
Lauharatanahirun, et al. [78]	30	DM > C	9
Stanley, et al. [67]	40	T (Black) > T (White)	8
	40	AI_T (Black) > AI_T (White)	2
	40	DM (Black) > DM (White)	2
	40	AI_{DM} (Black) > AI_{DM} (White)	3
Aimone, et al. [27]	30	DM (Hum) > DM (C)	3
	30	T > D	2
	30	T (Hum) > T (C)	2
	8	T_{BA} (Hum) > T_{NBA} (Hum)	1
		Multi-Round IG	
Delgado, et al. [108]	12	T > D	8
	12	T (B) > D (R)	16
King-Casas, et al. [59]	48	T (R) > D (B)	1
Krueger, et al. [47]	44	[T (S1) + T (S2)] > [C(S1) + C(S2)]	2
	22	$[T(S2) > R(S2)]_T > [T(S2) > R(S2)]_D$	2
Krueger, et al. [102]	44	[T (S1) + T (S2)] > [C(S1) + C(S2)]	2
	44	$[T(S2) > R(S2)]_T > [T(S2) > R(S2)]_D$	3
Baumgartner, et al. [103]	24	$[(T > C)_{OXT} > (T > C)_{PLA}]$ (Before Feedback)	1
	24	$[(T > C)_{PLA} > (T > C)_{OXT}]$ (Before Feedback)	2
	24	$[(T > C)_{PLA} > (T > C)_{OXT}]$ (After Feedback)	10
Sripada, et al. [91]	26	T_{HC} (Hum) > T_{HC} (C)	9
Fareri, et al. [65]	18	T > D	3
	18	DM (Coop) > DM (Uncoop) > DM (Neut) > DM (C)	14
Xiang, et al. [62]	44	$(DM_{PE\ HIGH} > DM_{PE\ LOW})_{IL\text{-}2} > (DM_{PE\ HIGH} > DM_{PE\ LOW})_{IL\text{-}0}$	2
	44	$(DM_{PE\ HIGH} > DM_{PE\ LOW})_{IL\text{-}2} > (DM_{PE\ HIGH} > DM_{PE\ LOW})_{IL\text{-}1}$	2
	44	$DM_{IL\text{-}2} > DM_{IL\text{-}0}$	1
Bereczkei, et al. [104]	12	$(IG_{Hi\text{-}MACH} > IG_{Lo\text{-}MACH}) > (C_{Hi\text{-}MACH} > C_{Lo\text{-}MACH})$	10
	12	$(DM_{Hi\text{-}MACH} > DM_{Lo\text{-}MACH}) > (C_{Hi\text{-}MACH} > C_{Lo\text{-}MACH})$	17
Fett, et al. [105]	45	DM > C, [Signal ↑, Age ↑] (R)	11
	45	DM > C, [Signal ↓, Age ↑] (R)	7
	45	DM > C, [Signal ↑, Age ↑] (B)	4
	45	DM > C, [Signal ↓, Age ↑] (B)	1
	45	DM > C, [Activation Level ↑, Age ↑] (R)	2
	45	DM > C, [Activation Level ↑, Age ↑] (B)	1
Fouragnan, et al. [66]	18	T > D	4
	18	T > D	8
Smith-Collins, et al. [109]	24	T (R) > D (R)	3
	24	[T (R) + D (B)] > [T (B) + D (R)]	11
Wardle, et al. [107]	29	DM (Coop) > DM (Uncoop) > DM (Neut) > DM (C)	8
Gromann, et al. [68]	33	$T_{HC} > T_{SIBLINGS}$	3
Riedl, et al. [69]	18	T > C	8

N/F, participant/foci count. T/D/R/B, trust/distrust/reciprocity/betrayal. DM/IG, decision-making (T+D)/(trustor + trustee). Trustee: Hum, human; Black/White. Trustor: BA/NBA, betrayal-averse/non-BA. C/HC, control/healthy C. OXT/PLA, oxytocin/placebo; WARM/COLD/NEUT, pre-game T° treatment. Hi/Lo MACH, high/low on MACH IV test; COOP/UNCOOP/NEUT, reciprocity rate, >74% / <26% / 50%. SIBLINGS, healthy ~ of patients. AI, amt. invested. IL, Investor Level: 0, trustee move is not predicted; 1/2, trustee move is predicted as Level-0/1. S1/2, partnership bldg. (Stage 1)/maintenance (Stage 2); Notation: Subscripts, experimental groups; Parentheses, conditions.

Table 3 Summary of selected experiment contrasts for the outcome phase (multi-round IG)

Study	N	Contrast	F
Delgado, et al. [108]	12	O (R) > O (B)	11
	12	O (B) > O (R)	1
Phan, et al. [6]	36	O (R) > O (B)	6
Fareri, et al. [65]	18	O (R) > O (B)	19
	18	O (Coop) > O (Uncoop) > O (Neutral) > O (C)	3
	18	O_{PE} (Coop) > O_{PE} (Uncoop) > O_{PE} (Neutral) > O_{PE} (C)	6
	18	O_{PE} (Coop) > O_{PE} (Uncoop) > O_{PE} (Neutral) > O_{PE} (C)	15
	18	O_{PE}	7
Xiang, et al. [62]	49	[O (PE Hi) > O (PE Lo)]$_{IL-0}$ > [O (PE Hi) > O (PE Lo)]$_{IL-1}$	4
	44	[O (PE Hi) > O (PE Lo)]$_{IL-0}$ > [O (PE Hi) > O (PE Lo)]$_{IL-2}$	4
	49	O (PE)$_{IL-0}$ > O (PE)$_{IL-1}$	1
	44	O (PE)$_{IL-2}$ > O (PE)$_{IL-0}$	1
Fouragnan, et al. [66]	18	O (PE, No-Prior) > O (PE, Prior)	2
	18	O (Consistent) > O (Inconsistent)	2
Smith-Collins, et al. [109]	24	O (Expected) > O (Unexpected)	1
	24	O (Unexpected R) > O (Unexpected B)	6
	24	O (Unexpected R) > O (Unexpected B), Increase	5
Gromann, et al. [68]	33	O_{HC} > $O_{SIBLINGs}$	3
Fareri, et al. [71]	26	O (R)$_{FRIEND}$ > O (R)$_{OTHER}$	4
	26	O (B)$_{FRIEND}$ > O (B)$_{OTHER}$	3
	26	O (PE)	6

N/F, participant/foci count. R/B, reciprocity/betrayal. C/HC, control/healthy C. Prior, info on trustee prior to decision. Trustee: Coop/Uncoop/Neut, reciprocity rate: >74% / 26% / 50%. SIBLINGS, healthy ~ of patient. FRIEND, ~ of participant. Expected/Un~/Consistent/In~, outcome. PE, prediction error. IL, Investor Level: 0, trustee move is not predicted; 1/2, trustee move is predicted as Level-0/-1. (Subscripts denote experimental groups; parentheses - conditions.)

Table 4 Summary of selected experiment contrasts for reciprocity (multi-round IG)

Study	N	Contrast	F
Krueger, et al. [102]	44	R > C	2
Li, et al. [70]	52	R Ratio, Correlation	5
	20	$R_{NO\text{-}SANCTION} > R_{SANCTION}$	12
	20	$R_{SANCTION} > R_{NO\text{-}SANCTION}$	10
Chang, et al. [80]	16	$R_{LESS} > R_{MATCH}$	11
	16	$R_{MATCH} > R_{LESS}$	25
	16	$R_{LESS} > R_{MATCH}$ (Parametric)	22
	16	$R_{MATCH} > R_{LESS}$ (Parametric)	33
	16	R_{LESS}, Correlation	6
	16	R_{MATCH}, Correlation	26
Bereczkei, et al. [104]	12	$R\ (IG)_{\text{Hi MACH} > \text{Lo MACH}} > R\ (CG)_{\text{Hi MACH} > \text{Lo MACH}}$	5
	12	$DM2_{\text{Hi MACH} > \text{Lo MACH}} > C_{\text{Hi MACH} > \text{Lo MACH}}$	10
Bereczkei, et al. [113]	16	$R_{FAIR} > R_{UNFAIR}$	9
	16	$R_{FAIR} > R_{CTRL}$	15
	16	$R\ (Unfair)_{\text{Lo MACH} > \text{Hi MACH}} > R\ (C)_{\text{Lo MACH} > \text{Hi MACH}}$	2
	16	$R\ (Fair)_{\text{Lo MACH} > \text{Hi MACH}} > R\ (C)_{\text{Lo MACH} > \text{Hi MACH}}$	7
Baumgartner, et al. [23]	26	$R_{DISHONEST} > R_{HONEST}$	3

N/F, participant/foci count. R, reciprocity. IG/CG, investment/control game (trustor + trustee). SANCTION, defection is punished. Hi/Lo-MACH, score on MACH IV test. Trustee: HONEST/DISHONEST, keeping/breaking promises; LESS/MATCH, returning smaller/expected amt. of money; C, control; FAIR/UNFAIR/C, reciprocity rate: >74% / <26% / 50%. (Subscripts denote experimental groups; parentheses - conditions.)

Table 5 Results of pooled ALE analysis of trust (one-shot + multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
R	Dorsal Anterior Cingulate Cortex (cingulate gyrus)	24/32	1	15	41	1.78	1576
R	Anterior Insula/Frontal Operculum	13/44	41	18	2	1.91	1008
R	Ventromedial PFC (medial frontal gyrus)	32/10	8	48	–13	1.89	416
R	Dorsolateral PFC (middle frontal gyrus)	9	40	25	28	1.43	464
R	Dorsolateral PFC (middle frontal gyrus)	9	50	27	30	1.42	216
L	Dorsolateral PFC (middle frontal gyrus)	9	–35	38	27	1.27	448
L	Rostrolateral PFC (superior frontal gyrus)	10	–27	56	6	1.31	176
L	Posterior Cingulate Cortex (cingulate gyrus)	31	–29	–71	25	1.83	408
R	Precuneus (medial parietal wall)	7	14	–38	50	1.28	168
L	Inferior Parietal Lobule	40	–43	–34	44	1.72	312
R	Middle Temporal Cortex	36	36	-4	–37	1.54	264
L	Middle Temporal Cortex	37	–62	–55	–2	1.31	184
R	Putamen (dorsal striatum)		24	10	–3	2.45	1512
L	Putamen (dorsal striatum)		–19	8	–5	2.08	656
L	Putamen (dorsal striatum)		–32	–19	–2	1.56	424
R	Body of Caudate Nucleus (dorsal striatum)		13	5	8	1.11	120
L	Head of Caudate Nucleus (ventral striatum)	25	–1	4	–8	1.35	248
L	Ventral Anterior Nucleus of Thalamus		–11	–3	1	1.30	336
R	Pulvinar (thalamus)		14	-30	8	1.23	144
R	Premotor Cortex (superior frontal gyrus)	6	28	13	55	1.29	176
R	Supplementary Motor Area (medial frontal gyrus)	6	3	10	57	1.46	272
R	Extrastriate Visual Area (lingual gyrus)	18	14	–70	–3	2.29	696
L	Extrastriate Visual Area (lingual gyrus)	18	0	–83	2	1.28	200
L	Extrastriate Visual Area (middle occipital gyrus)	18	–22	–90	22	1.22	160

Lat, laterality: L, left; R, right.

Table 6 Results of ALE single-dataset analyses of trust (one-shot and multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
	a) Unconditional Trust, One-Shot IG (ALE)						
R	Anterior Insula/Frontal Operculum	13/44	42	18	2	1.88	1520
R	Dorsal Anterior Cingulate Cortex (cingulate gyrus)	32	1	15	41	1.09	736
L	Hippocampus (parahippocampal gyrus)	27	–21	–30	–4	1.02	176
	b) Conditional Trust, Multi-Round IG (ALE)						
R	Ventromedial PFC (medial frontal gyrus)	32/10	8	48	–13	1.89	456
L	Dorsolateral PFC (middle frontal gyrus)	9	43	26	27	1.34	840
L	Rostrolateral PFC (superior frontal gyrus)	10	–27	56	6	1.31	256
L	Posterior Cingulate cortex (cingulate gyrus)	31	–29	–71	25	1.83	472
R	Precuneus (medial parietal wall)	7	14	–38	50	1.27	224
L	Middle Temporal Cortex	37	–63	–55	–3	1.31	264
R	Inferior Temporal Cortex	37	47	72	7	1.13	136
L	Inferior Parietal Lobule	40	–43	–34	45	1.72	384
R	Putamen (dorsal striatum)		24	10	–3	2.45	1680
L	Putamen (dorsal striatum)		–19	8	–5	2.07	736
L	Putamen (dorsal striatum)		–32	–20	–2	1.55	488
R	Body of Caudate Nucleus (dorsal striatum)		13	5	8	1.11	176
R	Body/Head of Caudate Nucleus (dorsal striatum)		11	13	3	1.05	144
L	Head of Caudate Nucleus (ventral striatum)	25	–1	4	–9	1.35	320
L	Ventral Anterior Nucleus of Thalamus		–11	–3	1	1.30	448
R	Pulvinar (thalamus)		14	–30	8	1.23	232
R	Premotor Cortex (superior frontal gyrus)	6	28	13	55	1.28	288
R	Supplementary Motor Area (medial frontal gyrus)	6	3	10	58	1.46	312
R	Culmen (cerebellum)		20	–41	–24	1.10	112
R	Extrastriate Visual Area (lingual gyrus)	18	14	–70	–3	2.29	776

Lat, laterality: L, left; R, right.

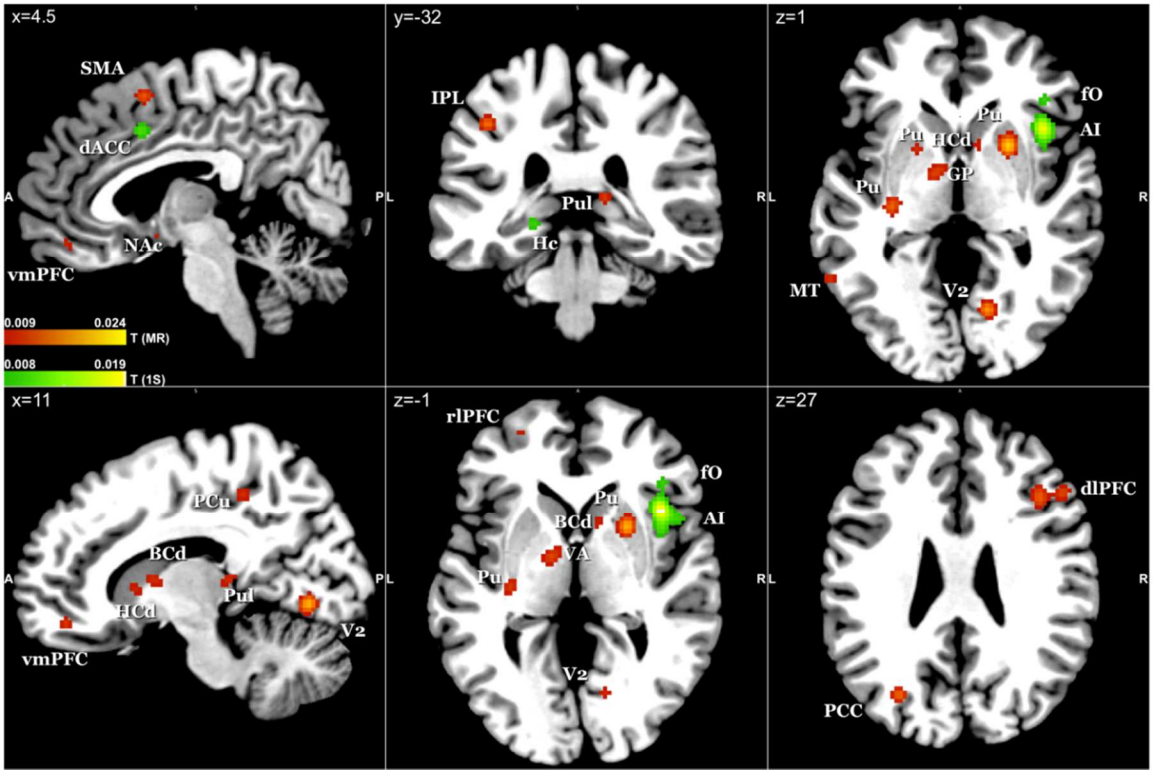


Figure 1. Results of ALE single-dataset analysis of trust (one-shot and multi-round IG).
Multi-Round Networks (Red): mOFC (vmPFC, NAc, HCd, GP, VA), ventral vis. stream (V2, MT/IT, Pul) & frontoparietal (rlPFC, dlPFC, PCC, PCu, IPL, PMC [not shown], SMA, BCd). One-Shot (Green): cingulo-opercular network (AI, dACC, fO) & Hip. Random-effects analysis, 5,000 permutations, ALE values, q(FDR)<0.05, min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.

Table 7 Results of ALE image-contrast analysis of trust (one-shot vs. multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			Z	Size (mm^3)
			x	y	z		
	a) One-Shot > Multi-Round (z-score)						
R	Anterior Insula	13	43	16	2	3.54	1184
	b) Multi-Round > One-Shot (z-score)						
R	Body/Head of Caudate Nucleus (dorsal striatum)		11	13	3	3.09	144
R	Body of Caudate Nucleus (dorsal striatum)		13	5	8	2.91	144
L	Head of Caudate Nucleus (ventral striatum)		-1	5	-8	3.24	256

Lat, laterality: L, left; R, right.

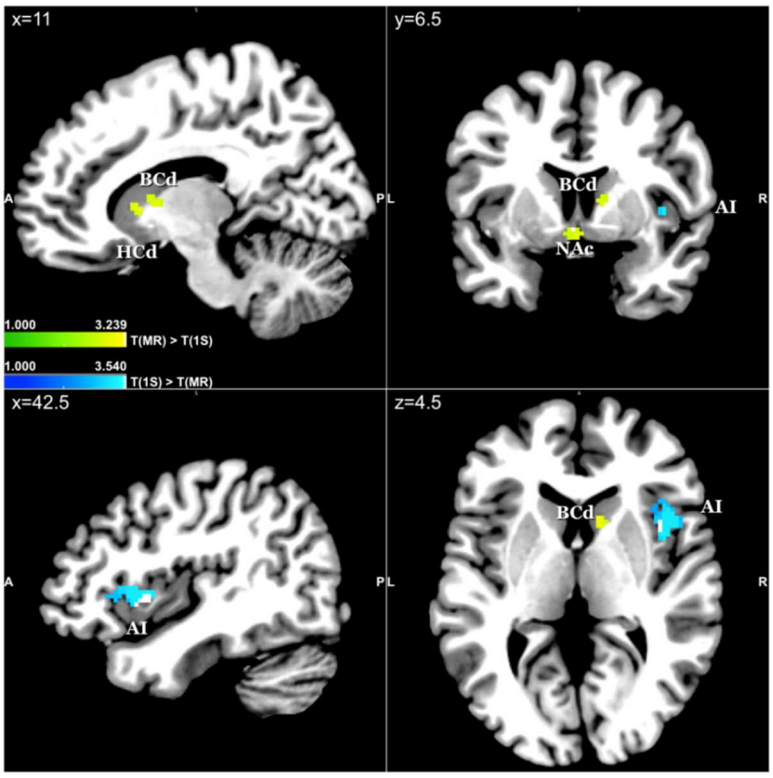


Figure 2. Results of ALE image-contrast analysis of trust (one-shot vs. multi-round IG).
Multi-Round > One-Round (Green): peaks in BCd, HCd & NAc. One-Round > Multi-Round (Blue): peak in AI. Random-effects analysis; 5,000 permutations; z-scores; q(FDR) < 0.05; min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.

Inquiry 2 concerned the issue of dissociable neural networks of trust during decisions and outcomes. For the outcome phase, ALE single-dataset analysis uncovered consistent peaks in the right lateral orbitofrontal cortex (lOFC), MT, V2 and caudate head (HCd); in the left IT, globus pallidus (GP), BCd and bilateral putamen (**Figure 3, Table 8**). ALE contrast analysis of the two phases (**Table 9**) revealed the conjunction (outcome ∩ decision) in the right BCd and bilateral putamen (**Figure 4, Table 9a**) and the (outcome > decision) contrast in the right HCd and left hippocampus (**Figure 4, Table 9b**).

Table 8 Results of ALE single-dataset analysis of trust outcome phase (multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
	Outcome, Multi-Round IG (ALE)						
R	Lateral Orbitofrontal Cortex (mid frontal gyrus)	10/11	32	50	–14	1.18	280
R	Middle Temporal Cortex	20	57	–37	–19	1.43	376
L	Inferior Temporal Cortex	20	–55	–37	–15	1.03	176
R	Head of Caudate Nucleus /Putamen (striatum)		11	16	–1	2.65	4784
L	Body of Caudate Nucleus (dorsal striatum)	33	–12	15	5	2.08	1120
R	Putamen (dorsal striatum)		24	4	0	1.97	616
L	Putamen (dorsal striatum)		–24	4	4	1.96	648
L	Globus Pallidus (ventral pallidum)		–18	4	–12	2.73	2064
L	Claustrum/Insula	13	–36	11	1	1.24	312
R	Extrastriate Visual Area (cuneus)	18	4	–88	17	1.55	736

Lat, laterality; L, left; R, right.

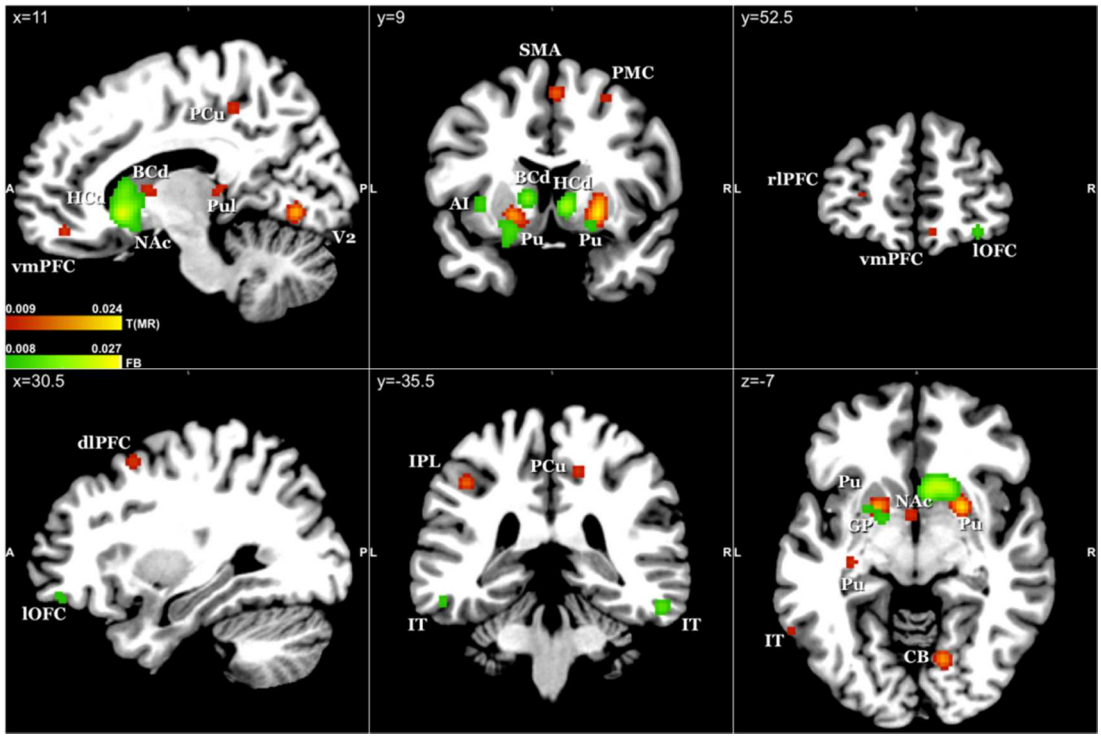


Figure 3. Results of ALE single-dataset analysis of trust decision and outcome phases (multi-round IG). Outcome Phase (Green): peaks in lOFC, BCd, HCd, NAc, Putamen, GP, CB, AI, IT & MT. Decision phase (Red). Random-effects ALE analysis, 5,000 permutations, ALE values, q(FDR) < 0.05, min. threshold of 100 mm^3. Image overlay on a normalized brain template using Mango.

Table 9 Results of ALE image-contrast analysis of trust decision vs. outcome phases (multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
	a) Trust Decision ∩ Outcome (ALE)						
R	Putamen (dorsal striatum)		24	4	0	1.97	656
R	Putamen (dorsal striatum)		19	12	–6	1.15	184
L	Putamen (dorsal striatum)		–19	7	–8	1.30	216
R	Body of Caudate Nucleus (dorsal striatum)		11	13	3	1.05	144
	b) Outcome > Trust Decision (Z score)					**Z**	
R	Head of Caudate Nucleus (dorsal striatum)		8	18	–1	3.16	520
R	Putamen /Hippocampus	34	–17	4	–15	2.85	352

Lat, laterality; L, left; R, right

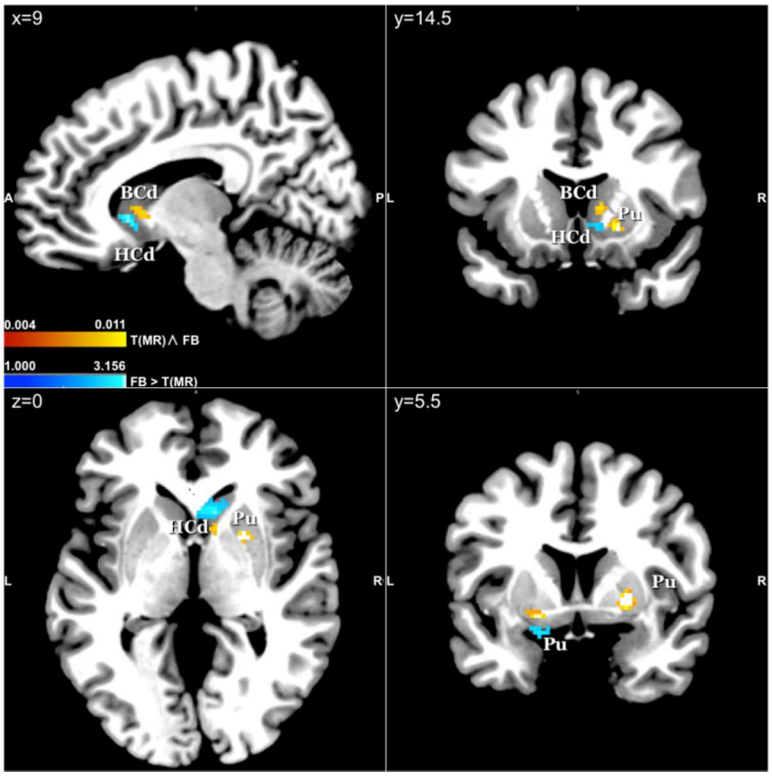


Figure 4. Results of ALE image-contrast analysis of trust decision vs. outcome phases (multi-round IG). Decision ∩ Outcome (Red): peaks in right BCd & HCd + bilateral putamen. Outcome > Decision (Blue): peaks in left HCd & Putamen. Random-effects analysis; 5,000 permutations; z-scores; q(FDR) < 0.05; min. threshold of 100 mm^3; Image is overlaid on a normalized brain template using Mango.

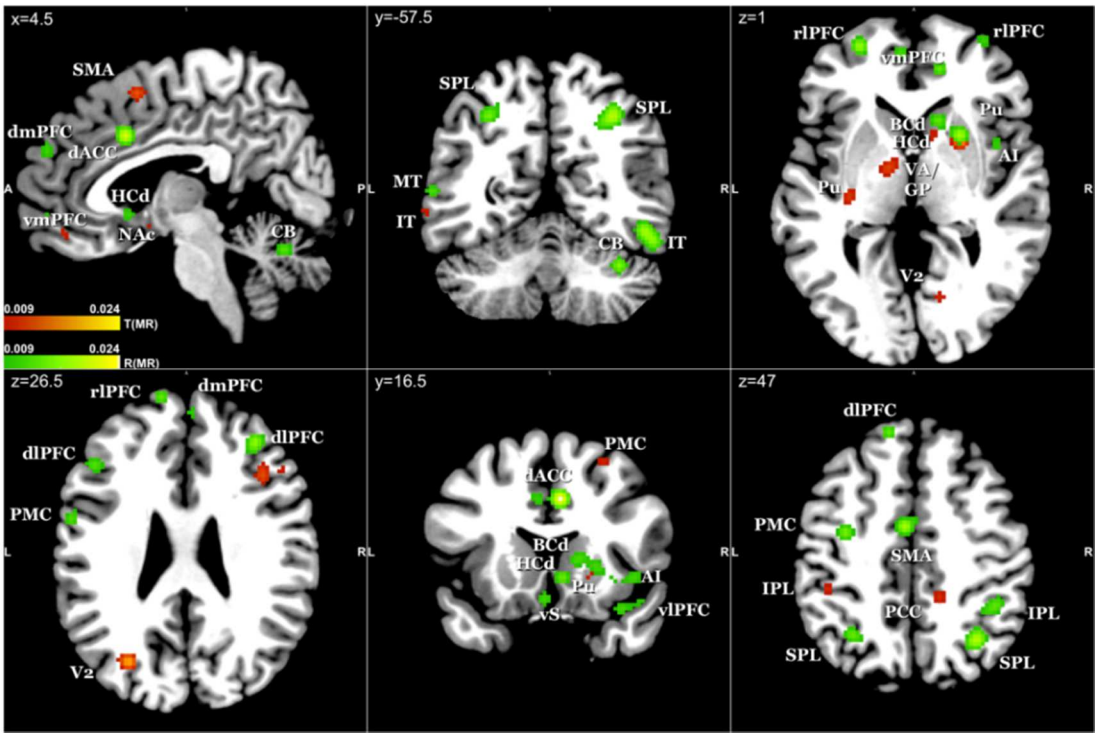


Figure 5. Results of ALE single-dataset analysis of trust vs. reciprocity (multi-round IG).
Reciprocity Networks (Green): frontoparietal (rlPFC, dlPFC, dmPFC, vlPFC, PCC, SPL/IPL, BCd & HCd) & mPFC (vmPFC, dACC, NAc, GP & VA) + AI, IT/MT, putamen & CB. Trust (Red). Random-effects analysis; 5,000 permutations; ALE values; q(FDR) < 0.05; min. threshold of 100 mm^3; Image is overlaid on a normalized brain template using Mango.

Inquiry 3 concerned the issue of common vs. dissociable neural networks of trust and reciprocity. Reciprocity (**Figure 5, Table 10**) engaged the right vlPFC, dmPFC, SMA, IT, putamen & HCd; the left SMA, MT & ventral striatum and bilaterally, dACC, PMC, ventromedial, rostrolateral/dorsolateral PFC, inferior/superior parietal lobules (IPL/SPL), V2 & cerebellum. The conjunction (trust ∩ reciprocity) involved the right putamen & BCd and the left rlPFC & NAc (**Figure 6, Table 11a**). The (reciprocity > trust) comparison identified consistent peaks in the right IPL/SPL, anterior insula, vlPFC & IT and in the left PMC (**Figure 6, Table 11b**).

Table 10 Results of ALE single-dataset analysis of reciprocity (multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
	Reciprocity, Multi-Round IG (ALE)						
R	Putamen (dorsal striatum)		29	17	–1	1.80	2912
R	Head of Caudate Nucleus (dorsal striatum)		7	15	–4	1.40	304
L	Head of Caudate Nucleus (dorsal striatum)	25	–5	12	–13	1.31	496
L	Tuber (cerebellum)		–42	–73	–23	2.03	1344
R	Tonsil (cerebellum)		28	–47	–38	1.77	680
R	Culmen (cerebellum)		39	–48	–20	1.42	488
R	Culmen (cerebellum)		6	–65	–22	1.48	352
R	Anterior Lobe (cerebellum)		36	–58	–30	1.58	304
R	Premotor Cortex (precentral gyrus)	6	–57	3	30	1.76	664
L	Premotor Cortex (middle frontal gyrus)	6	–34	–5	48	1.48	352
R	Supplementary Motor Area (medial frontal gyrus)	6	–3	–2	50	1.80	760
R	Ventrolateral PFC (inferior frontal gyrus)	47	41	17	–19	1.18	352
L	Dorsomedial PFC (medial frontal gyrus))	10	–5	55	2	1.60	624
R	Dorsomedial PFC (medial frontal gyrus)	9	3	56	29	1.12	176
R	Dorsal Anterior Cingulate Cortex	32/24	6	16	36	2.36	704
L	Dorsal Anterior Cingulate Cortex	32	–6	16	36	1.09	128
R	Ventromedial PFC (medial frontal gyrus)	32/10	14	46	0	1.57	264
R	Dorsolateral PFC (middle frontal gyrus)	9	36	41	28	1.83	768
L	Dorsolateral PFC (middle frontal gyrus)	9	–46	30	25	1.27	296
L	Dorsolateral PFC (superior frontal gyrus)	9	–12	64	24	1.57	264
L	Dorsolateral PFC (superior frontal gyrus)	8	–12	46	50	1.57	224
L	Rostrolateral PFC (superior frontal gyrus)	10	–28	58	7	2.02	1200
R	Rostrolateral PFC (superior frontal gyrus)	10	36	61	4	1.17	184
L	Middle Temporal Cortex	21	–58	–63	7	1.97	616
R	Inferior Temporal Cortex	37	51	–60	–13	1.83	2000
L	Inferior Parietal Lobule	40	–48	–33	37	1.39	440
R	Inferior Parietal Lobule	40	41	–43	47	1.39	528
R	Superior Parietal Lobule	7	31	–60	45	2.15	1584
L	Superior Parietal Lobule	7	–30	–57	47	1.48	464
R	Extrastriate Visual Area (lingual gyrus)	18	30	–94	–7	2.00	952
L	Extrastriate Visual Area (inferior occipital gyrus)	18	–32	–94	–9	1.04	152

Lat, laterality; L, left; R, right.

Table 11 Results of ALE image-contrast analysis of trust vs. reciprocity (multi-round IG)

Lat	Brain Regions	BA	MNI Coordinates (mm)			ALE $\times 10^{-2}$	Size (mm^3)
			x	y	z		
	a) Trust ∩ Reciprocity (ALE)						
L	Rostrolateral PFC (superior frontal gyrus)	10	–27	56	6	1.31	248
R	Putamen (dorsal striatum)		24	12	1	1.79	456
	b) Reciprocity > Trust (z score)					**Z**	
R	Ventrolateral PFC (inferior frontal gyrus)	47	42	18	–18	3.72	160
R	Anterior Insula	13	42	10	–3	3.72	392
R	Precuneus /Superior Parietal Lobule	7	31	–60	45	3.54	1568
R	Inferior Parietal Lobule	40	38	–45	46	3.72	200
R	Inferior Temporal Cortex (fusiform gyrus)	37	51	–65	–9	2.62	152
R	Inferior Temporal Cortex (fusiform gyrus)	37	43	–49	–21	3.04	112
L	Premotor Cortex (precentral gyrus)	6	–56	4	31	3.54	352
R	Primary Visual Cortex (inf. occipital gyrus) / Extrastriate Vis. Area (lingual, mid occip. & fusiform gyri)	17/ 18	30	–94	–7	3.04	904

Lat, laterality; L, left; R, right

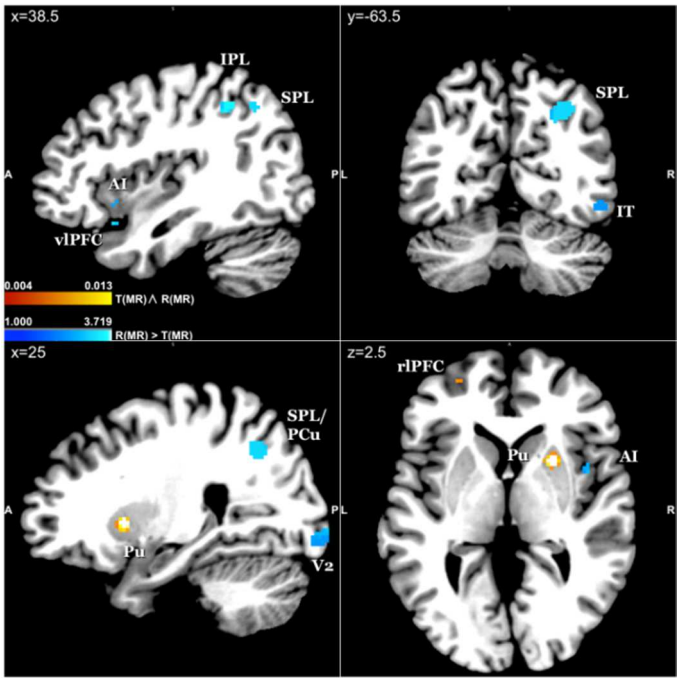


Figure 6. Results of ALE image-contrast analysis of trust vs. reciprocity (multi-round IG).
Trust ∩ Reciprocity (Red): peaks in rlPFC, NAc (not shown) & putamen. Reciprocity > Trust (Blue): peaks in AI, vlPFC, SPL, IPL, IT & V2. Random-effects analysis: 5,000 permutations; *z*-score values; q(FDR) < 0.05; min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.

Section 2.4 Discussion

Subsection 2.4.1 Scope

The combined product of a series of meta-analyses discussed above is a structure and layout of three domain-general neurocognitive networks of trust – the result of a research mission that has never been previously undertaken. In Chapter 2, trust was examined in its two most prominent forms – unconditional (dispositional) and conditional (trust in reciprocity). The key constituent property of trust was the degree, to which participants would be willing to invest (entrust) a share of their endowment with others under conditions of social uncertainty. The results obtained in the meta-analysis for Inquiry 1 are the reflection of varying dynamics between the two forms of trust, the forms that are similar in some of their cognitive constituents but are dissimilar in terms of conditions that might sustain them.

Correspondingly, in this analysis, the dynamic of these two neurocognitive systems was modeled through the use of an experiment design specifically tailored to highlight the distinction between the unconditional (one-shot IG) and conditional (multi-round IG) trust. Accordingly, in a multi-round IG – in a paradigm with a repeated delegation, the expectations about the trustee would be confirmed by the trustee's continuous response. In contrast, in a one-shot IG, there would be no such assurance and trustors would be deliberately forced to unconditionally rely on their default positive beliefs.

Subsection 2.4.2 Trust Dichotomy

The dichotomy of trust was reflected in the dissociation between the neurocognitive patterns elicited by the two variants of IG. Brain activity modulated by unconditional trust was confined to a concise set of regions that included the anterior insula, fO, dACC and hippocampus. Conditional trust on the other hand, would reveal a far more complex pattern (or rather, a series of patterns) involving cortico-striato-thalamo-cortical loops along with the occipito-parietal (dorsal) and occipito-temporal (ventral) visual streams. Among the regions that were strongly consistent for the one-shot IG was anterior insula. Ventral striatum had strong showing for the multi-round IG. The anterior insula vs. ventral striatum dichotomy in IG is in parallel with a series of earlier studies implicating the insular-striatal "switch" in inverse relationships between corresponding pairs of cognitive traits in some other social learning tasks [80,114]. Intriguingly, some of the regions putatively implicated in trust in earlier studies have not been detected in this analysis. The examples include dmPFC, temporal-parietal junction and amygdala.

Subsection 2.4.3 Unconditional Trust

The unconditional trust analysis revealed a generalized pattern of activation, in which three of the detected regions, namely dACC, anterior insula and frontal operculum, appeared to lay within a brain system known as cingulo-opercular network [115,116]. This network has been theorized as a neural substrate of mental phenomena known as "social value orientation" (SVO) or "cooperative phenotype" [117,118]. Prior evidence, which finds the activity in this network to be predictive of other-regarding, cooperative tendencies in

a behavior, is in agreement with several fMRI-IG studies linking this pattern to both trust [108] and trust responsiveness (reciprocity) [95,110].

In terms of the specific relative functional contribution of each of the regions in this network, the anterior insula had the strongest effect. However, the precise role of insula in trust is subject of ongoing debate in the literature, where competing theories link insula to either social aversion [27,53,79,95], prosocial orientation [110] or attention salience [95]. The ALE method does not have the capacity to test relative significance of different theories in their predictive power of behavior, but an in-depth review of the fMRI-IG literature on the subject shows that the majority favors the "aversion" hypothesis, whereby aversion is caused by ambiguity and lack of a-priori knowledge about the trustee. In the present analysis, the conditions were also created for ruling out the salience hypothesis, as the experiments selected for the analysis were counterbalanced for the effects of stimulus salience.

Subsection 2.4.4 Conditional Trust

The analysis of conditional trust sets the wheels in motion for a more in-depth analysis, while capitalizing on the knowledge of unconditional trust – a more basic form of the cognitive function. Unlike the unconditional trust, the existence of which is merely confined to the decision phase, the conditional trust presumably develops across two phases of the IG lifecycle, decision and outcome. In theory, the "meaning" of conditional trust is deeper and its properties are more dynamic than those of unconditional trust. Unlike in a one-shot IG, in which a decision to trust is as good as a guess, one's trust in a multi-round IG is not only evidence-based and derived from the direct experience with

the trustee, but is also instrumentally reinforced with every iteration. Repayment (reciprocity) of trust generates trustworthiness if the rewards and the other player's positive intentions are consistently confirmed [47,102].

Subsection 2.4.5 Goal-Setting in Trust

In Chapter 1 a theoretical groundwork for Chapter 2 was laid out the by emphasizing the importance of goal setting in a trust belief system. In support of this claim, the meta-analysis provided evidence of significant trust-modulated brain activity in rlPFC, which implies that trust in IG is manifest by goal-directed action. This conclusion is based on what is known about the anatomy and function of rlPFC [119-122]. The rlPFC may represent a "seat" of goal-guidance for performance of socially interactive tasks. In general, goal-guidance is about associating past actions and anticipated outcomes according to their value and keeping them online. But crucially, goal-guidance is recruited when a desire for an immediate reward must be suppressed for a potential of more socially significant future rewards. As suggested by the growing evidence from prior studies, the ability to selectively integrate goal-relevant features of stimuli and actions, to abstract information relevant to global needs and to apply these representations to the dynamic social context can be all linked to the rlPFC [104].

Subsection 2.4.6 Learning Trust

Importantly, goals have to be learned. High cognitive demands of executing well-motivated goal-directed actions require that cognitive control work in tandem with learning to guide actions that are implemented in the brain's motor systems. A goal-directed

cognitive function is fundamentally facilitated by a framework of instrumental and associative learning functions. This framework involves four critical elements – stimulus, response, value and goal [123]. In order to accomplish a purposeful behavioral sequence, there has to be a learning mechanism in place where a stimulus value is inspected and later transformed into an efficient behavioral plan for future actions. In support of these assumptions, the results obtained in the meta-analysis for the Inquiry 2 revealed strong activation in ventral striatum (HCd and ventral putamen) for the comparison of (outcome > decision) phases and in dorsal striatum (BCd and dorsal putamen) for the (outcome ∩ decision) phase conjunction. The revealed functional pattern of activation along the ventral-dorsal axis of the striatum allows for a potential mechanism of reinforcement learning (RL) across the two phases of the game. In an RL-based model, the role of the striatum is often framed within a context of a putative "Actor-Critic" neurocognitive architecture [123,124].

In Actor-Critic, the decision choices (policies) are managed by a neurocognitive module called "Actor", while value encoding of the choices is managed by "Critic". The RL model takes as its starting point, the notion of a "learning agent" in contact with his outside environment or with another agent. In terms of IG, the overall task for the agent (trustor) is to consistently learn and perform actions that are useful in delivering sustainable monetary and social rewards. Critic is viewed as an inspector of the potential reward value for both the actions (trust decisions) and stimuli (outcomes of trust). Actor is viewed as the brain's gating system meant to safeguard the selection of reward-maximizing moves at the expense of suboptimal action plans. The results suggest disso-

ciable contributions of ventral and dorsal striatum to the process of learning the rewards, with the ventral striatum (HCd) corresponding to the Critic reward-prediction role and the dorsal striatum (BCd and putamen) corresponding to the Actor. Subsequently for trust, Actor and Critic would be essential in predicting rewards of trustworthiness and correcting errors in the policy selection on a round-by-round basis in IG. Since learning is expected to benefit the trustor via a step-by-step improvement of his action-selection policy, the IG decision phase can be focused on the task of optimizing the policies towards best possible move (Actor, dorsal striatum). In turn, optimization is based on the outcomes of the previous moves, which are "inspected" by the Critic (ventral striatum). Critic steps in, to detect the degree and valence of an error between the reward estimates and the actions aimed at benefiting the selection. The error valence is then used as a basis for updating both the stimulus value and the policy value. If trust is met with an unfavorable response (no payback), the Actor is signaled a negatively valued error. Negativity would then drive the policy-adjustment towards lesser trust (avoidance). Positivity on the other hand, steers the response policy in the direction of greater trust. Meanwhile, the values of both, action and its outcome are promptly updated on the trial-by-trial basis facilitating a learning mechanism, whereby the conditions for a decision to trust undergo a continuous transformation towards increasingly beneficial decisions.

In terms of social learning, trust is not only a skill but also a trait strongly affected by the trustor's concern for socially appropriate personal conduct. Leading to this conclusion is the evidence from the analysis of the outcome phase revealing neural activity in

the putative OFC cortico-striatal circuitry, previously associated with mediating socially appropriate behavior (for review, see [125]).

Subsection 2.4.7 Interaction in Trust

In theory, trust in IG is predicated on the expectation that it will be eventually reciprocated. Reciprocity shall result in further intent to delegate potentially rewarding tasks to the trustee. Yet, one-shot IG and multi-round IG must be markedly different in terms of the degree of reciprocity expected. Unlike in a "starved of knowledge" one-shot IG interaction, where trust essentially amounts to acting "as if" reciprocity have already occurred and to overcoming a credibility gap, in the repeated IG, participants face a "world full of action choices" [126], while being fully-informed about each other intentions. Thus multi-round IG shall lead to overall high expectations of reciprocity and in case the ensuing reciprocity is consistent, the trustee should be earning high confidence and "strong delegation" from the trustor. The results of meta-analysis for Inquiry 3 are providing the insights into how reciprocity is differentially modulated by various motives of trust and how reciprocity compares to trust in terms of shared (e.g., benevolence) and distinct (e.g., betrayal vs. guilt) neurocognitive properties. Both trust and reciprocity where mediated by activation in ventral and dorsal striatum and rlPFC. For reciprocity in particular, the network included medial and lateral PFC, anterior insula, parietal cortex, posterior temporal cortex and the striatum.

Prior neuroimaging evidence points at distinct roles for the ventral and dorsal striatum in the network. For example, NAc in the ventral striatum has been strongly linked with a role of a "Critic" in the RL model. This role is to learn how to improve social in-

teraction by strengthening relational bonds [127] and how to ensure that everything goes as expected in terms of both values and costs and among the costs, in terms of both monetary losses and social contingencies (e.g., betrayal) [128,129]. Putamen in the dorsal striatum on the other hand, fits the role of an "Actor" by transforming a stimulus-value association into a goal-instrumental action [130,131]. While the ventral striatum as Critic is affiliated with social stimuli disambiguation, the dorsal striatum is more of an Actor, involved in action evaluation and selection. NAc has been correlated to players' having to deal with a socio-cognitive ambiguity such as a concern for a material payoff mixed with the concern for reputation with others [132,133] or their attempts to predict rewards in a mix of risky and positive value choices [78]. Intriguingly, NAc has been selectively activated when those participants who were categorized as "defectors" betrayed trust and those categorized as "cooperators", would honor it [80,110]. From this evidence, the overall mission for NAc appears to be that of a benefit inspector as it entails a strategic monitoring of earning long-term reward and minimizing costs. Dorsal striatum on the other hand, is involved with evaluating actions in terms of a successful prediction of a reward-action association [134].

Subsection 2.4.8 Reciprocated Trust

So far, the analysis of interaction between trust and reciprocity has highlighted their "mutually agreeable", commonly shared cooperative properties. Nevertheless strong evidence from the analysis has also revealed that they are distinct types of cooperative behavior. The (reciprocity > trust) comparison showed reliable activation in the parietal cortex (IPL, SPL), anterior insula, posterior MT, pars orbitalis (vlPFC), PMC and visual cortex.

The role of the parietal regions in reciprocity is unclear as IPL and SPL are among those highly interconnected and multi-modal regions that represent a complex mix of factors contributing to distinct cognitive domains. The jury is still out for a suitable explanation of a role for IPL. One possible explanation would be supported by strong neuroimaging evidence of IPL's involvement in cognitive control of hand movements for grasping [135]. This evidence has been interpreted by some as "social grasping" or "mirroring" and even mentalizing, because the way humans grasp an object presumably varies depending on the intention, with which the object is grasped [136-138]. In the context of IG, this kind of social mirroring could be an indicator of a reliably reciprocal and confidence-inducing social bond ("grasping"). Some neuroimaging studies have hypothesized a neural correlate of "social grasping" and provided evidence in favor of a "mirror neuron network" underlying this behavior [139-141]. If confirmed in further studies, the network would link into a single reciprocity network three of the regions identified in this dissertation – vlPFC, PMC and temporal cortex.

However, a competing theory linking behavioral response to actively choosing from social options rather than to passive mirroring has challenged the notion of "mirror neuron network". Proponents of the "social affordances" theory have argued [126,142] that a large portion of the "canonical" mirror neurons in reality does not respond to observation of objects as much as they do to social interactions with the environment ("social affordances"). Instead, a network, distinct in its neuronal properties from the "mirror neuron network" has evolved to produce adaptive response to the "world full of action choices" [126]. In the process of adaptation to the environment and pursuing various opportuni-

ties to fulfill the organism's goals, the initial response to the choices (i.e., "social affordances") is divided between the parieto-occipital (dorsal) and temporal-occipital (ventral) visual streams. Thus, while the identity of non-social objects is processed by the ventral stream (IT in this analysis), neurons in the dorsal stream (IPL/SPL) are more sensitive to social information. In IG, a player receives "social affordances" in terms of his payment options in the game. The payment option information enters the player's visual cortex and travels to the parietal cortex, where a set of alternative candidate actions is created to provide a response appropriate for the options at hand.

Alternatively, IPL has been previously linked with representation of numeric quantities and mathematical operations such as adding, subtracting and comparing one~ to two-digit numbers [143-145] which is the case in IG. IPL has also been observed to be engaged increasingly stronger with an increasing difficulty of tasks such as computing two operations instead of one, for example [146-149]. On that evidence, the involvement of IPL can be interpreted as serving players' capacity to evaluate the economic utility of both, their options and potential actions for the purpose of earning maximum possible reward.

Among the regions of the parietal cortex, a puzzling lack of an effect in temporal-parietal junction (TPJ) previously linked with mentalizing tasks, appears to disagree with those mentalizing studies that implicate TPJ in enabling "shared intentionality" – i.e., motivation to interact cooperatively in a social exchange. All things considered however, this outcome can be explained by the notion of selectivity of TPJ involvement in mentalizing tasks, where it's known to get selectively engaged by face-to-face human commu-

nications [150], and therefore in the IG "face-blocked", anonymous interaction it should be selectively inhibited.

While (reciprocity > trust) comparison revealed a number of consistent activations, there were no differences in activation maxima for the reversed (trust > reciprocity) comparison. The observed asymmetry in neural patterns may mark a certain distinction in the psychological causes of reciprocity as compared to trust. For example, the reciprocity-related modulation of insula might suggest that the players feel more averse to uncertainty arising from reciprocity than to uncertainty arising from trust. For example, Andreoni [151] proposed a theory whereby a trustee is considered a more vulnerable party in IG than the trustor. This proposal was explained by the asymmetry inherent in the IG paradigm design in which the trustor's benefits from the advantage of the first move. Subsequently and unwillingly the trustee gets exposed to a potentially unfair treatment and might experience inequity aversion. This explanation could be useful for the asymmetry of reciprocity but not for the asymmetry of trust in IG, whereby the trustor gets on the receiving end of a potentially unfair treatment.

Another, more plausible and empirically testable theory has also implicated a role for insula in mediating inequity aversion but on dissimilar grounds. On that account, reciprocity is driven by a dichotomy of two competing motivational forces. One motivational force concerns self-relevant gain and the other – equity for the partner [152-154]. These two forces might be the basis for the trustee interest in the partner's perspective (1^{st}-order beliefs). They can also motivate the players' to predict consequences of their actions from the players' mutual perspective (2^{nd}-order beliefs) [110,155,156]. The theory links inequity

aversion with norm-compliance tendency to avoid potential guilt for the consequences of "letting trustor down" [157,158]. The inequity aversion may stem from two different beliefs – a concern for the trustor's disappointment with the trustee payoff (1st order belief) and a concern for failing the trustor's expectations of trustworthiness (2nd order belief). Both beliefs may be rooted in the trustee's fairness-driven social norm compliance. The latter theory is in agreement with some of the neuroimaging data generated in the fMRI-IG studies. Thus, the trustee's 1st-order beliefs [53] and 2nd-order beliefs [80] have been linked to the enhanced activation in the insular cortex. In the study of 1st-order beliefs, the insula is engaged in monitoring unfairness of payoffs, while in the study of 2nd-order beliefs, the insula is linked to guilt aversion. Taken together, this evidence indicates that aversion (or avoidance) of social adversity is the most plausible common factor among the causal influences on reciprocity. At the same time, it is also apparent that the present meta-analysis points in the direction of guilt aversion as a possible explanation of motivation for reciprocity but it fails to generate any evidence to support the extension of this hypothesis to the causes of trust.

The activity in vlPFC (BA 47) could indicate a role of a logical and neuronal station in the trustee's chain of action selection processing, as vlPFC has been largely associated with resolving conflicts at the level of selection of behavioral response. The activity in vlPFC has been linked to the activity in the insula as the two work together to inhibit irrelevant behavioral responses in favor of contextually appropriate action plans [159-162]. Specifically in the IG task, this region has been linked with players' ability to resolve a decision conflict of having to choose between two competing beliefs – one based on the

response history and another based on prior beliefs [66]. The PMC region identified by the reciprocity analysis could be the ultimate associative link between the sensory stimuli encoded in visual and parietal cortices and the pragmatic response instructions encoded in the motor cortex [163].

Subsection 2.4.9 Significance

In summary, this meta-analysis has linked converging activation patterns in the domain-general neurocognitive network to trust in reciprocity and facilitated a generalized model of trust in the context of economic exchange. Overall, these findings emphasize the idea that trust is a conflict between anticipation of future rewards (NAc) and avoidance of ambiguity due to risk (insula). This conflict is resolved in the process of learning how much the trustees can be relied upon. At the neurocognitive level, trust is mediated by a wide network of frontal and parietal brain regions that are associated with stimulus-value-response i.e., instrumental or goal-directed learning. Per iteration and through feedback learning, the mental model of the trustee (Actor/Critic) is updated (striatum) with respect to goal-directed strategies (rlPFC). In turn, the iterative improvement in knowing the trustee intentions facilitates a long-term cooperative exchange. In conclusion, the presented conceptual framework provides a detailed characterization of the neural bases of trust in reciprocity and highlights their important roles in promoting trust and reciprocity during economic exchange. The model presented in this chapter is aimed to characterize the generalized cognitive and neurobiological model of trust and reciprocity in three dimensions i.e., believing in trust, learning trust and reciprocating trust. Revealed

here is the topology of the neural networks supporting the cogntive models across the IG experimental conditions.

The study of trust neurocognitive networks will now proceed to Chapter 3 exploring the role of mentalizing in the communication of value to neural circuits of choice. The model presented in Chapter 3 will demonstrate effective connectivity i.e., causal directed relations among neuronal populations. The effective connectivity will be modeled through the use of a data-driven MVGC framework and the existing hyperscan-fMRI data, to allow testing of the proposed trust models more fully i.e., at both, the individual ("within-brain") and social ("between-brains") activity levels.

Subsection 2.4.10 Limitations

The meta-analysis conducted in Chapter 2 improves the existing conceptual understanding of neural signatures of trust. However, there exist several limitations that need to be acknowledged. Only a relatively small number of papers eligible for inclusion in the analysis were ultimately selected. This, and the lack of reported parametric analyses linking brain activity and key behavioral variables of IG, led this dissertation to considering only the basic comparisons between the measured variables (e.g., 'trust' vs. 'control'). The phrase "key behavioral variables" is to denote the most important aspects of the IG experiment design, such as the amount sent to the trustee or the amount paid back to the trustor, for example. The analysis was based on the reported activation coordinates and subject counts. However, many other variables, such as the fMRI scanning parameters, data analysis parameters and some potential mediator variables were omitted, because of their high cross-study variability. The variation may have impacted the final

results along certain dimensions of the IG design, including for example, the participant category (e.g., "healthy" vs. "patient") [68,89-91], player's role-changing (e.g., "permanent" vs. "alternating") [47,102] or treatment type (e.g., "oxytocin" vs. "placebo") [103].

To account for such undesirable but sometimes unavoidable deficits, future studies should model their experiment design after the seminal paper of Baumgartner, et al. [103], in which the approach is to expand on the number and quality of demographic, psychological and cultural measurements and to bring more clarity to the study of the brain's neurocognitive networks. Based on that approach, further progress in quantitative verification of theoretical hypotheses about trust can be achieved by increasing the statistical power, external validity and refinement of trust generalized models. Raising the effectiveness of future fMRI-IG studies will also help to advance the understanding of causal and temporal relationships among the brain regions engaged by trust.

CHAPTER 3. Effective Brain Connectivity in Trust Networks

Section 3.1 Introduction

Subsection 3.1.1 Scope

The topology and functional mapping of trust neural networks identified in Chapter 2 provides the structural basis for further analysis of the networks' dynamic characteristics. Analysis in the upcoming chapters will start with defining key terms and conceptual understanding (e.g., "de-convolution", "Granger causality", "connectivity") of the effective connectivity problem domain. The analysis will then proceed with a brief description of current research as it relates to decision-making in general and trust decision-making in particular. Next, a more detailed, technical account of a novel approach to studying trust i.e., effective connectivity modeling, will be laid out and the methods that the model is built upon will be described. A composition of customized algorithms and scripts implementing the model will be analyzed in depth. The goal of the model is to compute strength of directed neuronal signaling within and between interacting brains during trust. At the end of the chapter, a brief summary of the results and the analysis of their importance to the cause of this research project, the methodological limitations as well as perspectives on future development are provided.

Subsection 3.1.2 Problem Domain

As has been noted in Chapters 1 and 2, a sound conceptual understanding of neural activity underlying trust decision-making must agree with the empirical evidence in

support of a generalized decision-making theory. In today's theory, decision-making is typically viewed as a product of two foundational neurocognitive mechanisms – one that supports choice and another that supports valuation of choices. Each of the neural systems has been in focus of independent interdisciplinary studies in psychology, economics and neuroscience, but the attempts to functionally bring together the evidence of the well-known biological substrates of the anatomically "isolated", but cognitively mutually influential constituents have not been successful [164].

Today, researchers agree on a putative neurobiological model of perceptual choice and a putative model of subjective valuation. In perceptual choice, a decision-maker is presented with sensory ambiguity of alternative signals. The decisions are then formed based on the outcomes of continuous sequential sampling of evidence related to the options at hand. The choice, i.e., physically committing decision-making to a single course of action occurs through perceptual integration, when the evidence favoring one of the options reaches a threshold for neuronal response and produces "winner-takes-all" neuronal response [9-11]. The constituent "ramping-to-threshold" and "accumulate-to-bound" brain activities appear to scale with respectively, the operation of evidence accumulation observed in lateral parietal cortex and caudate [13-16] and a subsequent activity in dlPFC, akin to integration of lower-level stimuli [12].

However, a significant limitation of the stimuli disambiguation paradigm restricts the use of the model in trust research. Thus, the model relies on the notion of decision-making dissociated from utility (i.e., relative desirability of the options) – the assumption disproved by the evidence that some kind of reward or aversion to loss is essential to de-

cision-making [17]. Evidence for a distinct set of frontal and striatal areas generating subjective value for choices on a common scale and for a wide range of rewards had emerged [18]. Studies of appetitive stimuli point at vmPFC – a target of critical inputs for valuing social, consumable and emotional rewards from the dlPFC [19], OFC [20], hypothalamus [21], amygdala [22,23] and the striatum [24]. Also, as shown in Chapter 2, the valuation-choice neural model can be expanded with the factor of "risk aversion", which introduces the notion of a trade-off between reward and aversion mediated by cingulo-opercular circuitry.

However, in spite of the abundant evidence, little is known about how the fronto-parietal (choice), medial prefrontal (valuation) and cingulo-opercular (aversion) neural circuits interact with one another. Whether the circuitry forms a tightly interconnected global network or its topology is composed of loosely coupled specialized functional modules – remains a matter of conjecture. The analysis in the upcoming chapter is set to provide evidence of functional links between the three constituents of decision-making.

Subsection 3.1.3 Method

To assist bridging the gap of knowledge described above, the "problem" of trust has to be taken out of the context of mere personal reward and choice processing and studied as a relational phenomenon of a social behavior reliant on a spectrum of interactional reward-risk trade-offs. Yet, among the existing fMRI-IG studies analyzed in Chapter 2, the majority are focused on the mental states of a single IG participant and only two, King-Casas, et al. [59] and Krueger, et al. [47], have attempted a neurocognitive model of trust as interaction. This is due to high cost of the hyperscan-fMRI technology used in the

interaction studies. However, the studies take advantage of a crucially important property of the hyperscan-fMRI technique as it relates to study of a social behavior like trust, in that the technique allows a researcher to concurrently image the brains on both sides of an interactive task [46].

However, while the hyperscan-fMRI-IG studies succeed in elucidating some of the time-order neurocognitive relations for trust, they produce little evidence on causal influence among the neurocognitive constituents at hand. The result is that today we know much more about where some of the trust functionally segregated signals are localized in the brain than about how the causal influences within and between brains are facilitated and communicated across neuronal populations and persons. Furthermore, despite the plethora of information on the functional mapping of trust, it can be easily predicted that any attempt to localize a cognitive function to a segregated set of brain regions will be met with difficulty. This is because a complex cognitive function is more often than not is mediated by a network of associated regions rather than by a segregated functional module in the brain [45].

Therefore, an improved understanding of the neurocognitive model of trust is predicated on the knowledge of effective connectivity or causal neurocognitive links among the regions of a functional network in the brain or causal cognitive influence that a neuronal population in one interacting brain can exert over the other. To test for the interpersonal and intertemporal neurocognitive dynamic influences within and across the participating brains, the effective connectivity analysis not yet undertaken in any of the earlier studies of IG has been performed as part of this dissertation. The ECA was im-

plemented using MVGC [51] – a robust data-driven method predicated on exploring the structure of a dataset in pursuit of task-related effects. This method was chosen for its capacity to detect causal dynamic influences between any pair of related stochastic processes. In that capacity, it was applied to the analysis of dynamic neuronal signal *a priori* deconvolved from the "raw" fMRI signal.

Subsection 3.1.4 Research Inquiry

Based on the converging evidence from the studies analyzed above, it can be conjectured that in the absence of direct anatomical links and strong presence of functional, cognitive links between the basic components of a decision-making system, there has to exist a third, "mediator" component of the model with the established anatomical and functional connections to both of the basic components. From what we know today about the functional anatomy of the "social brain" a candidate region can be identified for its unique connections to both, the reward substrate vmPFC and action selection neuronal substrate, the parietal cortex. This region specifically, dmPFC could play a role of a "missing link" i.e., the anatomical and functional connectivity between the brain circuits of action choice (parietal cortex) and valuation of choices (vmPFC). The analysis of evidence in support of this assumption is provided in the upcoming Chapter 3 analysis.

Section 3.2 Data Preprocessing

Subsection 3.2.1 Data Source

From what has been noted above, trust as a relational (social, interactive) phenomenon can be measured in two neurocognitive dimensions: within an individual (trus-

tor or trustee) or between the subjects i.e., in their brains during their interaction with each other. In Chapter 3 it is attempted to build the brain maps of causal regional connectivity underlying the cognitive dimensions of trust interaction. The analysis relies on existing data from one of the two earlier hyperscan-fMRI experiments mentioned above [47]. To make the data source consistent with the goals of the present analysis, BOLD time series for the ROI during trust and reciprocity decision-making (experiment and control conditions) were selectively retrieved from the source dataset and a priori de-convolved to make the data suitable for the subsequent ECA.

Subsection 3.2.2 Participants

At the start, the dataset was divided in the "defectors" and "cooperators" data, based on the valence of participants' response during the experiment. In line with the final goal of this dissertation, which is to explore the incentives for cooperative and not divisive behavior, the outcomes for the "cooperators" were selectively included in the new dataset and the "defectors" data were omitted. The cohort comprised 44 data records, representing 22 pairs of participants who were healthy, native English speakers, matched by sex (22 males, 22 females), age (28.3 ± 7.1 [mean age ± s.d.], range 21-51 years) and education (17.3 ± 2.2 [mean level of education ± s.d.], range 12-23 years), right-handed as determined by Edinburgh Handedness Inventory (95.3 ± 8.7 [mean right-handedness ± s.d.], range 65-100 points) with normal or corrected-to-normal vision and no history of neurological/psychiatric disorder or medication use. Participants provided a written informed consent approved by Institutional Review Board of the National Institute of Neurological Disorders and Stroke (NINDS).

Subsection 3.2.3 Functional Analysis

In the original experiment, continuous whole brain imaging was performed as the participants were engaged in the IG task. For this dissertation, the outcomes were analyzed using whole-brain and GLM data analyses, with the criteria for activation to reach a threshold for the main effects of trust and reciprocity. The results were thresholded to correct for multiple comparisons using $q(\text{FDR}) < 0.005$ at a minimum cluster volume of 100 mm^3. Images were analyzed using BrainVoyager QX (Brain Innovation) and custom MATLAB (The MathWorks Inc., Natick, MA, US) scripts. Statistical images were overlaid onto a normalized brain template (Talairach space) and thresholded at $P < 0.005$, uncorrected, with extent threshold of 100 mm^3 ($t = 3.00$, random effects). Data preprocessing involved slice scan-time correction (‘‘sinc’’ interpolation), linear trend removal, temporal high-pass filtering of low-frequency nonlinear drifts of three or fewer cycles in a time course, spatial smoothing (8-mm FWHM), 3D motion correction of small head movements and spatial alignment of all participants to the first volume by rigid-body transformation procedure. Estimated translation and rotation parameters were inspected and never exceeded 2 mm or 2°.

Subsection 3.2.4 Data Normalization

Throughout the analysis the data was encoded in different ways. The fMRI functional analysis produced a single “one size fits all” Excel data sheet to allow plenty of redundancy for the pragmatic purposes of legibility to humans. Such data structure was as good for human perception as it was bad for computer processing. Redundancy would have violated a number of data normalization principles [173] and would have likely result-

ed in errors and endless back and forth patch-up and maintenance. With that in mind, a normalized Excel-based database system was put in place. Also, an additional script was created to convert and import the source data into the new database. For the sake of significant data reduction, normalization and efficiency, the behavioral and neuroimaging (BOLD time series) data were split in two separate tables and logically linked back together at the run time by enforcing a set of association rules in the code. Further, the number of data elements was reduced to a necessary minimum and each data element was represented by a set of unique data records. To speed up processing, all the text data elements were enumerated. These efforts allowed for a significant reduction in the overall volume of data and hence for considerable improvement in performance. Once the new database structure was in order, a preprocessing script was created, to import the data into the new database. At the end of this initial preprocessing sequence, another script was created to upload the converted data into the memory at run time.

Subsection 3.2.5 Suitability of fMRI Data

The main goal of ECA was to produce pairwise connectivity values for the ROI modulated by the implied neurocognitive constituents of trust. Accomplishing this goal was predicated on overcoming a large number of technical challenges to the task of bridging differences between the fMRI and MVGC analysis techniques. The selected ROIs' BOLD signal time series representing brain's HR were extracted, averaged across voxels and normalized across the participants. However, proceeding with the analysis was predicated on a successful data conversion and preprocessing. The rationale for adding yet another step to the already complex computational cascade was to avoid data re-

dundancy as well as missing and faulty data – the potential causes of computational errors affecting the subsequent stages of the modeling process. Another important factor was suitability of data for the MVGC analysis model.

One of the challenges to achieving that objective concerned a series of drawbacks associated with BOLD – the primary measure in fMRI. The signal measured in fMRI neuroimaging experiments is a result of linear convolution between the brain's neuronal and subsequent HR to stimuli. BOLD represents only a secondary, indirect manifestation of the primary neuronal signal of interest. As a consequence, eligibility of BOLD data for a MVGC analysis is hampered by a reduced spatial resolution, lowered signal-to-noise ratio and lag in relation to the latent neuronal signal. Another obstacle to using BOLD data in a MVGC analysis is that it comes as a nonlinear and non-stationary HR, which is also characterized by high level of variability across participants and brain regions [165,166]. The nonlinear and non-stationary properties of HR complicate estimation of the interregional connectivity by distorting the underlying latent neuronal signal.

Subsection 3.2.6 Deconvolution

To account for issues like the HR's nonlinearity and variability across participants and regions [167], the fMRI data must first be converted into a properly approximated biophysical representation of the latent neuronal signals. This intermediate but critical goal can be achieved via a complex cascade of signal processing techniques termed "blind deconvolution" [168], which includes a combination of signal processing algorithms such as dynamic expectation maximization (DEM) [169] and cubature Kalman filtering (CKF) [170]. In the past, this technology has proven highly robust in application to nonlinear dynamic

systems, representing hidden neuronal states. In this dissertation, the efficiency of forward-pass CKF was finessed via the use of backward-pass cubature Rauch–Tung–Striebel smoother [171]. In addition, the efficient square root formulation of these algorithms facilitated consistent and accurate estimates of the hidden neuronal time series, well beyond the temporal resolution of fMRI. Using the blind deconvolution techniques in this model has appeared extremely efficient in estimating the latent neuronal signaling needed for the input into the subsequent MVGC analysis [50,172].

Section 3.3 Method

Subsection 3.3.1 IG Task

In the experiment conducted by the source study, participants were matched in pairs of "trustor" vis-a-vis "trustee" and were involved in a series of repeated voluntary IG, with players randomly alternating their roles in every round. Payoffs in cents were presented as a binary decision tree with an option to quit and an option to share. A two-player IG can proceed in three possible scenarios resulting in three possible outcomes. Most mutually rewarding for the players is to share with each other. A less rewarding scenario is for the trustor to abstain from investing and to trigger equally split but reduced earnings. The third scenario is for the trustee to abstain from repaying trust and benefit at the trustor's expense. Thus, for each round of the game, trustor's withholding the money would get equal but lower [5, 5] payoff for both players. Trustor's investing and being paid back on the investment would result in a greater share [10, 15] for each player. Trus-

tor's investing and not being paid back would yield nothing to the trustor and everything to the trustee [0, 25].

Subsection 3.3.2 IG Timeline

The timeline for a round of the game (**Figure 7**) included a 2 second introductory phase (a display informing the players of their roles), a sequence of two 6 second decision phases (a display informing one player of his payment options, while the other player is waiting), 4 second feedback phase (decision outcomes displayed to both players) followed by the blank screen with a jittered inter-stimulus interval lasting 2 to 6 seconds. No information about cumulative earnings was provided during the experiment. Timeline of the experiment was *a priori* divided in two distinct stages: "partnership building" (Stage 1) and "partnership maintenance" (Stage 2). Primary goal of the subsequent analysis was to provide connectivity evidence in support of such division. Within each stage, opponents sequentially performed 36 rounds of the IG task and 16 rounds of a control task (designed to control for sensorimotor and material aspects of decision-making). In the experimental condition, participants randomly alternated their roles in rounds of a voluntary IG, playing half of the time as "trustor" and half of the time as "trustee". In the control task with the identical timeline and screen layout, participants were to merely choose from a pair of computer generated monetary rewards. The source data were obtained through the use of event-related hyperscan-fMRI of unacquainted participant pairs matched by sex. While in the process of performing the task, participants on both sides of the exchange were simultaneously imaged in separate scanners.

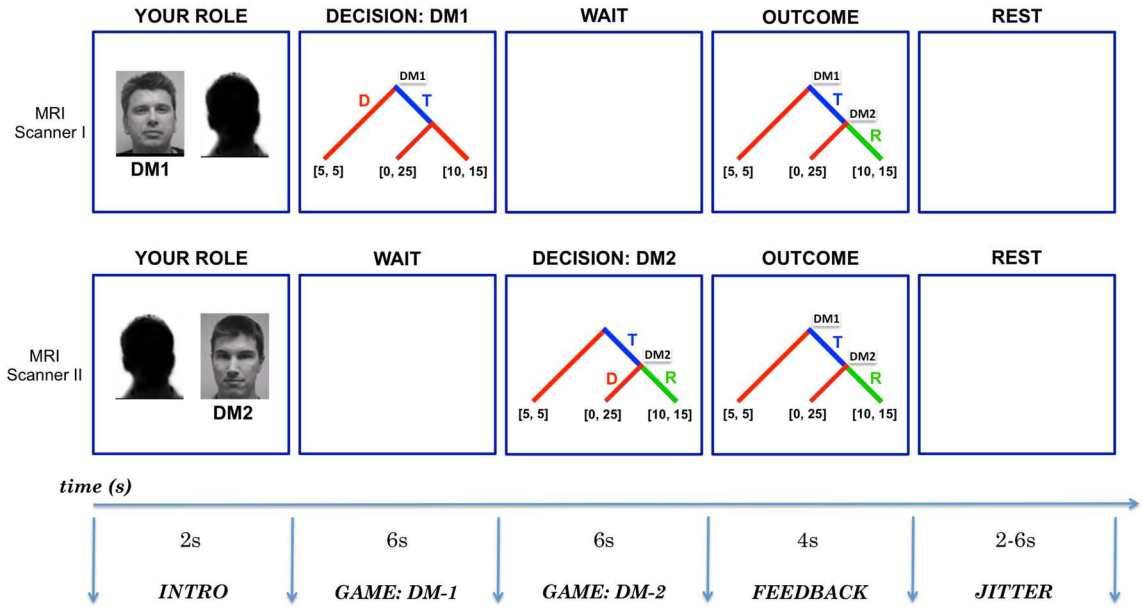


Figure 7. Timeline of a round of the multi-round voluntary IG
Adapted from "Neural Correlates of Trust" by Krueger et al., 2007, Proc Natl Acad Sci U S A, 104 (50), pp. 20084-9. Copyright 2007 by The National Academy of Sciences of the USA.
"Intro", a 2 second display informing players of their roles (DM1, DM2, decision-makers (players) 1 and 2); "Game: DM", a sequence of two 6 second decision-making phases – one for DM1 followed by the one for DM2; "Wait", 6 second wait while the partner is deciding; "Feedback", a 4 second outcome processing phase; followed by a blank screen with a jittered inter-stimulus interval of 2–6 seconds.

Subsection 3.3.3 Multivariate Granger Causality

MVAR Modeling. ROI generated by the whole brain data analysis were subsequently tested to quantify the strength and causal influence (i.e., effective connectivity) of their interactions across the network. The assumptions were tested using the MVGC mapping [174]. A property of MVGC key to this analysis is its capacity for predicting an unobservable signal based on a functional approximation of intertemporal relationships between the signal and relevant observable signals. This possibility can be realized in multivariate autoregressive (MVAR) modeling [175] – a practical implementation of MVGC tailored to fMRI needs. In MVAR a process "X" is said to have causal influence

on process "Y", if considering past values of both time series "X" and time series "Y" improves the future prediction of time series "Y" (**Equation 1**).

Equation 1. MVAR Principle

$$Y(t) = \sum_{k=1}^{p} a_k Y(t-k) + \sum_{k=1}^{p} b_k X(t-k) + e(t)$$

In terms of brain connectivity, this principle translates into a concept where one region in the brain is assumed to have a directed influence on another, if a combination of time courses of activation in the source and target regions allows predicting the temporal progression of activity in the target region. The between-brain relationships implicate causal flow between two regions in two different brains. As such, they cannot be interpreted in terms of physical connectivity and instead, should be viewed as a predictive relationship between the brains such that activity in one brain at one instant is replicated in another brain after a time lag.

Algorithm. A series of MATLAB scripts were developed in close collaboration with Dr. Gopikrishna Deshpande at the Department of Electrical Engineering, Auburn University, AL. The proposed method of ECA involved four major steps as illustrated in **Figure 8**. First, the source data required conversion and normalization. Second, the BOLD time series extracted from the ROI functional analysis had to be de-convolved to further extract the underlying neuronal response time series. Third, the resulting neuronal time series were fit into the MVAR model to compute the network of connectivity path-

ways for trust and reciprocity and for within and between brains. After obtaining the connectivity measures the statistical testing was applied to extract significant effects.

Subsection 3.3.4 Deconvolution

Proceeding with the analysis was predicated on converting the source fMRI data to a form suitable for the MVAR model. Due to the limitations of the fMRI method described above any enhancement of the input temporal resolution would have to rely on a series of complex signal processing algorithms termed "deconvolution" [176]. The algorithm adopted from Dr. Deshpande, was tailored to the needs of this analysis in a customized script. The output of the script is a 3-D matrix (N_{TP} X N_{PR} X [2 X N_{ROI}]) of the neuronal time series that are used as input for the next, connectivity stage of the analysis. N_{TP} denotes the number of points in a time series, N_{PR} is the number of participant pairs and 2 *X* N_{ROI} is the number of regions of interest in a pair.

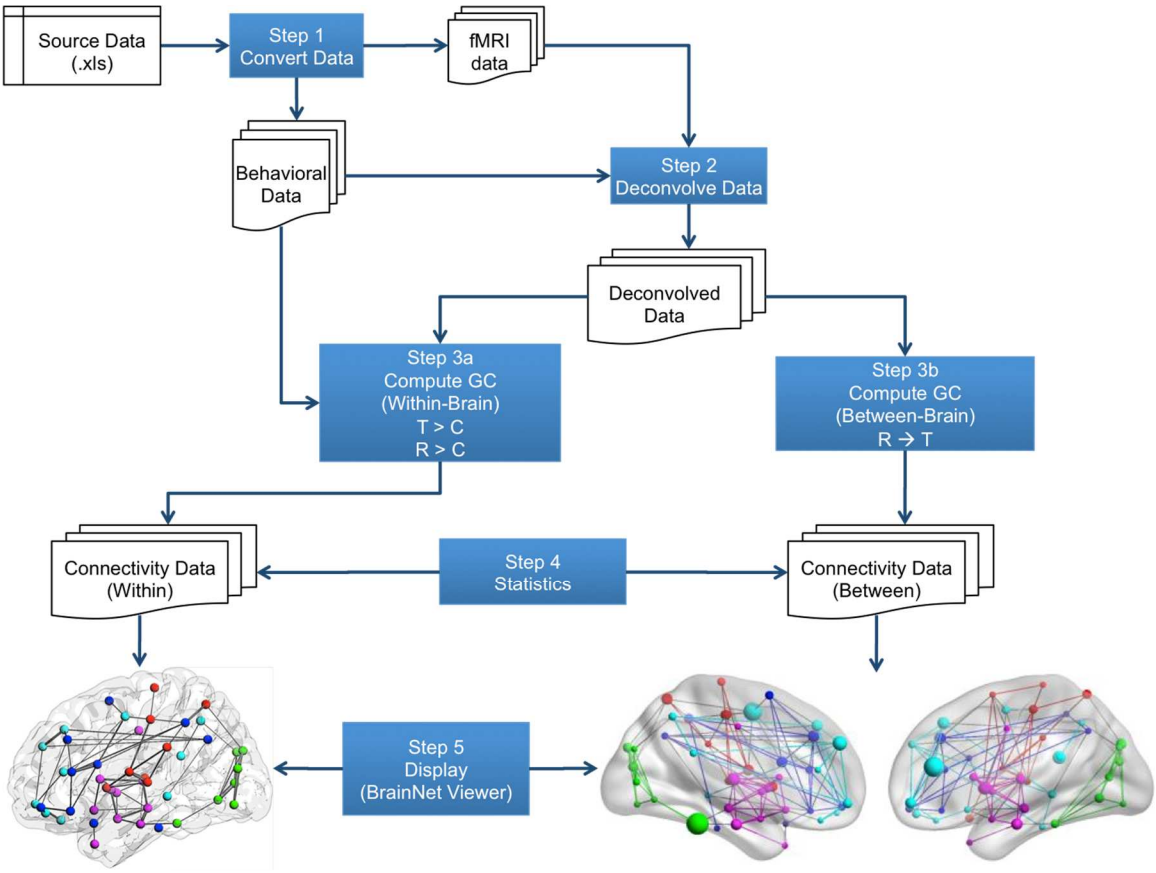


Figure 8. Computing Effective Connectivity.
Within brain: "T > C", trust; "R > C", reciprocity. Between brains: "R → T", reciprocity to trust.

Subsection 3.3.5 Effective Connectivity Analysis

In the next step, a novel algorithm building on the strength of the original MVAR model was developed to provide cross-region connectivity predictions for the deconvolved neuronal signals [177,178]. The model coefficients were allowed to vary as a function of time, so that the condition-specific connectivity values could be obtained [179]. Boxcar functions corresponding to the trust and reciprocity conditions were subsequently used to extract connectivity values for the within- and between-brain contexts. Three ma-

jor steps in the analysis included: (1) Optimization of the model: choosing the right values for Bayesian information criterion (BIC) i.e., the "order" or "memory" of the system and for "Forgetting Factor" (ff) i.e., how quickly the model would converge on a predicted connectivity value; (2) Obtaining the connectivity measures; (3) Testing for statistical significance.

Choosing the optimal value for BIC was predicated on how many predictors (lag values) were available within a decision window (**Figure 9**). With brain images taken every 2 seconds during a 6 second decision phase, there is a total of 3 time points in a decision. The diagram does not allow depicting any of the within-brain mapping, but the process of obtaining a connectivity measure is illustrated by the arrows at the bottom (blue, "trust"; green, "reciprocity", red "idle"). The middle arrows illustrate how the neuronal signal time series are mapped between the trustee and trustor brains. An estimate of connectivity at a given point in a neuronal signal time series (e.g., t_{T3} for the trustor, t_{R3} for the trustee) is based on the current value of the signal and a maximum number of lag values (e.g., t_{T1} and t_{T2} for the trustor, t_{R1} and t_{R2} for the trustee).

Key to ECA was the notion of connectivity pathway defined by strength (numeric value) and direction. Both can be derived from the output of the ECA function – a 4-D connectivity matrix (N_{TP} X N_{PR} X N_{ROI} X N_{ROI}) – according to Equation 2.

Equation 2. MVAR Matrix
2, number of stages, N_{TP}, number of time points per stage. N_{PR}, number of participant pairs. 4, number of conditions of interest = 2 within-brain + 2 between-brain; N_{ROI}, number of ROI = N_{ROI} for trustor + N_{ROI} for trustee.

$$M = 2 \times N_{TP} \times N_{PR} \times 4 \times (N_{ROI} \times N_{ROI})$$

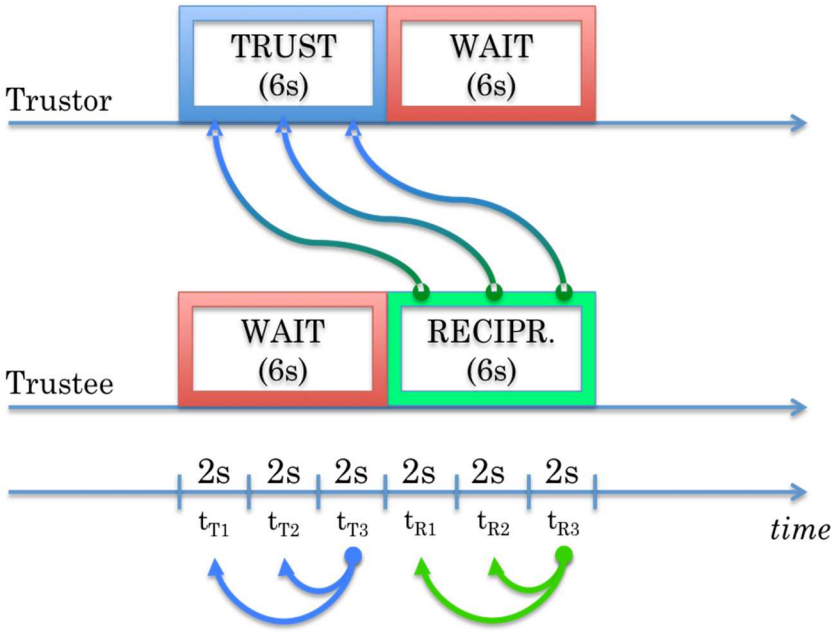


Figure 9. Defining the Order of the Model.
Within Brain: (Blue) Trust (t_{T3}, connectivity estimate; t_{T1} & t_{T2}, connectivity predictors); (Green) Reciprocity (t_{R3}, connectivity estimate; t_{R1} & t_{R2}, connectivity predictors. Between Brains: (Green-to-Blue) Reciprocity-to-Trust.

Equation 2 and **Figure 10** illustrate how the matrix defines connectivity mapping for a pair of ROI as a function of condition (within-brain/between-brain), player's role (trustor/trustee) and placement (scanner 1/2) at a particular instant in a time series. Thus, a within-brain connectivity value is found in a quadrant representing one player, while the between-brain values are found in a crosshair of two players *vis-à-vis* each other. The direction of connectivity is set as "row-to-column".

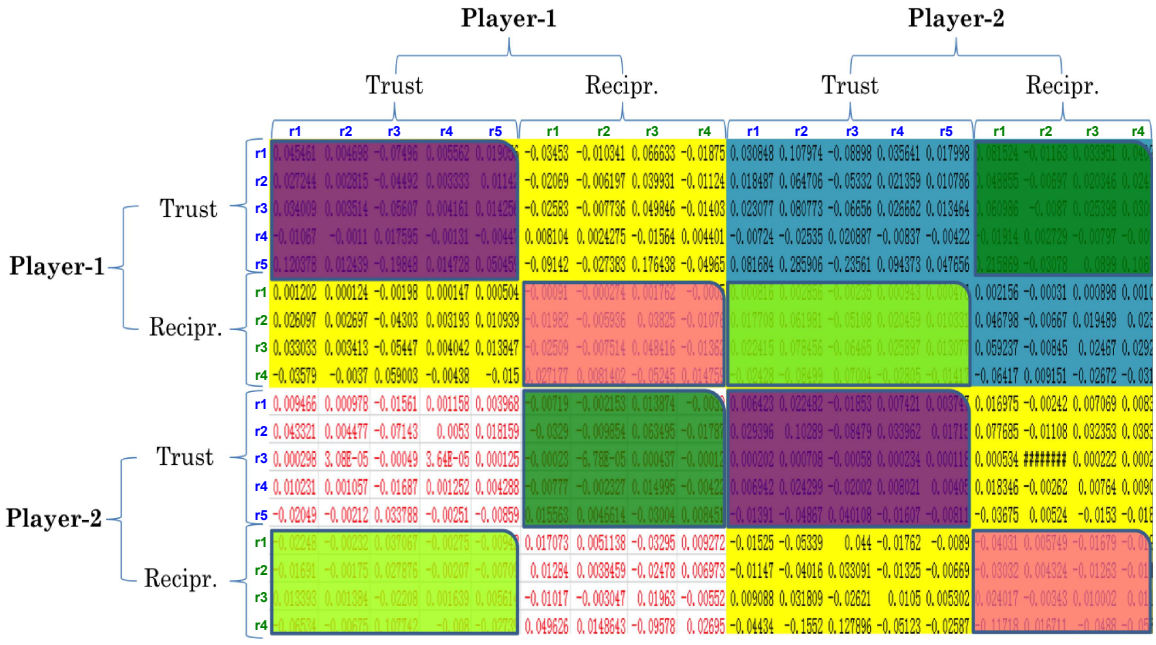


Figure 10. Connectivity Strength and Direction (Row-to-Column) Selection Criteria.
Within brain: trustor (purple), trustee (pink). Between brains: trustor *vis-a-vis* trustee (forest green), trustee *vis-a-vis* trustor (Lime Green). ROI: r1-r5 (blue) – trustor; r1-r4 (green) – trustee.

In the final step of the analysis, the connectivity values were tested for statistical significance in a human-to-human condition of interest (i.e., trust-within, reciprocity-within, reciprocity-to-trust-between) compared to a human-to-nonhuman or baseline control condition. Pairwise and between sample t-tests were performed for the within-brain and between-brain contexts, respectively with the FDR-correction for multiple comparisons thresholded at $q(\text{FDR}) < 0.05$. The outcomes of the analysis were then converted from a numeric form into a user-friendly graph format (termed "glass-brain plot") and overlaid onto a normalized brain template using "BrainNet Viewer" [180] – a MATLAB-based toolbox designed to visualize brain connectomes through the use of Statistical Parametric Mapping 8 functions (SPM, http://www.fil.ion.ucl.ac.uk/spm/).

Section 3.4 Results

Subsection 3.4.1 Functional Analysis

Neuroimaging data pertinent to the analysis was retrieved from the Krueger et al. (2007) study. The ROIs were selected on the basis of their enhanced activation for a behavioral condition (trust or reciprocity) and formed distinct neural networks for trust: right dmPFC (medial frontal gyrus, BA 9); PCC (cingulate gyrus, BA 23); temporal polar cortex (TP, superior temporal gyrus, BA 38); hippocampus (parahippocampal gyrus, BA 35) and hypothalamus (**Table 11**); and for reciprocity: right dmPFC (superior frontal gyrus, BA 9), dlPFC (middle frontal gyrus, BA 8), lOFC (inferior frontal gyrus, BA 11) and precuneus (BA 7) (**Table 12**).

Table 12 Results of whole brain functional analysis of trust and reciprocity

Lat	Brain Regions	BA	MNI Coordinates (mm)			t
			x	y	z	
	a) Trust (Trustor)					
R	Dorsomedial PFC (medial frontal gyrus)	9	6	45	23	4.7
R	Posterior Cingulate Cortex	23	5	–15	35	5.3
R	Temporal Polar Cortex (superior temporal gyrus)	38	43	7	–33	4.7
R	Hippocampus (parahippocampal gyrus)	35	24	–35	–8	4.6
R	Hypothalamus	–	9	–3	–8	4.7
	b) Reciprocity (Trustee)					
R	Dorsomedial PFC (superio frontal gyrus)	9	9	49	40	3.2
R	Dorsolateral PFC (middle frontal gyrus)	8/9	38	7	47	3.1
R	Orbitofrontal Cortex (inferior frontal gyrus)	11	33	32	–37	3.2
R	Precuneus (medial parietal wall)	7	5	–79	50	3.0

Voxel-wise $P < 0.001$ in conjunction with a cluster-size threshold of 40 voxels, voxel size = 2 × 2 × 2 mm. Lat, laterality; R, right hemisphere; L, left hemisphere.

Subsection 3.4.2 Behavioral Data

Participants in the source study had a strong tendency for prevalence of cooperative incentives. Trustors opted to trust more often (84%) than not (16%) and the trustees opted to repay trust more often (77%) than defect (7%). However, the trustees differed in levels of their concern for the partner ("cooperators" vs. "defectors"). For the "cooperators" group no trustee ever defected their partner's decision to trust, whereas for the "defectors" group trustors experienced some defections.

Subsection 3.4.3 Connectivity within Brain

To identify the effective connectivity of interest, a data-driven MVGC analysis was performed for the regions surviving the statistical threshold of $q(\text{FDR}) < 0.05$. The analysis revealed varying connectivity across the trust/reciprocity/within-/between-brain conditions. Significant causal pathways are depicted graphically in **Figures 11-16**. For trust, three regions in the network (TP, hypothalamus and hippocampus) formed two star-like topologies, one with dmPFC and another with PCC. The latter two regions established a bidirectional connection with each other. This connectivity pattern was preserved across stages, but the connectivity strength varied. In Stage 1, dmPFC was first in the order of strength, followed by TP, PCC, hypothalamus and hippocampus (**Figure 11, Table 13**). In Stage 2, TP was first and dmPFC was second in the order of strength (**Figure 12, Table 14**). For reciprocity, precuneus, dmPFC and lOFC formed a star-like connection [181] with the central dlPFC region and both in Stage 1 (**Figure 13; Table 15**) and Stage 2 (**Figure 14; Table 16**).

Table 13 Effective Connectivity during Trust (Stage 1, Partnership Building)

ROI (BA)		Connectivity	
Source	**Target**	**Stage 1**	**P**
Trust (Trustor)			
dmPFC (9)	PCC (23)	–0.100650	0.00470
PCC (23)	dmPFC (9)	–0.068874	0.00012
TP (38)	dmPFC (9)	–0.092693	0.01270
TP (38)	PCC (23)	–0.077045	0.00003
Hippocampus (35)	dmPFC (9)	–0.067079	0.00002
Hippocampus (35)	PCC (23)	–0.068064	0.00002
Hypothalamus	dmPFC (9)	–0.073249	1.06×10^{-7}
Hypothalamus	PCC (23)	–0.074405	1.84×10^{-7}

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

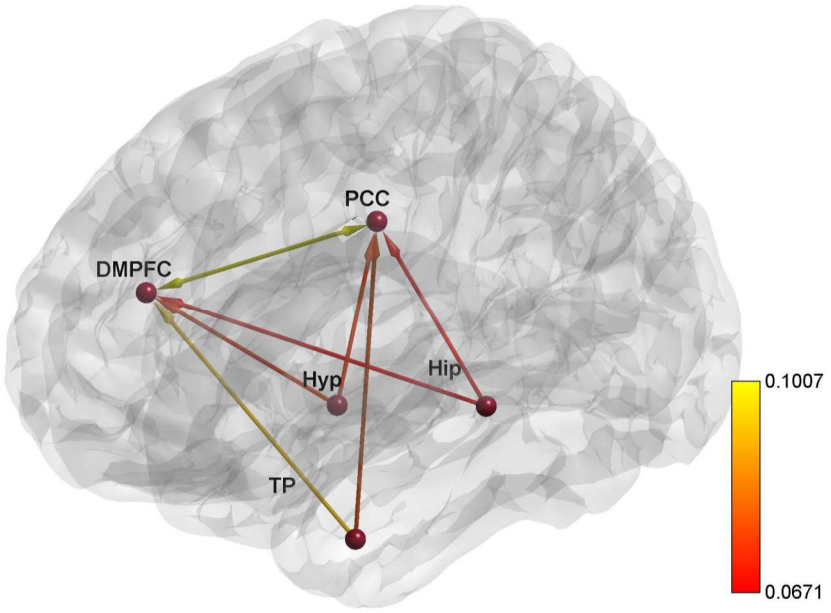


Figure 11. Effective Connectivity during Trust (Stage 1, Partnership Building).

Table 14 Effective Connectivity during Trust (Stage 2, Partnership Maintenance)

ROI (BA)		Connectivity	
Source	**Target**	**Stage 2**	**P**
Trust (Trustor)			
dmPFC (9)	PCC (23)	–0.090525	0.00006
PCC (23)	dmPFC (9)	–0.070257	0.00310
TP (38)	dmPFC (9)	–0.101540	0.00034
TP (38)	PCC (23)	–0.076091	1.07×10^{-6}
Hippocampus (35)	dmPFC (9)	–0.064154	1.30×10^{-9}
Hippocampus (35)	PCC (23)	–0.065877	1.49×10^{-9}
Hypothalamus	dmPFC (9)	–0.064582	0.00680
Hypothalamus	PCC (23)	–0.066605	0.00019

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

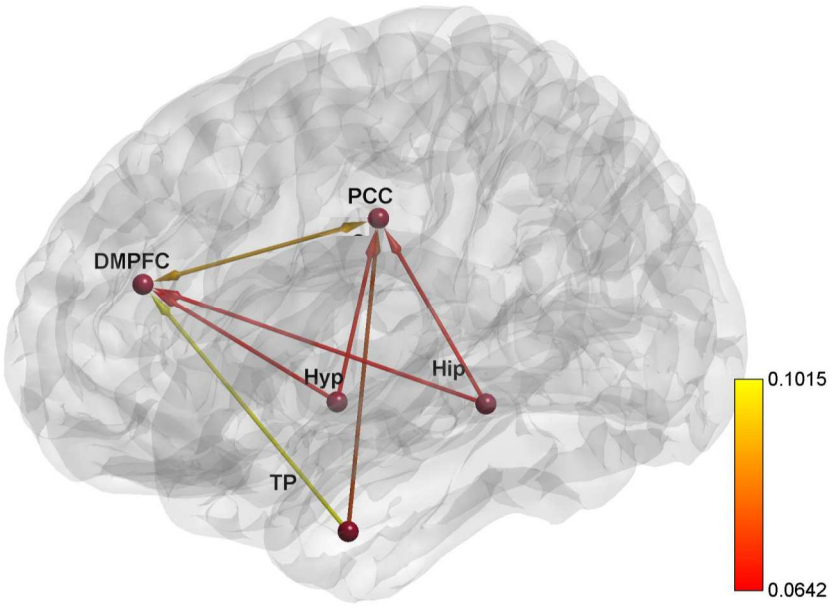


Figure 12. Effective Connectivity during Trust (Stage 2, Partnership Maintenance).

Table 15 Effective Connectivity during Reciprocity (Stage 1, Partnership Building)

ROI (BA)		Connectivity	
Source	**Target**	**Stage 1**	**P**
Reciprocity (Trustee)			
dmPFC (9)	dlPFC (8)	–0.059914	4.50×10^{-6}
Precuneus (7)	dlPFC (8)	–0.061478	1.41×10^{-8}
OFC (11)	dlPFC (8)	–0.082516	0.00024

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

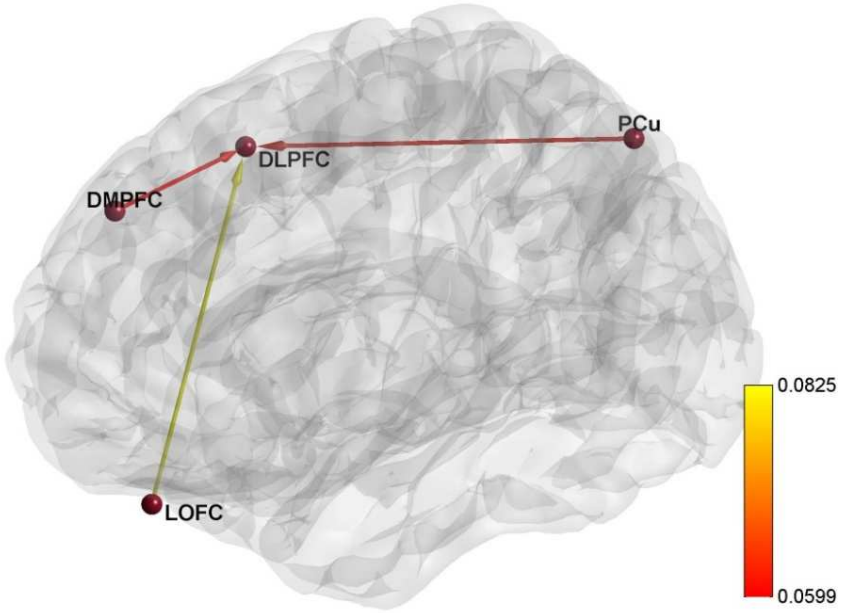


Figure 13. Effective Connectivity during Reciprocity (Stage 1, Partnership Building).

Table 16 Effective Connectivity during Reciprocity (Stage 2, Partnership Maintenance)

ROI (BA)		Connectivity	
Source	**Target**	**Stage 2**	**P**
Reciprocity (Trustee)			
Precuneus (7)	dlPFC (8)	–0.067855	0.00007
OFC (11)	dlPFC (8)	–0.070404	0.00001

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

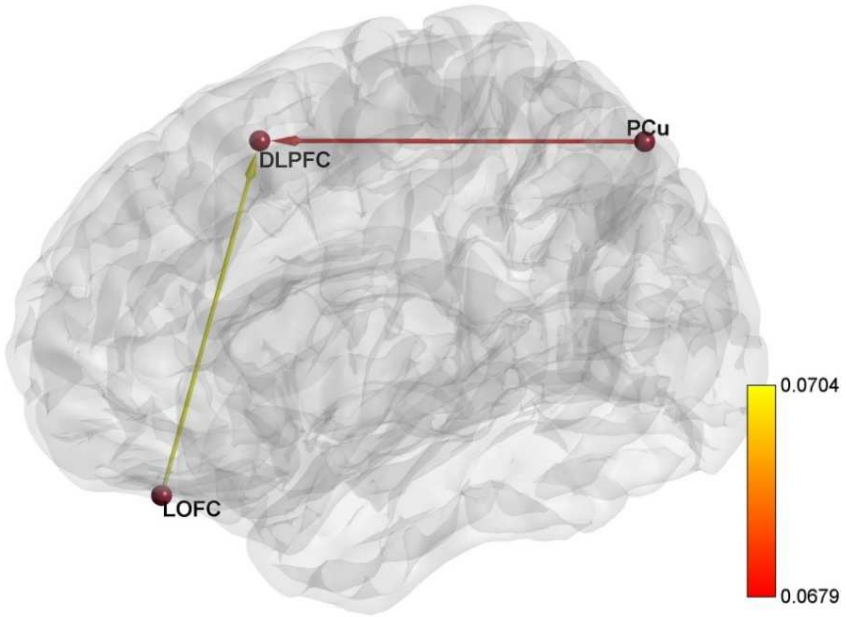


Figure 14. Effective Connectivity during Reciprocity (Stage 2, Partnership Maintenance).

Subsection 3.4.4 Connectivity between Brains

Stage 1. The precuneus and lOFC (trustee) were causally linked to the dmPFC (trustor). The precuneus was also linked to the PCC (trustor) (**Figure 15, Table 17**).

Table 17 Effective Connectivity between the Trustor and Trustee (Stage 1, Partnership Building)

ROI (BA)		Connectivity	
Trustee	**Trustor**	**Stage 1**	**P**
Reciprocity to Trust			
Precuneus (7)	dmPFC (9)	–0.102570	0.00370
Precuneus (7)	PCC (23)	–0.099779	0.00830
OFC (11)	dmPFC (9)	–0.105010	0.00830

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

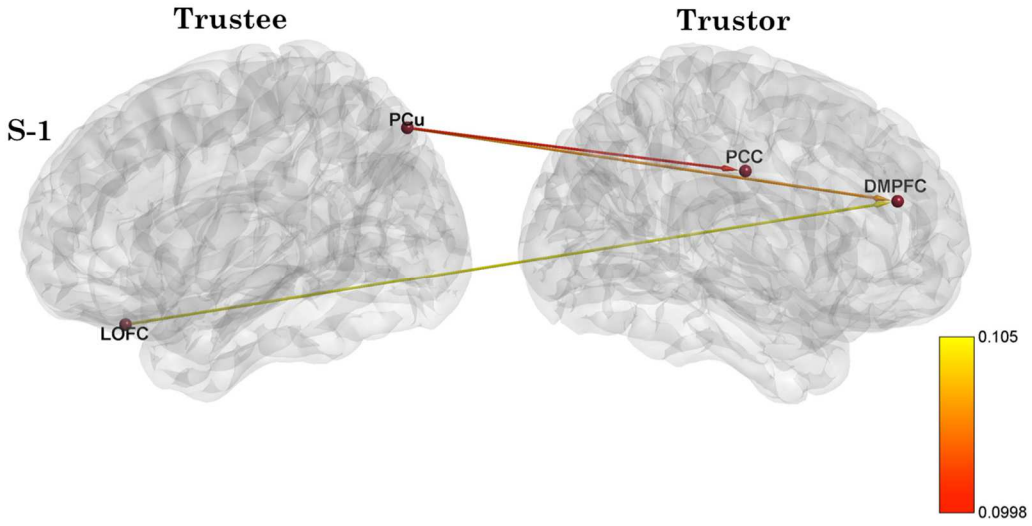


Figure 15. Effective Connectivity between the Trustor and Trustee (Stage 1, Partnership Building).

Stage 2. The topology linked three regions (i.e., precuneus, lOFC and dlPFC) in the trustee brain with the dmPFC of the trustor and two of the regions i.e., precuneus and lOFC, with the PCC of the trustor (**Figure 16, Table 18**).

Table 18 Effective Connectivity between the Trustor and Trustee (Stage 2, Partnership Maintenance)

ROI (BA)		Connectivity	
Trustee	**Trustor**	**Stage 2**	**P**
Reciprocity to Trust			
Precuneus (7)	dmPFC (9)	–0.099815	0.00026
Precuneus (7)	PCC (23)	–0.095545	0.00039
OFC (11)	dmPFC (9)	–0.099500	0.00400
OFC (11)	PCC (23)	–0.102060	0.00510
dlPFC (8)	dmPFC (9)	–0.102700	0.01910

t-test, [Human > Control], q(FDR) = 0.05 corrected for multiple comparisons.

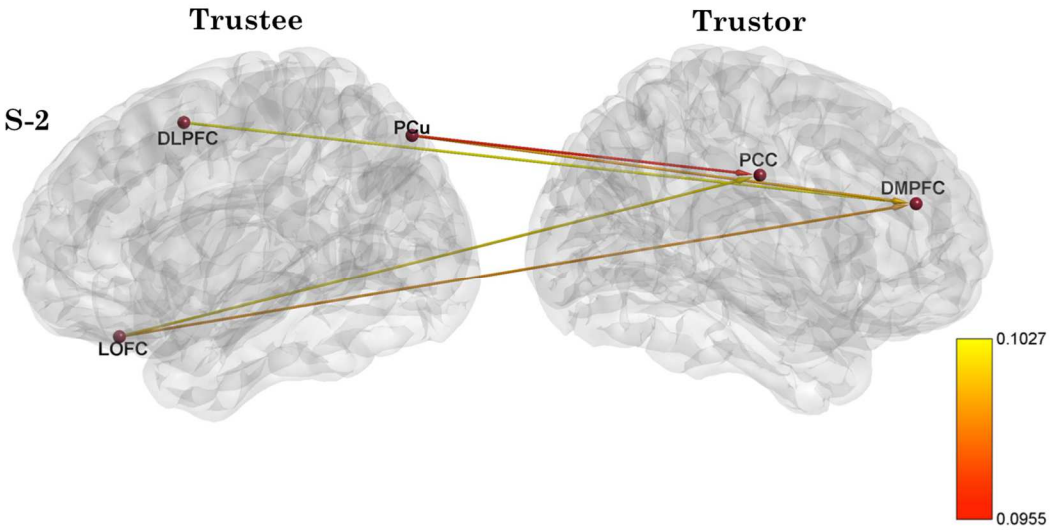


Figure 16. Effective Connectivity between the Trustor and Trustee (Stage 2, Partnership Maintenance).

Stage 1 vs. Stage 2. On the trustor's side, dmPFC and PCC were the only two regions involved in any communication in any of the stages. For the trustee, dlPFC was involved in Stage 2, but not in Stage 1, while dmPFC was absent in both. Connections of

the precuneus and lOFC (trustee) to the dmPFC (trustor) and of the precuneus (trustee) to the PCC (trustor) persisted across the stages. The connectivity strength varied cross-stage. In Stage 1, the strongest was lOFC (trustee) to dmPFC (trustor) connection, while the precuneus (trustee) to PCC (trustor) was the weakest. In Stage 2, the strongest was the dlPFC (trustee) to dmPFC (trustor) connection, while the precuneus (trustee) to PCC (trustor) was the weakest.

Section 3.5 Discussion

Subsection 3.5.1 Approach

If a generalized neurocognitive model of trust is to be built, the principles of its construction have to agree with the existing evidence on functional components of the model. One of the main functional components of trust is decision-making. Yet, the initiative to match the trust model developed in this dissertation against the existing neurobiological models of decision-making was met with a controversy within the domain of decision-making research itself. The research has produced an overall decision-making schema that involves two putatively related components. They include a neural circuitry of "choice" and "valuation", fundamental to any decision-making behavior. However, the two neurobiological models were produced by two compartmentalized lines of research and over the course of several years. As a result, to date, no apparent functional and neuronal links between the two systems have been empirically established [25]. This is despite the fact, that a wealth of functional evidence implicates the brain's hidden processes in the ability to aggregate subjective values for use in choice [19,182]. The exploratory analysis

in Chapter 3 set out to provide evidence assisting in closing this gap in the knowledge of decision-making in general and in trust decision-making in particular.

To achieve this goal, the analysis had to apply a combination of methods. First, for a connectivity analysis to be valid, the data must be collected on both sides of the interactive task. This was achieved by selecting a dataset from a hyperscan-fMRI study and by an extensive search of neurobiological literature for preliminary evidence. Based on the evidence, it was conjectured that a group of regions collectively known as "social brain" (mentalizing) regions possess the structural and functional properties necessary to provide the links between the two decision-making systems. Accordingly, the analysis schema would have to be designed in such a way as first, to functionally test the assumption that all three component systems would indeed engage for trust and second, while testing, to control for basic mechanisms of reward perception and choice. First step in that direction was made in Chapter 2, where the traits were manipulated using IG to extract the functional maps of reward, choice and mentalizing, specifically for trust and reciprocity. The next step in the analysis was to implement a schema as to allow the control for the effects of "choice" and "valuation" and subsequently to allow exploring the "middle-man" assumption. There were two known approaches to solving the problem: a hypothesis-driven approach, which would require an experiment for testing validity of a hypothesis and a data-driven approach, which would require a researcher to examine the existing data structure for possible effects for the task.

This dissertation has adopted the latter approach to examine the source data. But first, a functional whole-brain analysis of the existing dataset was conducted to tailor the

data to the needs of this analysis. While controlling for the confounding effects of basic reward-perception and choice behavior, the analysis found the effect of mentalizing on trust in the putative "social brain" ROI. This discovery motivated the choice of this dataset for the analysis in the pursuit of anatomical and functional evidence of the "missing" functional links between the reward and choice systems in the brain. Based on the source study evidence of strong presence of social brain in trust, it was conjectured that the mentalizing networks might serve as functional links between reward and choice.

Subsection 3.5.2 Effective Connectivity

The ROI time series were subsequently fit into the MVGC model to ascertain effective connectivity between the parietal and prefrontal regions – key to value-choice neuronal interactions. As a result, strong frontoparietal effective connectivity pathways for both trust and reciprocity were revealed. Furthermore, for trust, the frontoparietal dmPFC-PCC pathway provided a bidirectional link. For reciprocity, the link was unidirectional and localized more laterally i.e., precuneus-dlPFC. Both networks formed a family of connections around some center of connectivity. For trust, the network formed around two central regions, dmPFC and PCC. For reciprocity, the network formed a topology of a star with only one center, dlPFC. A functional role implied by the dmPFC-PCC pathway could be defined as "Network Bridge". By definition a "network bridge" is understood as a provider of 2-way communication exchange between autonomous and functionally significant segments of a network [183]. On these grounds, the bridge is in pole position at center of communications between the reward and choice segments of the network [19,184].

Subsection 3.5.3 Connectivity during Trust

The importance of the dmPFC-PCC connectivity is supported by the evidence implicating these regions in tracking subjective value of delayed monetary rewards [182] and in representing subjective value under conditions of ambiguity [185]. Ambiguity and delay are embedded in the IG reward structure, which is described by an implicit delay-magnitude trade-off. The trade-off is an ambiguous intertemporal choice between a larger but postponed and an immediate but smaller reward. This evidence indicates that dmPFC and PCC are not only anatomically connected to both the reward-processing and action-selection systems but also play a significant functional role in those systems.

For PCC, the evidence points at a role of a highly heterogeneous functional and anatomical hub, which links together a number of networks "that are functionally distinct but that require coordinated changes in activity to allow for efficient cognitive function" [186-188]. Of a particular interest is one of the PCC many links – a structural association with the parietal area LIP – key player in the putative "winner take all" frontoparietal model of action selection. According to this model, a decision is made based on a highest intrinsic subjective value [184,189].

The dmPFC is similarly known to form alliances with many other segments of the decision-making infrastructure such as vmPFC (valuation), lateral parietal cortex (action choice) and PCC (mentalizing) [190]. Numerous studies of mentalizing and empathy have emphasized the convergence underlying perspective taking to dmPFC [191,192]. In an experiment that may elucidate a possible role of dmPFC in the IG decision-making, a striking similarity between the dmPFC activation dynamic and the reward-learning dopaminergic

activity was observed [193]. Distinctively however, the dmPFC is engaged for social, but not the sensory context and might be able for example, to predict at once a reward and the valence of advice used for locating the reward. The activity in dmPFC has also been found to correlate with reward prediction error (PE), but only when the PE is linked to a trustworthy advice rather than to the reward's scalar value.

The TP connectivity findings in this dissertation converge with the earlier data relating together TP, mPFC and PCC [194]. Together, these regions comprise a mentalizing network involved in decoding information associated with personal memories in the context of interpersonal interactions [195,196]. Specifically, TP has been linked to one's ability to interpret social cues (others' actions) and to relate the social signals and intrinsic emotional states according to a putative "social script" [197]. In the IG task, the mentalizing triad may thus provide another crucial link between the "action choice" and "valuation" circuits, which can be established by means of mapping intentions to actions in the context of social norms (social script).

Less clear is the meaning of the dmPFC–hippocampus connection, as there is no direct evidence of the structural connectivity between the two regions. The apparent causal covariance between the regions is likely mediated by a third party structural coupling with the medial temporal lobe and the third party could be PCC [198-201]. The dmPFC-hippocampus connectivity pattern points at a critical utility of episodic social memories in the dmPFC global function of acquiring social knowledge about partner's intentions.

The results for hypothalamus would also agree with the previous findings confirming that the entire "valuation cascade" is strongly modulated by the hypothalamus –

the source of "trust hormone" oxytocin. Oxytocin is used by the brain as a means of controlling predispositions to trust in a social context dependent manner [202].

Subsection 3.5.4 Connectivity during Reciprocity

The reciprocity network analysis arrived at a rather different connectivity outcome than has the trust network analysis by identifying a different pattern of "social brain" [203,204], whereby the dlPFC was a site of critical input from the dmPFC for valuing social cooperation and self-control of impulsivity [205,206]. Further, the anatomical overlap of the dlPFC cluster with the frontal eye fields area in this analysis, points at a role in guiding saccade movements during choice [25].

The precuneus forming anatomical connections with dlPFC and area LIP [207,208] has been linked to first person perspective taking [209]. In terms of the IG paradigm, that might imply a role in evaluating partner's intentions in relationship to the player's own intentions. Together, the connectivity patterns of the precuneus-dlPFC bridge suggest a cognitive link between the parieto-centric "action selection" and the prefronto-centric "valuation" circuits during reciprocity.

The dlPFC connectivity with lOFC might reflect player's motivation for cooperative conduct given the design of the IG reward structure. In IG, the trustee is tempted to defect in favor of a higher payback, but is held back by the concern for social norms and by avoidance of betrayal-induced guilt [80]. Based on the neuroimaging evidence of the lOFC role in encoding for negative reinforcement by punishment or by omission of reward [210], its recruitment may indicate a concern for trustor's punishment, which can mo-

tivate the trustee's norm-compliance behavior even further. In this regard, a surprising result of the analysis is the absence of insula implicitly "averse" activity in the ROI for either reciprocity or trust, which is in contrast to the results generated in Chapter 2. This outcome can be explained by Chapter 3 selective focus on the multi-round IG and the regions that are associated with mutually benevolent and cooperative decision-making (i.e., trust and reciprocity, but not defection or betrayal).

Subsection 3.5.5 Connectivity during Reciprocity to Trust

The importance of the dmPFC-PCC functional tandem in trustor's decision-making was emphasized further by the results of the between-brain connectivity analysis, where these regions were the only ones ever involved in the interaction with the trustee. The dmPFC activity was predictive of the activity in all but one (dlPFC in Stage 1) of the three regions involved on the trustee side. The PCC activity was predictive of the activity in two of the trustee regions, precuneus (Stage-1 and 2) and lOFC (Stage 2).

The between-brain analysis provided evidence on what the within-brain analysis failed to show – a higher relative contribution of the dmPFC region in trust. For trustor, this evidence points at the greater importance of gathering social knowledge and accounting for the trustee's intentions when making trust decisions. The latter would engage the trustee's concern for positive reputation with the trustor (precuneus) mediated through a desire to not disappoint the trustor (lOFC). The relationship between the dmPFC (trustor) and dlPFC (trustee) in Stage 2 and lack thereof in Stage 1 highlight the differences between the partnership building stage, when the trustee could be more focused on building the reputation (precuneus and lOFC) sought by the trustor (dmPFC) and the relationship

maintenance stage, when the temptation to violate trust could be stronger (due to less need for the trustee to win reputation towards the end of the game) and would have to be controlled (dlPFC).

The PCC connections indicate both players going through the social adjustment process with each other. As noted before, the PCC has been implicated in policy switching while adjusting to changes in others' behavior as compared to expectations. The PCC (trustor) link to precuneus and lOFC (trustee) could very well model the relationship between the trustor's expectations about the trustee trustworthiness and his adjustment to the trustee's actual responses. Additionally, lack of the lOFC "reaction" to PCC in Stage 1 but not in Stage 2 provides converging evidence that the temptation and subsequent guilt feeling towards the trustor could be more prominent in Stage 2 than in Stage 1.

Subsection 3.5.6 Summary of Outcomes

The Chapter 3 analyses collectively revealed a series of important aspects of the economic interaction in IG. Based on this evidence, it can be concluded that trust and reciprocity are driven by different motives for favoring cooperation. For the trustee who has a stake in trustor's lasting cooperation, the greatest concern is to establish good reputation (precuneus) with the trustor. But equally important is the affective motive of avoiding guilt (lOFC), which is induced by the trustee temptation to defect and at the same time the desire to not let the trustor down. The trustor on the other hand, might be encouraged by the partner's adoption of mutually rewarding, albeit ambiguous, interpersonal decision-making strategy (dmPFC-PCC-TP-Hypothalamus).

CHAPTER 4. General Discussion

Section 4.1 Objectives

Subsection 4.1.1 Scope

The overall goal of the research presented in the current thesis is to examine dynamic characteristics of human interpersonal trust and reciprocity. The problem was approached by asking, what are the neural mechanisms that actually motivate people to delegate their goals and tasks to other individuals and to take into account others' motives in order to motivate their own behavior? In Chapter 1, these questions were framed in terms of beliefs and decisions of an agent who is pondering whether to turn to another individual in pursuit of his goals. In the subsequent chapters, this class of behaviors was examined in light of the data collected in earlier studies. Source studies were conducted in a natural social interaction of the IG task. IG was designed to study trust and reciprocity under the conditions, in which the interacting individuals are expected to demonstrate increased tendency for the alignment of their incentives to cooperate.

However, inherent in the problem domain of trust and reciprocity is complexity. Not only the majority of psychological causes of these behaviors are hidden mental states, but also, the course of a cooperative relationship is an intricate sequence of interdependent decision strategies and motivations that vary in time. To deal with the increasing complexity of the varied alternative neurocognitive paths, the necessity to inquire into domain-general properties of trust and to construct a comprehensive neurocognitive mod-

el of trust has emerged. These goals were achieved through the use of data-driven analyses, ALE and MVGC, in which the existing data structures were examined in pursuit of task relevant effects.

Subsection 4.1.2 Meta-Analysis

First, an "aerial", more global view on trust and reciprocity was adopted in Chapter 2 and then, a more focused, single-study approach followed in Chapter 3. The study in Chapter 2 provided a toolbox that housed all the important instruments to assist in generalizing the discrete phenomena across multiple fMRI studies. For example, the transition of trust attitudes throughout the interaction was traced from the offset of the task, when players' anxiety is presumably high, into the later stages, when the anxiety gives way to confidence.

Subsection 4.1.3 Connectivity Analysis

The combined contribution of the studies reported in Chapter 2 and Chapter 3 is a greater comprehension of the incentives to cooperate and their evolution throughout the exchange. These findings illustrate that cooperative incentives (as theory would predict) while being initially formed as unconditional beliefs, transform throughout the experience of an interaction into the attitudes that are conditional on the opponent's response history [2]. Specific contribution of Chapter 3 is a comprehensive resolution of a long-standing issue in neuroeconomics study of decision-making. When in decision-making research the evidence of strong impact of value signaling on choice-selection became ap-

parent [164], researchers have made several attempts to combine the two systems and to arrive at a generalized model that is biologically and psychologically valid.

So far, such attempts have not been very successful. No neurobiological evidence had been produced to date of how these two seemingly interdependent systems connect to each other. The precise mechanism of how the value signals are projected to the choice circuits of the brain has been poorly understood. This dissertation is approaching the problem from a novel perspective in the hope to contribute to a resolution. First, Chapter 2 provides confirmatory evidence of strong links between the two systems by correlating lateral parietal cortex to choice and vmPFC to reward-processing in trust. Then Chapter 3 reveals further computational evidence of effective connectivity between the two systems. In a series of exploratory studies leading to the proposed analysis it was initially theorized that the "missing" functional link could be found in the areas of the brain that are anatomically connected to both circuits. Subsequently, an equivalent of a "mediator" brain system with a mission to abstract and integrate reward values into behavioral responses was conceived [211]. Since then, strong neuroimaging evidence implicated a network of regions, including dmPFC, medial parietal and temporo-parietal cortices in a mental process commonly known as mentalizing. Based on the evidence, it was theorized further, that the mentalizing network is anatomically in pole position to mediate the neural communication between the anatomically distant frontoparietal (choice) and mPFC (valuation) networks and must be key to reinforcing trust beliefs nurtured by cooperative decision-making. In the next step, a study linking the putative "social brain" to trust was

identified and a dataset based on that study was generated subject to further examination in the present data-driven analysis.

Section 4.2 Methods

Subsection 4.2.1 ALE

Chapter 2 espoused IG – a useful economic game paradigm to examine trust from the economics perspective of utility in decision-making. The analysis in Chapter 2 also examined trust from a neurobiological perspective by looking at the neurocognitive mechanisms of disambiguating trust's cognitive dichotomies. The leading forces behind the emergence of trust and reciprocity beliefs and decisions (more broadly described as "cooperative phenotype" [118]) were analyzed in depth. The focus of the analysis was first, on the selection of fMRI-IG studies suitable for validating the neuroimaging evidence of human propensity for cooperation and second, on linking patterns of individual trust and trust repayment choices to the images of brain activity during the exchange. However, no single fMRI study would allow a sampling size large enough and a signal-to-noise ratio strong enough to ensure reliable generalizations of the behavioral and neuroimaging findings. To overcome the limitations, an ALE meta-analysis approach was implemented to inspect the evidence of regional activation in correlation with core cognitive components of trust in reciprocity. Comprised of three independent analyses comparing unconditional vs. conditional trust, expected (decision) vs. experienced (outcome) utility and trust vs. reciprocity, Chapter 2 generated the evidence of consistent activation maxima for trust across a selection of fMRI-IG studies.

Subsection 4.2.2 Multivariate Granger Causality

Chapter 3 embarked on a powerful method of analyzing neuroimaging data, namely the MVGC analysis, to see whether converging evidence in favor of the new conjectures could be provided. This allowed a novel approach as no prior studies of IG have addressed either within-brain or between-brain effective connectivity during trust and reciprocity in IG. MVGC is a data-driven method and unlike the standard hypothesis-driven data analysis methods does not require an accurate estimate of the relationship between the fMRI signal and performance of the experimental task. MVGC provides a complimentary approach by testing a hypothesis about the time course of activation in an experimental condition vs. control condition. What makes this method relevant to the analysis of interaction in economic exchange like IG is its ability to estimate the influence i.e., effective connectivity of neuronal populations on each other not only within, but also between the interacting brains. The outcome of the ECA in Chapter 3 was the topology of task-related effective connectivity pathways among the active regions.

Subsection 4.2.3 Hyperscan-fMRI

The source study for Chapter 3 applied a paradigm in which multi-round IG was played twice in a row with the same partner. The two runs represented two hypothetical stages in the formation of trust relationship, "partnership building" and "partnership maintenance". Importantly, the neural imaging was carried out by simultaneously scanning (i.e., "hyperscanning") the brains of both partners who were alternating their roles in the exchange. The combined methodological approach allowed for truly interactive exploration of brain activity in two cognitive dimensions, intertermporal and interactional.

Section 4.3 Outcomes

Subsection 4.3.1 ALE

The results of Chapter 2 demonstrated significant differences in regional activation between one-shot and multi-round IG. The unconditional trust (one-shot) was linked to activation maxima in cingulo-opercular network and hippocampus. The most robust maxima were found in anterior insula – an indicator of strong effects of ambiguity aversion on mental state of the trustor. In contrast, the conditional trust was strongly affected by how the opponent's reputation emerged from the course (decision → outcome) of the game. The rationale behind using the Chapter 2 findings as a baseline for Chapter 3 was to expand on the knowledge of trust dynamics and to analyze the potential mapping between the tendencies in the behavior and brain activity underlying its cognition. The results of Chapter 2 put the research inquiry in a new place, allowing an important new question: could this kind of understanding of the global network of trust be used to make progress in understanding the dynamics of trust?

Subsection 4.3.2 Connectivity for dmPFC

The results of Chapter 3 contribute to such understanding. Connectivity analysis in Chapter 3 demonstrated that multiple brain areas, found to be part of the trust decision-making process, showed a strong tendency to interconnect and produce compound connectivity patterns. Indeed, in the trust network the analysis revealed strong bidirectional connectivity and therefore strong causal relationship between the dmPFC and medial parietal cortex. Both the trust-mediated network and reciprocity-mediated network involved

dmPFC and medial parietal areas. Intriguingly however, the dmPFC activation, while so prominent in Chapter 3 has only shown consistency for reciprocity but not for trust in Chapter 2 study. Also, dmPFC and medial parietal cortex were directly connected for trust, but not for reciprocity where the two were connected through a third-party region, namely dlPFC.

These results may have several implications. According to earlier studies, the dmPFC pattern of activation was shown to be predicated on the already known dopaminergic activity during reward learning, but it manifests itself exclusively for socially salient information such as that pertaining to theory of mind [193]. Individual connectivity studies of the rostral cluster of dmPFC, matching the one identified in Chapter 3, agree with the notion that dmPFC has a role in episodic social memory and mental scene construction implicit in this region's connectivity to the PCC and hippocampus [190]. The dmPFC structural connectivity pattern is known to be distinct from that of the adjacent vmPFC. The vmPFC is weakly interconnected with motor areas and while in the right place to compute stimulus values, is not in the position to directly influence decisions. In contrast, the dmPFC is heavily interconnected with both SMA and the vmPFC involved in valuation [212,213]. The dmPFC neurobiological evidence is conducive with the socio-cognitive theory claiming that trust is a relationship between the agents who assume "intentional stance" [2,214]. Based on the overwhelming evidence supporting the notion of a value-to-choice functional connection, the Chapter 3 outcomes for the dmPFC imply a role for this region in the neural communication between the medial frontal (valuation) and parietal (choice) decision-making circuits.

Subsection 4.3.3 Connectivity for PCC

As for the medial parietal side of the frontoparietal connection, neurobiological literature also implicates medial parietal areas (PCC for trust and precuneus for reciprocity) in mentalizing but for a different constituent role. In contrast to the dmPFC focus on social experience, medial parietal cortex is focused on decoding personal experiences in social contexts. For example, neuroimaging evidence links PCC and the ability to integrate one's awareness of self-location and body ownership [215] and to modulate changes in subjective motivational state by reacting strongly to salient but not neutral rewards [216]. The PCC has dense structural connections to many other brain networks, suggesting a role as a cortical hub with its primary function to perform integration of distributed neural communications [187]. Both, PCC and precuneus are highly heterogeneous in structure [208,217] and facilitate "transitional" connectivity and functional coordination of the adjacent networks representing dissociable cognitive domains [218]. Such pattern of structural connectivity may reflect a role of a comparator for the medial parietal hub regions, which integrate and also relay the outcomes of value computations from the vmPFC to intraparietal regions to guide choices.

Subsection 4.3.4 dmPFC-PCC Bridge

Chapter 3 study has produced an important and novel finding in relationship to the dmPFC connectivity with PCC. This connectivity pathway interpreted as a "communication bridge" between the circuits of value and choice is strategically situated in the brain to compare stimulus values (e.g., monetary options in the task) against action costs (e.g., reward delay or betrayal of trust) and to integrate the outcomes of the comparison

into the process of computing optimal courses of action [219]. Chapter 3 provides evidence of the dmPFC serving as the crucial sought-after functional link – a neural system that might play a crucial role in mediating the valuation and choice brain systems during decision-making in trust.

According to the results, the dmPFC-PCC network bridge was also linked with the activity in TP – "a transitional region where many different cortical regions meet" [220]. The connectivity pattern of TP identified in the Chapter 3 study is in many respects in line with the structural connectivity pattern of the anterior temporal lobe, which receives anatomical projections from OFC, inferior frontal, perirhinal and insular cortices. Anatomically segregated sensory inputs (e.g., ventral visual stream and OFC) converge in TP to integrate perceptual input into the ability to recognize, infer and respond to a host of social signals. TP has also been implicated in a large number of mentalizing tasks and appears to be associated with socially relevant episodic memory encoded in the adjacent hippocampus and amygdala [197]. Neuroimaging findings implicating TP in theory of mind tasks may reflect the linkage of human ability to recognize social cues and self-regarding interpretations and reactions [221]. Selectively however, TP responds to only those social stimuli that feature a "story", a narrative or in psychological terms, "social script" [222]. From the TP evidence, it appears that the IG players learn from their opponent's past responses and put their views of the opponent in line with a cognitively developed and constantly updated interaction scenario. These outcomes thus point to the role of TP in keeping track of an event sequence during social interaction, which is done according to expectations the participants might have about the future of the exchange.

Subsection 4.3.5 Connectivity for Reciprocity

Reciprocity connectivity in Chapter 3 formed a pattern considerably different from that of trust. The center of the topology is localized to dlPFC, which receives projections from lOFC and the mentalizing regions i.e., dmPFC and precuneus. The distinct reciprocity pattern might indicate that, in spite of the apparent behavioral similarities, players' inferences about each other's intentions could be driven by quite dissimilar forces. While the trustor is busy betting on the trustee desire to cooperate, the trustee, when trusted, is awarded with more money than the trustor and could be caught in two minds between the temptation to defect and avoidance of feeling guilty if defected. The latter is evidenced by activation in lOFC and precuneus for reciprocity. The lOFC region has been previously linked to aversion reinforcement by punishment or guilt [210]. Precuneus on the other hand, has been implicated in modeling one's own behavior based on social norms and reputation with others [209]. The center of the network, dlPFC, has been previously implicated in controlling impulsive behaviors and in time discounting of short-term rewards in favor of a long term goal [206] – a solid candidate for a position in mediating conflicts caused by the trustee decision-making dilemma.

Subsection 4.3.6 Summary of Outcomes

In summary, the study of predisposition to trust characterized by ambiguity over potential violation of trust (Chapter 2) points at the anterior insula as a critical input to mPFC for negative subjective valuation and avoidance of ambiguity. The study of interactional trust, characterized by conditional i.e., reputation-based beliefs (Chapter 2 and Chapter 3), reveals a complex pattern of encoding for reward in vmPFC, ventral striatum

and PCC (Chapter 2). Mentalizing is likely to engage dmPFC, TP and precuneus (Chapter 3). IPL, dlPFC, SMA and dorsal striatum are recruited for behavioral response selection (Chapter 2). The study in Chapter 2 also implicates vmPFC and NAc in expected utility (decision phase). Lateral OFC and dorsal striatum are the likely neural correlates of experienced utility during the outcome (feedback) phase. The results of the reciprocity study in Chapter 3 point at lOFC, dmPFC and precuneus as providers of social knowledge inputs for the dlPFC. The dlPFC has a role in gating the inputs to control and produce socially appropriate behavior. Thus, what emerges is a fairly complex network of brain areas that collectively construct subjective value, social knowledge and action selection signals to guide cooperation-driven, inter-personal and inter-temporal choice.

Section 4.4 Implications and Future Directions

Subsection 4.4.1 Conclusions

The most significant findings of this research can be summarized as follows. First, the cognitive model of trust was defined to a greater degree of abstraction and clarity. Second, the model revealed a number of new generalizable patterns of brain activity and connectivity that support trust. Third, it showed strong evidence in support of a third-party link between the anatomically distant neural circuits underlying the functions of reward mapping and choice decision-making in trust. The identified fronto-parietal neural connectivity might prove critical to integrating reward values in choice and should be examined further. Having inherited what was originally a compartmentalized theory of value and choice, consisting of a mechanistic account of stimulus disambiguation on the one

hand, and an anatomically unsubstantiated account of the utility's role in choice on the other, this dissertation developed a unified theory of a domain-general brain system underlying value-guided choice in trust. Fourth, this dissertation provided an improved functional dissociation between unconditional and conditional trust, between phases of learning trust and between trust and reciprocity by revealing some commonly shared as well as dissimilar functional patterns such as the anatomically dissociable brain systems engaged for ambiguity resolution and theory of mind.

Subsection 4.4.2 Outlook

An improved conceptual understanding of trust introduced in this dissertation provides a natural platform and guide to future research. To achieve a higher level of utility the new studies could contribute to this line of research by addressing the measurement of alternative forms of trust, such as trust in information (e.g. trust in advice or testimony) and comparing those forms to the one addressed in this dissertation i.e., trust in actions. Or they could improve on the existing measures of trust that have received less attention in trust literature such as deficiency of trust in certain psychiatric disorders (e.g., generalized social anxiety disorder or bipolar personality disorder).

Alternatively, there is still room for improvement on the characteristics of the proposed model itself. A future meta-analysis could enhance the validity of the results reported here by including a comparison of IG to some other less-studied paradigms of trust, such as communication game. A potential effective connectivity analysis, on the other hand, could benefit from an enhanced efficiency of the existing fMRI studies or from the inclusion of a greater number of hyperscan-fMRI studies of trust should those

studies become more readily available in the future. The network model could be expanded further by conducting a connectivity analysis of unconditional trust in addition and in contrast to the connectivity findings for conditional trust provided in this dissertation. Collectively, such a thourough and comprehensive approach to trust experimentation could serve to support the existing neurocognitive models of trust and enhance their validity by demonstrating the causal role of functional brain activation and connectivity in cognitive modulation of human-to-human trust and reciprocity.

APPENDIX A. META-ANALYSIS STUDY SELECTION

Aimone, J. A., Houser, D., & Weber, B. (2014). Neural signatures of betrayal aversion: an fMRI study of trust. Proceedings of the Royal Society B: Biological Sciences, 281(1782), 20132127.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. Neuron, 64(5), 756-770.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron, 58(4), 639-650.

Bereczkei, T., Deak, A., Papp, P., Perlaki, G., & Orsi, G. (2013). Neural correlates of Machiavellian strategies in a social dilemma task. Brain Cogn, 82(1), 108-116.

Bereczkei, T., Papp, P., Kincses, P., Bodrogi, B., Perlaki, G., Orsi, G., & Deak, A. (2015). The neural basis of the Machiavellians' decision making in fair and unfair situations. Brain Cogn, 98, 53-64.

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron, 70(3), 560-572.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci, 8(11), 1611-1618.

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. Front Neurosci, 6, 148.

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. Journal of Neuroscience, 35(21), 8170-8180.

Fett, A. K., Gromann, P. M., Giampietro, V., Shergill, S. S., & Krabbendam, L. (2013). Default distrust? An fMRI investigation of the neural development of trust and cooperation. Soc Cogn Affect Neurosci, 9(4), 395-402.

Fouragnan, E. (2013). The Neural Computation of Trust and Reputation. University of Trento.

Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., & Coricelli, G. (2013). Reputational Priors Magnify Striatal Responses to Violations of Trust. Journal of Neuroscience, 33(8), 3602-3611.

Gromann, P. M., Shergill, S. S., de Haan, L., Meewis, D. G., Fett, A. K., Korver-Nieberg, N., & Krabbendam, L. (2014). Reduced brain reward response during cooperation in first-degree relatives of patients with psychosis: an fMRI study. Psychol Med, 44(16), 3445-3454.

Kang, Y., Williams, L. E., Clark, M. S., Gray, J. R., & Bargh, J. A. (2011). Physical temperature effects on trust behavior: the role of insula. Soc Cogn Affect Neurosci, 6(4), 507-515.

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(5718), 78-83.

Krueger, F., Grafman, J., & McCabe, K. (2008). Neural correlates of economic game playing. Philos Trans R Soc Lond B Biol Sci, 363(1511), 3859-3874.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., . . . Grafman, J. (2007). Neural correlates of trust. Proc Natl Acad Sci U S A, 104(50), 20084-20089.

Lauharatanahirun, N., Christopoulos, G. I., & King-Casas, B. (2012). Neural computations underlying social risk sensitivity. Front Hum Neurosci, 6, 213.

Li, J., Xiao, E., Houser, D., & Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. Proc Natl Acad Sci U S A, 106(39), 16835-16840.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci U S A, 98(20), 11832-11835.

Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. Journal of Neuroscience, 35(8), 3412-3419.

Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. Proc Natl Acad Sci U S A, 107(29), 13099-13104.

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2014). Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. Journal of Management Information Systems, 30(4), 83-114.

Smith-Collins, A. P. R., Fiorentini, C., Kessler, E., Boyd, H., Roberts, F., & Skuse, D. H. (2013). Specific neural correlates of successful learning and adaptation during social exchanges. Soc Cogn Affect Neurosci, 8(8), 887-896.

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), 984-989.

Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. Philos Trans R Soc Lond B Biol Sci, 367(1589), 744-753.

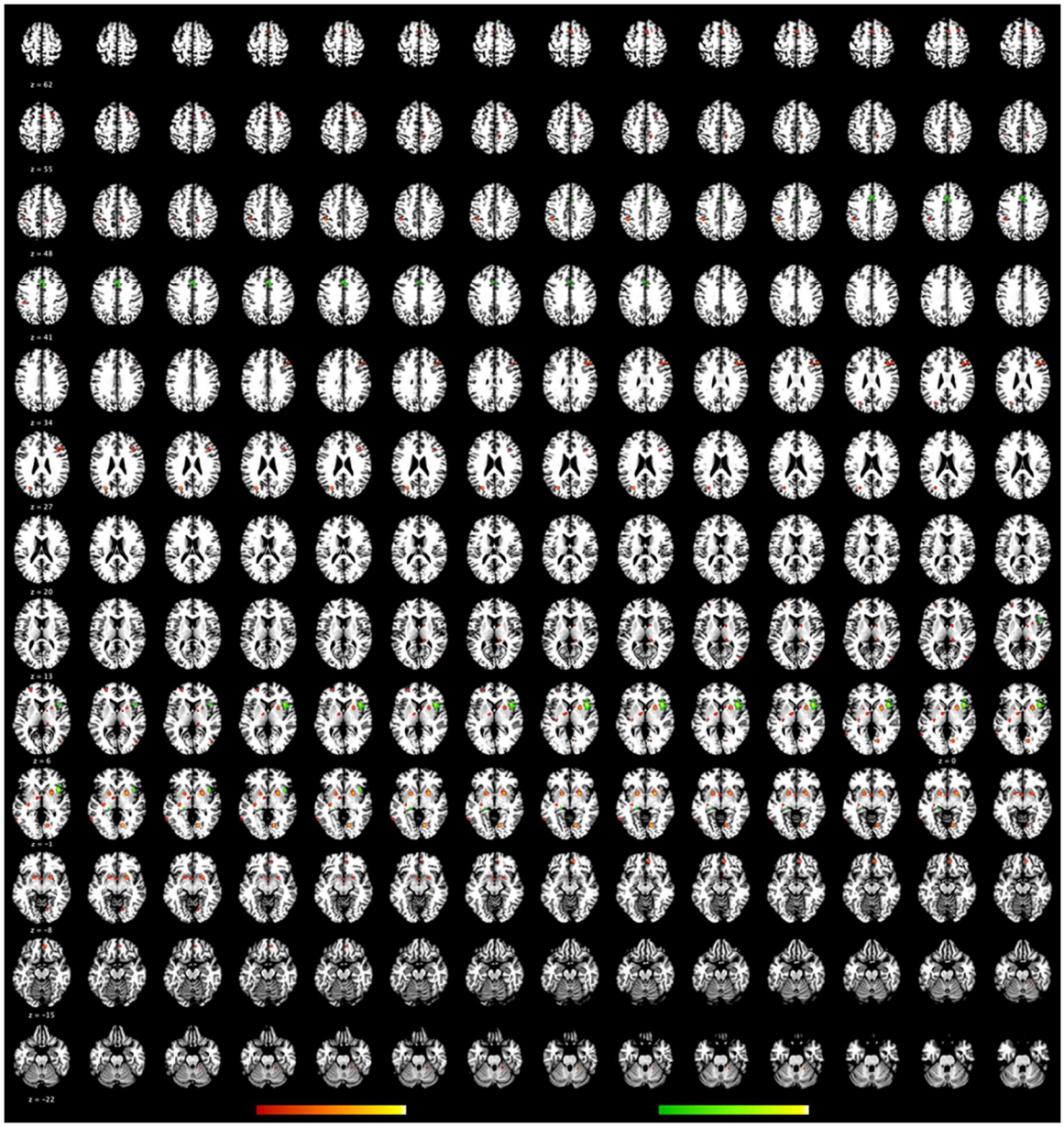
van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., & Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. Soc Cogn Affect Neurosci, 4(3), 294-304.

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., & Crone, E. A. (2011). Changing brains, changing perspectives: the neurocognitive development of reciprocity. Psychol Sci, 22(1), 60-70.

Wardle, M. C., Fitzgerald, D. A., Angstadt, M., Sripada, C. S., McCabe, K., & Luan Phan, K. (2013). The caudate signals bad reputation during trust decisions. PLoS One, 8(6), e68884.

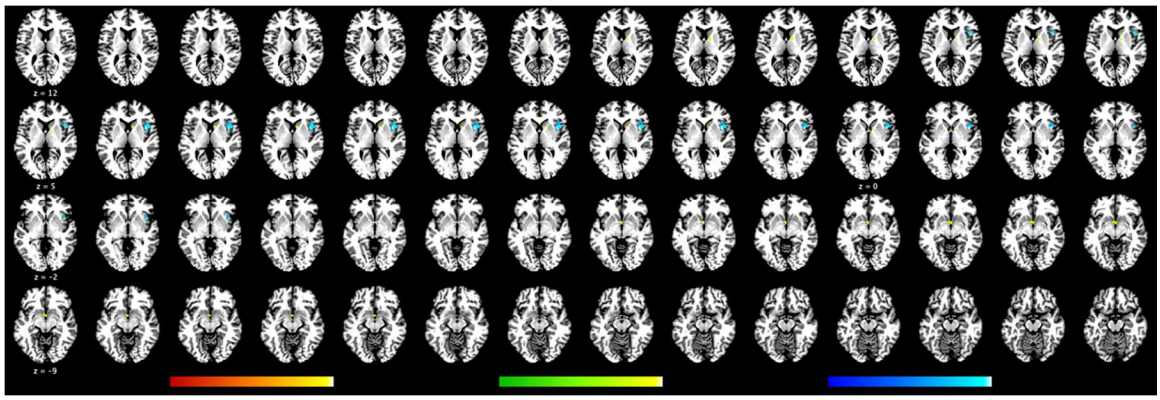
Xiang, T., Ray, D., Lohrenz, T., Dayan, P., & Montague, P. R. (2012). Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. PLoS Comput Biol, 8(12), e1002841.

APPENDIX B. SUPPLEMENTAL INFO

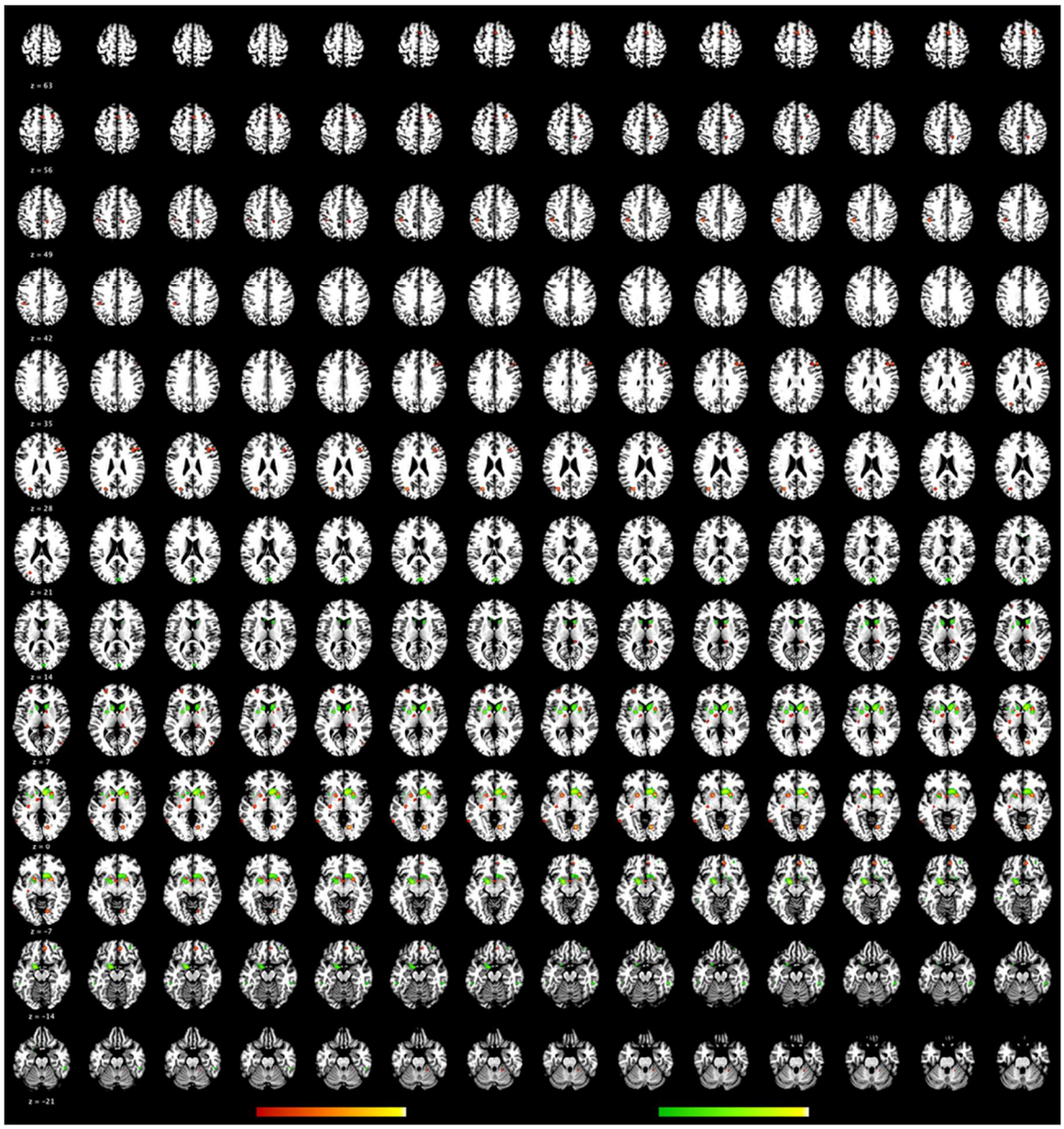


Supplemental Figure 1 Results of ALE singe-dataset analysis of trust (one-shot IG vs. multi-round)
Multi-Round Networks (Red): mOFC (vmPFC, NAc, HCd, GP, VA), ventral vis. stream (V2, MT/IT, Pul) & frontoparietal (rlPFC, dlPFC, PCC, PCu, IPL, PMC [not shown], SMA, BCd). One-Shot (Green): cingulo-

opercular network (AI, dACC, fO) & Hip. Random-effects analysis, 5,000 permutations, ALE values, q(FDR)<0.05, min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.

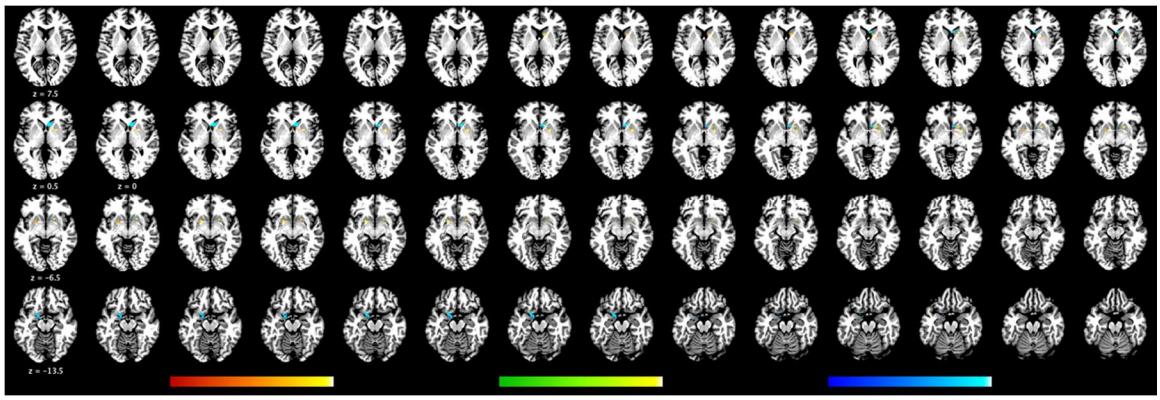


Supplemental Figure 2 Results ALE image-contrast analysis of trust (one-shot vs. multi-round IG)
Multi-Round > One-Round (Green): peaks in BCd, HCd & NAc. One-Round > Multi-Round (Blue): peak in AI. Random-effects analysis; 5,000 permutations; z-scores; q(FDR) < 0.05; min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.



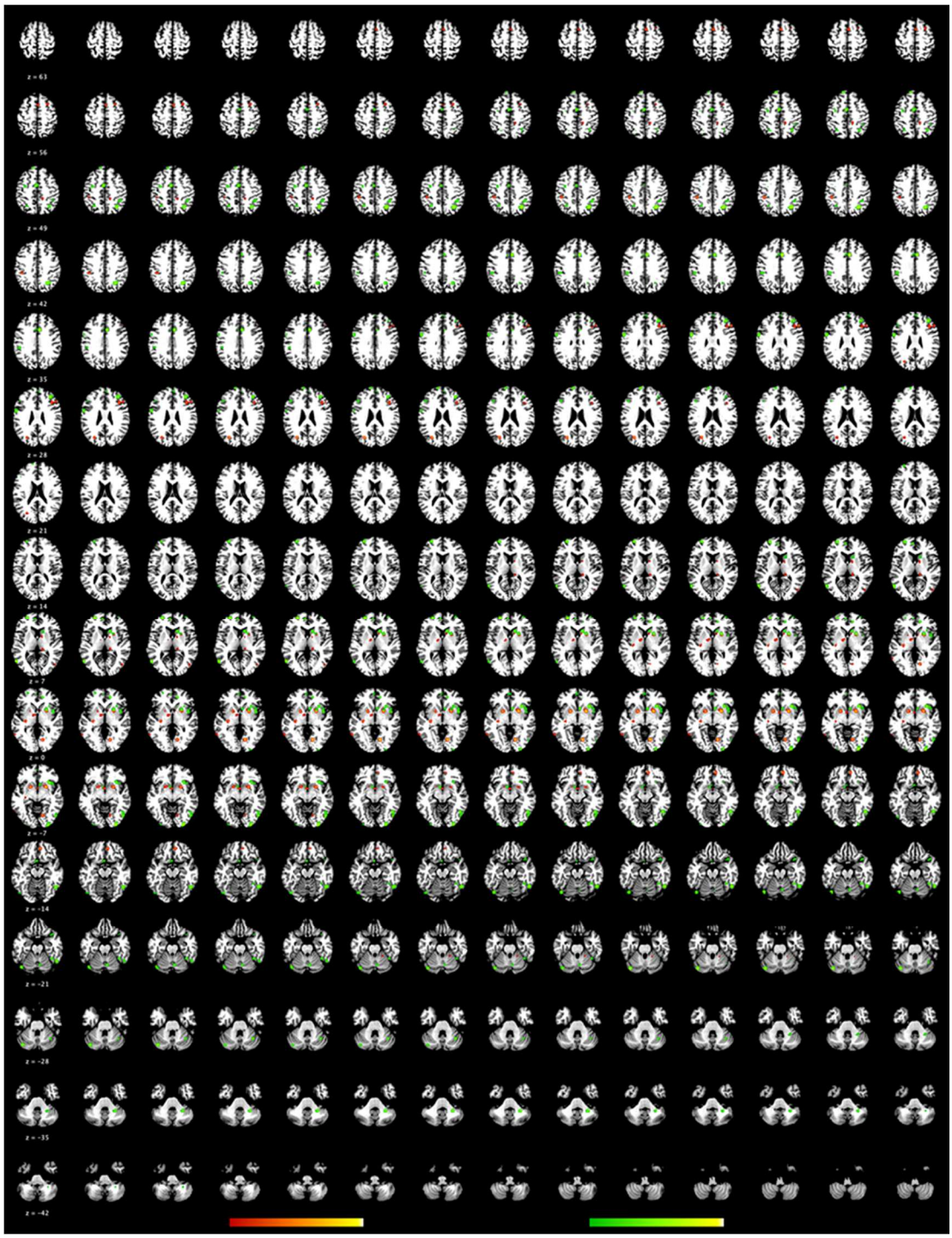
Supplemental Figure 3 Results of ALE singe-dataset analysis of trust decision & outcome phases (multi-round IG)

Outcome Phase (Green): peaks in lOFC, BCd, HCd, NAc, Putamen, GP, CB, AI, IT & MT. Decision phase (Red). Random-effects ALE analysis, 5,000 permutations, ALE values, q(FDR) < 0.05, min. threshold of 100 mm^3. Image overlay on a normalized brain template using Mango.

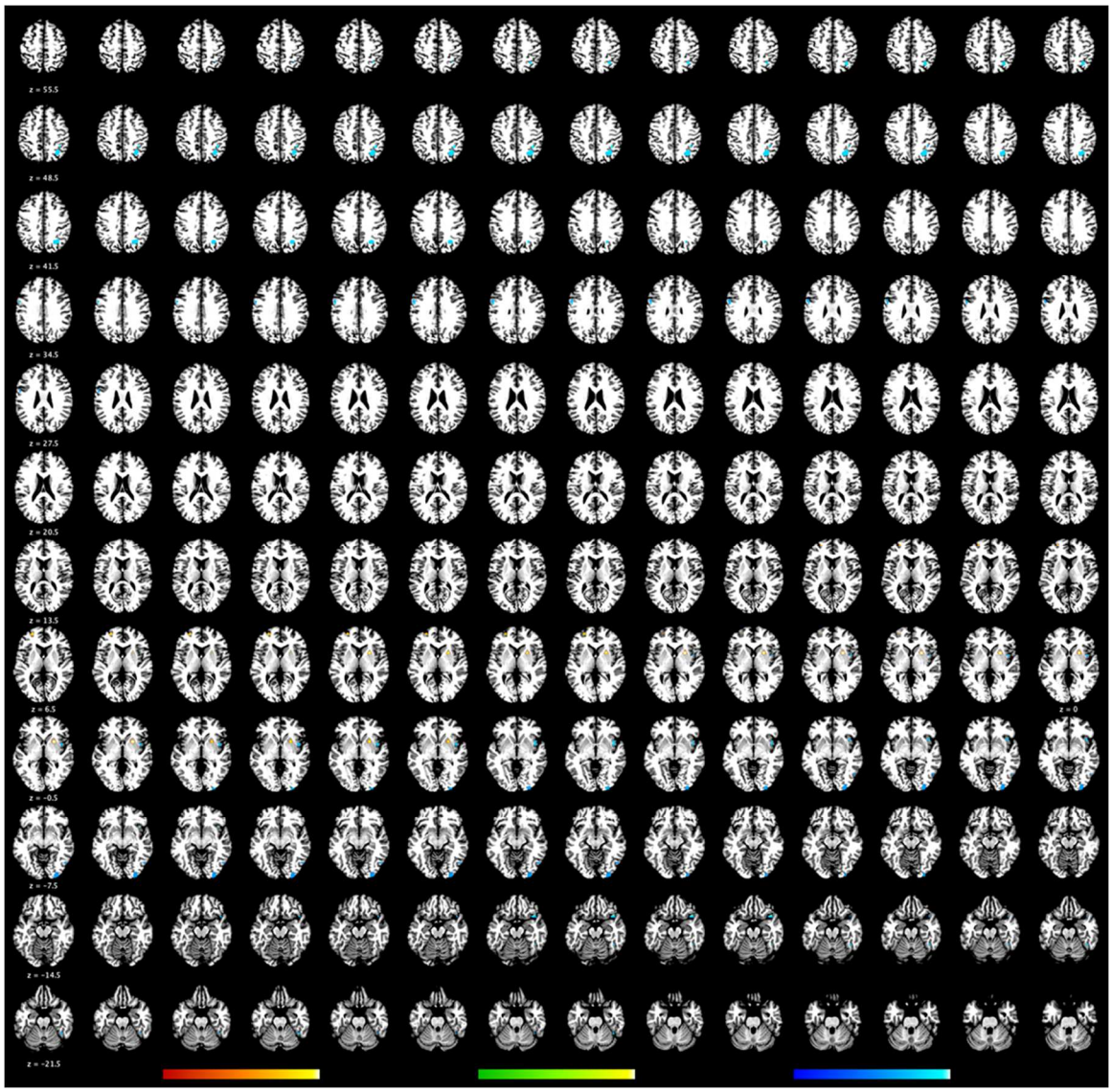


Supplemental Figure 4 Results of ALE image-contrast analysis of trust decision vs. outcome phases (multi-round IG)

Decision ∩ Outcome (Red): peaks in right BCd & HCd + bilateral putamen. Outcome > Decision (Blue): peaks in left HCd & Putamen. Random-effects analysis; 5,000 permutations; z-scores; *q*(FDR) < 0.05; min. threshold of 100 mm^3; Image is overlaid on a normalized brain template using Mango.



Supplemental Figure 5 Results of ALE singe-dataset analysis of trust vs. reciprocity (multi-round IG) Reciprocity Networks (Green): frontoparietal (rlPFC, dlPFC, dmPFC, vlPFC, PCC, SPL/IPL, BCd & HCd) & mPFC (vmPFC, dACC, NAc, GP & VA) + AI, IT/MT, putamen & CB. Trust (Red). Random-effects analysis; 5,000 permutations; ALE values; q(FDR) < 0.05; min. threshold of 100 mm^3; Image is overlaid on a normalized brain template using Mango.



Supplemental Figure 6 ALE image-contrast analysis of trust vs. reciprocity (multi-round IG)
Trust ∩ Reciprocity (Red): peaks in rlPFC, NAc (not shown) & putamen. Reciprocity > Trust (Blue): peaks in AI, vlPFC, SPL, IPL, IT & V2. Random-effects analysis: 5,000 permutations; *z*-score values; *q*(FDR) < 0.05; min. threshold of 100 mm^3. Image is overlaid on a normalized brain template using Mango.

REFERENCES

1 Gambetta, D. *Trust: Making and Breaking Cooperative Relations*. Vol. 52 (Blackwell, 1988).
2 Castelfranchi, C. & Falcone, R. *Trust theory: A socio-cognitive and computational model*. Vol. 18 (John Wiley & Sons, 2010).
3 Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. F. Not so Different after All: A Cross-Discipline View of Trust. *The Academy of Management Review* **Vol. 23,**, 393-404 (1998).
4 Mayer, R. C., Davis, J. H. & Schoorman, F. D. An Integrative Model of Organizational Trust. *The Academy of Management Review* **20**, 709-734 (1995).
5 Lewis, J. D. & Weigert, A. Trust as a social reality. *Social forces* **63**, 967-985 (1985).
6 Phan, K. L., Sripada, C. S., Angstadt, M. & McCabe, K. Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13099-13104 (2010).
7 McAllister, D. J. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* **38**, 24-59 (1995).
8 Gallotti, M. & Michael, J. *Perspectives on Social Ontology and Social Cognition*. Vol. 4 (Springer, 2014).
9 Gold, J. & Heekeren, H. Neural mechanisms for perceptual decision making. *Neuroeconomics: decision making and the brain, Ed* **2**, 355-372 (2013).
10 Klein, S. A. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics* **63**, 1421-1455 (2001).
11 Strasburger, H. Converting between measures of slope of the psychometric function. *Perception & Psychophysics* **63**, 1348-1355 (2001).
12 Heekeren, H. R., Marrett, S., Bandettini, P. & Ungerleider, L. A general mechanism for perceptual decision-making in the human brain. *Nature* **431**, 859-862 (2004).
13 Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299-308 (2002).
14 Ding, L. & Gold, J. I. Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, bhr178 (2011).
15 Ding, L. & Gold, J. I. Caudate encodes multiple computations for perceptual decisions. *The Journal of Neuroscience* **30**, 15747-15759 (2010).
16 Ding, L. & Gold, J. I. Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron* **75**, 865-874 (2012).
17 Summerfield, C. & Tsetsos, K. Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Frontiers in neuroscience* **6** (2012).
18 Levy, D. J. & Glimcher, P. W. The root of all value: a neural common currency for choice. *Curr Opin Neurobiol* **22**, 1027-1038 (2012).
19 Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* **10**, 1625-1633 (2007).
20 Padoa□Schioppa, C. & Cai, X. The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Annals of the New York Academy of Sciences* **1239**, 130-137 (2011).
21 Levy, D. J. & Glimcher, P. W. Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *Journal of Neuroscience* **31**, 14693-14707 (2011).
22 Haruno, M. & Frith, C. D. Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature neuroscience* **13**, 160-161 (2010).
23 Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K. & Fehr, E. The neural circuitry of a broken promise. *Neuron* **64**, 756-770 (2009).
24 Delgado, M. R. Reward□related responses in the human striatum. *Annals of the New York Academy of Sciences* **1104**, 70-88 (2007).
25 Glimcher, P. W., Dorris, M. C. & Bayer, H. M. Physiological utility theory and the neuroeconomics of

choice. *Games and economic behavior* **52**, 213-256 (2005).
26 Ferrin, D. L. & Dirks, K. T. The use of rewards to increase and decrease trust: Mediating processes and differential effects. *Organization science* **14**, 18-31 (2003).
27 Aimone, J. A., Houser, D. & Weber, B. Neural signatures of betrayal aversion: an fMRI study of trust. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132127 (2014).
28 Kim, Y. A. & Song, H. S. Strategies for predicting local trust based on trust propagation in social networks. *Knowledge-Based Systems* **24**, 1360-1371 (2011).
29 Von Neumann, J. & Morgenstern, O. Game theory and economic behavior. *Princeton, Princeton University* (1944).
30 Fehr, E. & Camerer, C. F. Social neuroeconomics: the neural circuitry of social preferences. *Trends in cognitive sciences* **11**, 419-427 (2007).
31 Hoffman, E., McCabe, K. A. & Smith, V. L. Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry* **36**, 335-352 (1998).
32 Camerer, C. & Weigelt, K. Experimental tests of a sequential equilibrium reputation model. *Econometrica: Journal of the Econometric Society*, 1-36 (1988).
33 Berg, J., Dickhaut, J. & McCabe, K. Trust, Reciprocity &Social History. *Games and Economic Behavior* **10**, 122-142 (1995).
34 Vogt, W. P., Gardner, D. C. & Haeffele, L. M. *When to use what research design*. (Guilford Press, 2012).
35 Fehr, E. On the Economics and Biology of Trust. *Journal of the European Economic Association* **7**, 235-266 (2009).
36 Tzieropoulos, H. The Trust Game in neuroscience: A short review. *Soc Neurosci* **8**, 407-416 (2013).
37 Riedl, R. & Javor, A. The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics* **5**, 63-91 (2012).
38 Price, C. J., Devlin, J. T., Moore, C. J., Morton, C. & Laird, A. R. Meta-analyses of object naming: effect of baseline. *Human brain mapping* **25**, 70-82 (2005).
39 Stark, C. E. & Squire, L. R. When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences* **98**, 12760-12766 (2001).
40 Wager, T. D., Lindquist, M. & Kaplan, L. Meta-analysis of functional neuroimaging data: current and future directions. *Soc Cogn Affect Neurosci* **2**, 150-158 (2007).
41 Feredoes, E. & Postle, B. R. Localization of load sensitivity of working memory storage: quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis. *Neuroimage* **35**, 881-903 (2007).
42 Raemaekers, M. *et al.* Test–retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage* **36**, 532-542 (2007).
43 Cabeza, R. & Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neuroscience* **12**, 1-47 (2000).
44 Fox, P. T., Parsons, L. M. & Lancaster, J. L. Beyond the single study: function/location metanalysis in cognitive neuroimaging. *Current Opinion in Neurobiology* **8**, 178-187 (1998).
45 Friston, K. J. Functional and effective connectivity: a review. *Brain connectivity* **1**, 13-36 (2011).
46 Montague, P. R. *et al.* Hyperscanning: Simultaneous fMRI during linked social interactions. *Neuroimage* **16**, 1159-1164 (2002).
47 Krueger, F. *et al.* Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20084-20089 (2007).
48 Laird, A. R. *et al.* ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Human brain mapping* **25**, 155-164 (2005).
49 Eickhoff, S. B., Amunts, K., Mohlberg, H. & Zilles, K. The human parietal operculum. II. Stereotaxic maps and correlation with functional imaging results. *Cereb Cortex* **16**, 268-279 (2006).
50 Havlicek, M., Friston, K. J., Jan, J., Brazdil, M. & Calhoun, V. D. Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *Neuroimage* **56**, 2109-2128 (2011).
51 Granger, C. W. Time series analysis, cointegration, and applications. *American Economic Review*, 421-425 (2004).
52 Kaminski, M., Ding, M., Truccolo, W. A. & Bressler, S. L. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics* **85**, 145-157 (2001).
53 King-Casas, B. *et al.* The rupture and repair of cooperation in borderline personality disorder. *Science* **321**, 806-810 (2008).

54 Miller, G. J. *Managerial Dilemmas: The Political Economy of Hierarchies*. 254 (Cambridge Univ. Press., 1992).
55 Knack, S. & Keefer, P. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics* **112**, 1251-1288 (1997).
56 Kramer, R. M. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology* **50**, 569-598 (1999).
57 Simpson, J. A. Psychological foundations of trust. *Current directions in psychological science* **16**, 264-268 (2007).
58 Camerer, C. F. Behavioural studies of strategic thinking in games. *Trends in Cognitive Sciences* **7**, 225-231 (2003).
59 King-Casas, B. *et al.* Getting to know you: reputation and trust in a two-person economic exchange. *Science* **308**, 78-83 (2005).
60 McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11832-11835 (2001).
61 Burnham, T., McCabe, K. & Smith, V. L. Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization* **43**, 57-73 (2000).
62 Xiang, T., Ray, D., Lohrenz, T., Dayan, P. & Montague, P. R. Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Comput Biol* **8**, e1002841 (2012).
63 McCabe, K. A., Rigdon, M. L. & Smith, V. L. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization* **52**, 267-275 (2003).
64 Johnson, N. D. & Mislin, A. A. Trust games: A meta-analysis. *Journal of Economic Psychology* **32**, 865-889 (2011).
65 Fareri, D. S., Chang, L. J. & Delgado, M. R. Effects of direct social experience on trust decisions and neural reward circuitry. *Front Neurosci* **6**, 148 (2012).
66 Fouragnan, E. *et al.* Reputational Priors Magnify Striatal Responses to Violations of Trust. *Journal of Neuroscience* **33**, 3602-3611 (2013).
67 Stanley, D. A. *et al.* Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philos Trans R Soc Lond B Biol Sci* **367**, 744-753 (2012).
68 Gromann, P. M. *et al.* Reduced brain reward response during cooperation in first-degree relatives of patients with psychosis: an fMRI study. *Psychol Med* **44**, 3445-3454 (2014).
69 Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D. & Heekeren, H. R. Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. *Journal of Management Information Systems* **30**, 83-114 (2014).
70 Li, J., Xiao, E., Houser, D. & Montague, P. R. Neural responses to sanction threats in two-party economic exchange. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16835-16840 (2009).
71 Fareri, D. S., Chang, L. J. & Delgado, M. R. Computational substrates of social value in interpersonal collaboration. *J Neurosci* **35**, 8170-8180 (2015).
72 Krueger, F., Moll, J., Zahn, R., Heinecke, A. & Grafman, J. Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cerebral Cortex* **17**, 2346-2353 (2007).
73 Raichle, M. E. & Snyder, A. Z. A default mode of brain function: a brief history of an evolving idea. *Neuroimage* **37**, 1083-1090 (2007).
74 Mitchell, J. P., Macrae, C. N. & Banaji, M. R. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* **50**, 655-663 (2006).
75 Fink, G. R. *et al.* Cerebral representation of one's own past: neural networks involved in autobiographical memory. *The Journal of Neuroscience* **16**, 4275-4282 (1996).
76 Northoff, G. *et al.* Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage* **31**, 440-457 (2006).
77 Buckner, R. L. & Carroll, D. C. Self-projection and the brain. *Trends in cognitive sciences* **11**, 49-57 (2007).
78 Lauharatanahirun, N., Christopoulos, G. I. & King-Casas, B. Neural computations underlying social risk sensitivity. *Front Hum Neurosci* **6**, 213 (2012).
79 Kang, Y., Williams, L. E., Clark, M. S., Gray, J. R. & Bargh, J. A. Physical temperature effects on trust behavior: the role of insula. *Soc Cogn Affect Neurosci* **6**, 507-515 (2011).

80 Chang, L. J., Smith, A., Dufwenberg, M. & Sanfey, A. G. Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron* **70**, 560-572 (2011).
81 Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychological review* **108**, 624 (2001).
82 Bush, G., Luu, P. & Posner, M. I. Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences* **4**, 215-222 (2000).
83 Brett, M. The MNI brain and the Talairach atlas. *MRC Cognition and Brain Sciences Unit* (1999).
84 Eickhoff, S. B. *et al.* Coordinate□based activation likelihood estimation meta□analysis of neuroimaging data: A random□effects approach based on empirical estimates of spatial uncertainty. *Human brain mapping* **30**, 2907-2926 (2009).
85 Laird, A. R. *et al.* A comparison of label-based review and ALE meta-analysis in the Stroop task. *Human brain mapping* **25**, 6-21 (2005).
86 Turkeltaub, P. E., Eden, G. F., Jones, K. M. & Zeffiro, T. A. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* **16**, 765-780 (2002).
87 Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870-878 (2002).
88 Rorden, C., Karnath, H. O. & Bonilha, L. Improving lesion-symptom mapping. *Journal of cognitive neuroscience* **19**, 1081-1088 (2007).
89 Sripada, C. S., Angstadt, M., Liberzon, I., McCabe, K. & Phan, K. L. Aberrant reward center response to partner reputation during a social exchange game in generalized social phobia. *Depress Anxiety* **30**, 353-361 (2013).
90 Gromann, P. M. *et al.* Trust versus paranoia: abnormal response to social reward in psychotic illness. *Brain* **136**, 1968-1975 (2013).
91 Sripada, C. S. *et al.* Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport* **20**, 984-989 (2009).
92 Yu, S., Beugelsdijk, S. & de Haan, J. Trade, trust and the rule of law. *European Journal of Political Economy* **37**, 102-115 (2015).
93 Declerck, C. H., Boone, C. & Emonds, G. When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and cognition* **81**, 95-117 (2013).
94 Engemann, D. A., Bzdok, D., Eickhoff, S. B., Vogeley, K. & Schilbach, L. Games people play—toward an enactive view of cooperation in social neuroscience. *Frontiers in human neuroscience* **6**, 148 (2012).
95 Caceda, R., James, G. A., Gutman, D. A. & Kilts, C. D. Organization of intrinsic functional brain connectivity predicts decisions to reciprocate social behavior. *Behav Brain Res* **292**, 478-483 (2015).
96 Cisler, J. M. *et al.* Brain and behavioral evidence for altered social learning mechanisms among women with assault-related posttraumatic stress disorder. *J Psychiatr Res* **63**, 75-83 (2015).
97 Kliemann, D., Young, L., Scholz, J. & Saxe, R. The influence of prior record on moral judgment. *Neuropsychologia* **46**, 2949-2957 (2008).
98 Lohrenz, T., McCabe, K., Camerer, C. F. & Montague, P. R. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences* **104**, 9493-9498 (2007).
99 Chiu, P. H. *et al.* Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron* **57**, 463-473 (2008).
100 Shao, R., Zhang, H.-j. & Lee, T. M. The neural basis of social risky decision making in females with major depressive disorder. *Neuropsychologia* **67**, 100-110 (2015).
101 Schilke, O., Reimann, M. & Cook, K. S. Effect of relationship experience on trust recovery following a breach. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15236-15241 (2013).
102 Krueger, F., Grafman, J. & McCabe, K. Neural correlates of economic game playing. *Philos Trans R Soc Lond B Biol Sci* **363**, 3859-3874 (2008).
103 Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U. & Fehr, E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* **58**, 639-650 (2008).
104 Bereczkei, T., Deak, A., Papp, P., Perlaki, G. & Orsi, G. Neural correlates of Machiavellian strategies in a social dilemma task. *Brain Cogn* **82**, 108-116 (2013).
105 Fett, A. K., Gromann, P. M., Giampietro, V., Shergill, S. S. & Krabbendam, L. Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Soc Cogn Affect Neurosci* **9**, 395-402 (2013).
106 Fouragnan, E. *The Neural Computation of Trust and Reputation*, University of Trento, (2013).

107 Wardle, M. C. *et al.* The caudate signals bad reputation during trust decisions. *PloS one* **8**, e68884 (2013).
108 Delgado, M. R., Frank, R. H. & Phelps, E. A. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* **8**, 1611-1618 (2005).
109 Smith-Collins, A. P. R. *et al.* Specific neural correlates of successful learning and adaptation during social exchanges. *Social Cognitive and Affective Neuroscience* **8**, 887-896 (2013).
110 van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. & Crone, E. A. What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Soc Cogn Affect Neurosci* **4**, 294-304 (2009).
111 van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. & Crone, E. A. Changing brains, changing perspectives: the neurocognitive development of reciprocity. *Psychol Sci* **22**, 60-70 (2011).
112 Nihonsugi, T., Ihara, A. & Haruno, M. Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *J Neurosci* **35**, 3412-3419 (2015).
113 Bereczkei, T. *et al.* The neural basis of the Machiavellians' decision making in fair and unfair situations. *Brain Cogn* **98**, 53-64 (2015).
114 Palminteri, S., Khamassi, M., Joffily, M. & Coricelli, G. Contextual modulation of value signals in reward and punishment learning. *Nature communications* **6** (2015).
115 Dosenbach, N. U. *et al.* Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences* **104**, 11073-11078 (2007).
116 Taylor, K. S., Seminowicz, D. A. & Davis, K. D. Two systems of resting state connectivity between the insula and cingulate cortex. *Human brain mapping* **30**, 2731-2745 (2009).
117 Hahn, T. *et al.* Reliance on functional resting-state network for stable task control predicts behavioral tendency for cooperation. *Neuroimage* **118**, 231-236 (2015).
118 Peysakhovich, A., Nowak, M. A. & Rand, D. G. Humans display a 'cooperative phenotype'that is domain general and temporally stable. *Nature communications* **5** (2014).
119 Frank, M. J. & Badre, D. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cerebral cortex* **22**, 509-526 (2012).
120 Braver, T. S. & Bongiolatti, S. R. The role of frontopolar cortex in subgoal processing during working memory. *Neuroimage* **15**, 523-536 (2002).
121 Halford, G. S., Wilson, W. H. & Phillips, S. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences* **21**, 803-831 (1998).
122 Christoff, K. *et al.* Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* **14**, 1136-1149 (2001).
123 Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction.*, (MIT Press, 1998).
124 Niv, Y. Reinforcement learning in the brain. *Journal of Mathematical Psychology* **53**, 139-154 (2009).
125 Koziol, L. F. & Budding, D. E. *Subcortical structures and cognition: Implications for neuropsychological assessment*. (Springer Science & Business Media, 2009).
126 Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience* **33**, 269-298 (2010).
127 Aragona, B. J. *et al.* Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. *Nature neuroscience* **9**, 133-139 (2006).
128 Botvinick, M. M., Huffstetler, S. & McGuire, J. T. Effort discounting in human nucleus accumbens. *Cognitive Affective & Behavioral Neuroscience* **9**, 16-27 (2009).
129 Carter, R. M., MacInnes, J. J., Huettel, S. A. & Adcock, R. A. Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. *Frontiers in behavioral neuroscience* **3** (2009).
130 Lewis, A. H., Porcelli, A. J. & Delgado, M. R. The effects of acute stress exposure on striatal activity during Pavlovian conditioning with monetary gains and losses. *Frontiers in behavioral neuroscience* **8** (2014).
131 Prévost, C., Liljeholm, M., Tyszka, J. M. & O'Doherty, J. P. Neural correlates of specific and general Pavlovian-to-Instrumental Transfer within human amygdalar subregions: a high-resolution fMRI study. *The Journal of Neuroscience* **32**, 8383-8390 (2012).
132 Meshi, D., Morawetz, C. & Heekeren, H. R. Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use. *Frontiers in human neuroscience* **7** (2013).
133 Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412-427 (2013).

134 Haruno, M. & Kawato, M. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks* **19**, 1242-1254 (2006).
135 Tunik, E., Rice, N. J., Hamilton, A. & Grafton, S. T. Beyond grasping: representation of action in human anterior intraparietal sulcus. *Neuroimage* **36**, T77-T86 (2007).
136 Becchio, C., Manera, V., Sartori, L., Cavallo, A. & Castiello, U. Grasping intentions: from thought experiments to empirical evidence. *Frontiers in human neuroscience* **6** (2012).
137 Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G. & Walter, H. Full Title:"Do you mean me? Communicative intentions recruit the mirror and the mentalizing system". *Social cognitive and affective neuroscience*, nst062 (2013).
138 Bonini, L. *et al.* Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. *Cerebral Cortex* **20**, 1372-1385 (2010).
139 Schilbach, L. *et al.* Toward a second-person neuroscience. *Behavioral and Brain Sciences* **36**, 393-414 (2013).
140 Rizzolatti, G. & Sinigaglia, C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews neuroscience* **11**, 264-274 (2010).
141 Van Overwalle, F. Social Cognition and the Brain: A Meta-Analysis. *Human brain mapping* **30**, 829-858 (2009).
142 Dezecache, G., Conty, L. & Grèzes, J. Social affordances: Is the mirror neuron system involved? *Behavioral and Brain Sciences* **36**, 417-418 (2013).
143 Dehaene, S., Piazza, M., Pinel, P. & Cohen, L. Three parietal circuits for number processing. *Cogn Neuropsychol* **20**, 487-506 (2003).
144 Krueger, F., Landgraf, S., van der Meer, E., Deshpande, G. & Hu, X. Effective connectivity of the multiplication network: a functional MRI and multivariate Granger Causality Mapping study. *Human brain mapping* **32**, 1419-1431 (2011).
145 Cohen Kadosh, R., Lammertyn, J. & Izard, V. Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Prog Neurobiol* **84**, 132-147 (2008).
146 Stanescu-Cosson, R. *et al.* Cerebral bases of calculation processes: Impact of number size on the cerebral circuits for exact and approximate calculation. *Brain* **123**, 2240-2255 (2000).
147 Chochon, F., Cohen, L., van de Moortele, P. F. & Dehaene, S. Differential contributions of the left and right inferior parietal lobules to number processing. *J Cogn Neurosci* **11**, 617-630 (1999).
148 Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D. & Dehaene, S. Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron* **33**, 475-487 (2002).
149 Menon, V., Rivera, S. M., White, C. D., Glover, G. H. & Reiss, A. L. Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* **12**, 357-365 (2000).
150 Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R. & Piefke, M. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Cognitive Neuroscience, Journal of* **19**, 1354-1372 (2007).
151 Andreoni, J. *Trust, Reciprocity, and Contract Enforcement: Experiments in Satisfaction Guaranteed*. (Social Systems Research Institute, University of Wisconsin, 2005).
152 Fehr, E. & Schmidt, K. M. Fairness and Incentives in a Multi□task Principal–Agent Model. *The Scandinavian Journal of Economics* **106**, 453-474 (2004).
153 Fehr, E. & Gintis, H. Human motivation and social cooperation: Experimental and analytical foundations. *Annu. Rev. Sociol.* **33**, 43-64 (2007).
154 Falk, A. & Fischbacher, U. A theory of reciprocity. *Games and economic behavior* **54**, 293-315 (2006).
155 McClintock, C. G. & Allison, S. T. Social value orientation and helping Behavior1. *Journal of Applied Social Psychology* **19**, 353-362 (1989).
156 Van Lange, P. A., De Bruin, E., Otten, W. & Joireman, J. A. Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *Journal of personality and social psychology* **73**, 733 (1997).
157 Dufwenberg, M. & Gneezy, U. Measuring beliefs in an experimental lost wallet game. *Games and economic Behavior* **30**, 163-182 (2000).
158 Bacharach, M., Guerra, G. & Zizzo, D. J. The self-fulfilling property of trust: An experimental study. *Theory and Decision* **63**, 349-388 (2007).
159 Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of*

neuroscience **24**, 167-202 (2001).
160 Aarts, E., Roelofs, A. & van Turennout, M. Attentional control of task and response in lateral and medial frontal cortex: Brain activity and reaction time distributions. *Neuropsychologia* **47**, 2089-2099 (2009).
161 Valentin, V. V., Dickinson, A. & O'Doherty, J. P. Determining the neural substrates of goal-directed learning in the human brain. *The Journal of neuroscience* **27**, 4019-4026 (2007).
162 Souza, M. J., Donohue, S. E. & Bunge, S. A. Controlled retrieval and selection of action-relevant knowledge mediated by partially overlapping regions in left ventrolateral prefrontal cortex. *Neuroimage* **46**, 299-307 (2009).
163 Wolfensteller, U. & Ruge, H. Frontostriatal mechanisms in instruction-based learning as a hallmark of flexible goal-directed behavior. *Frontiers in psychology* **3** (2012).
164 Glimcher, P. W. & Fehr, E. *Neuroeconomics: Decision making and the brain*. (Academic Press, 2013).
165 Handwerker, D. A., Ollinger, J. M. & D'Esposito, M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* **21**, 1639-1651 (2004).
166 Friston, K. J. *et al.* Event-related fMRI: characterizing differential responses. *Neuroimage* **7**, 30-40 (1998).
167 Deshpande, G., Sathian, K. & Hu, X. Effect of hemodynamic variability on Granger causality analysis of fMRI. *Neuroimage* **52**, 884-896 (2010).
168 Lam, E. Y. & Goodman, J. W. Iterative statistical approach to blind image deconvolution. *JOSA A* **17**, 1177-1184 (2000).
169 Friston, K. Hierarchical models in the brain. *PLoS Comput Biol* **4**, e1000211 (2008).
170 Arasaratnam, I. & Haykin, S. Cubature kalman filters. *Automatic Control, IEEE Transactions on* **54**, 1254-1269 (2009).
171 Rauch, H. E., Striebel, C. & Tung, F. Maximum likelihood estimates of linear dynamic systems. *AIAA journal* **3**, 1445-1450 (1965).
172 Havlicek, M., Jan, J., Brazdil, M. & Calhoun, V. D. Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *Neuroimage* **53**, 65-77 (2010).
173 Codd, E. F. A relational model of data for large shared data banks. *Communications of the ACM* **13**, 377-387 (1970).
174 Granger, C. Investigating Causal Relations by Econometric Models. *Econometrica* **37**, 424-438 (1969).
175 Roebroeck, A., Formisano, E. & Goebel, R. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* **25**, 230-242 (2005).
176 Glover, G. H. Deconvolution of impulse response in event-related bold fmri 1. *Neuroimage* **9**, 416-429 (1999).
177 Deshpande, G., LaConte, S., James, G. A., Peltier, S. & Hu, X. Multivariate Granger causality analysis of fMRI data. *Human brain mapping* **30**, 1361-1373 (2009).
178 Deshpande, G., Hu, X., Stilla, R. & Sathian, K. Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *Neuroimage* **40**, 1807-1814 (2008).
179 Sathian, K., Deshpande, G. & Stilla, R. Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. *The Journal of Neuroscience* **33**, 5387-5398 (2013).
180 Xia, M., Wang, J. & He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one* **8**, e68910 (2013).
181 Chiang, M. & Yang, M. in *Proc., Allerton Conf. on Comm., Control, and Computing.*
182 Kable, J. W. & Glimcher, P. W. An "As Soon As Possible" Effect in Human Intertemporal Decision Making: Behavioral Evidence and Neural Mechanisms. *Journal of Neurophysiology* **103**, 2513-2531 (2010).
183 Ball, E., Linge, N., Kummer, P. & Tasker, R. Local area network bridges. *Computer Communications* **11**, 115-117 (1988).
184 Dorris, M. C. & Glimcher, P. W. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* **44**, 365-378 (2004).
185 Levy, I., Snell, J., Nelson, A. J., Rustichini, A. & Glimcher, P. W. Neural representation of subjective value under risk and ambiguity. *Journal of neurophysiology* **103**, 1036-1047 (2010).
186 Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* **137**, 12-32 (2014).
187 Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS biology* **6**, e159 (2008).
188 Vogt, B. A., Vogt, L. & Laureys, S. Cytology and functionally correlated circuits of human posterior cingulate areas. *Neuroimage* **29**, 452-466 (2006).

189 Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y. & Platt, M. L. Posterior cingulate cortex: adapting behavior to a changing world. *Trends in cognitive sciences* **15**, 143-151 (2011).
190 Eickhoff, S. B., Laird, A. R., Fox, P. T., Bzdok, D. & Hensel, L. Functional segregation of the human dorsomedial prefrontal cortex. *Cerebral cortex*, bhu250 (2014).
191 Bzdok, D. *et al.* Segregation of the human medial prefrontal cortex in social cognition. *Front Hum Neurosci* **7**, 232 (2013).
192 Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* **7**, 268-277 (2006).
193 Behrens, T. E., Hunt, L. T. & Rushworth, M. F. The computation of social behavior. *science* **324**, 1160-1164 (2009).
194 Fan, L. *et al.* Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cerebral cortex* **24**, 3365-3378 (2014).
195 Fletcher, P. C. *et al.* Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* **57**, 109-128 (1995).
196 Frith, C. D. & Frith, U. The neural basis of mentalizing. *Neuron* **50**, 531-534 (2006).
197 Olson, I. R., Plotzker, A. & Ezzyat, Y. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* **130**, 1718-1731 (2007).
198 Morris, R. G. *et al.* Spatial working memory in Asperger's syndrome and in patients with focal frontal and temporal lobe lesions. *Brain and cognition* **41**, 9-26 (1999).
199 Maddock, R. J., Garrett, A. S. & Buonocore, M. H. Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience* **104**, 667-676 (2001).
200 Lavenex, P., Suzuki, W. A. & Amaral, D. G. Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *Journal of Comparative Neurology* **447**, 394-420 (2002).
201 Kobayashi, Y. & Amaral, D. G. Macaque monkey retrosplenial cortex: II. Cortical afferents. *Journal of Comparative Neurology* **466**, 48-79 (2003).
202 Krueger, F. *et al.* Oxytocin receptor genetic variation promotes human trust behavior. *Frontiers in Human Neuroscience* **6** (2012).
203 Petrides, M. & Pandya, D. Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience* **11**, 1011-1036 (1999).
204 Goulas, A., Uylings, H. B. & Stiers, P. Unravelling the intrinsic functional organization of the human lateral frontal cortex: a parcellation scheme based on resting state fMRI. *The Journal of Neuroscience* **32**, 10238-10252 (2012).
205 McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503-507 (2004).
206 McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. Time discounting for primary rewards. *The Journal of Neuroscience* **27**, 5796-5804 (2007).
207 Zhang, S. & Chiang-shan, R. L. Functional connectivity mapping of the human precuneus by resting state fMRI. *Neuroimage* **59**, 3548-3562 (2012).
208 Margulies, D. S. *et al.* Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of the National Academy of Sciences* **106**, 20069-20074 (2009).
209 Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564-583 (2006).
210 O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature neuroscience* **4**, 95-102 (2001).
211 Krueger, F., Barbey, A. K. & Grafman, J. The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences* **13**, 103-109 (2009).
212 Beckmann, M., Johansen-Berg, H. & Rushworth, M. F. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *The Journal of neuroscience* **29**, 1175-1190 (2009).
213 Picard, N. & Strick, P. L. Imaging the premotor areas. *Current opinion in neurobiology* **11**, 663-672 (2001).
214 Dennett, D. C. *The intentional stance*. (MIT press, 1989).
215 Guterstam, A., Björnsdotter, M., Gentile, G. & Ehrsson, H. H. Posterior cingulate cortex integrates the senses of self-location and body ownership. *Current Biology* **25**, 1416-1425 (2015).
216 Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C. & Jones-Gotman, M. Changes in brain activity related to eating chocolate. *Brain* **124**, 1720-1733 (2001).
217 Leech, R., Kamourieh, S., Beckmann, C. F. & Sharp, D. J. Fractionating the default mode network: distinct

contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *The Journal of Neuroscience* **31**, 3217-3224 (2011).
218 Vincent, J. L. *et al.* Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of neurophysiology* **96**, 3517-3531 (2006).
219 Platt, M. L. & Plassmann, H. in *Elsevier Inc.*
220 Olson, I. R., McCoy, D., Klobusicky, E. & Ross, L. A. Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social cognitive and affective neuroscience*, nss119 (2012).
221 Andrews-Hanna, J. R., Reidler, J. S., Huang, C. & Buckner, R. L. Evidence for the default network's role in spontaneous cognition. *Journal of neurophysiology* **104**, 322-335 (2010).
222 Gallagher, H. L. & Frith, C. D. Functional imaging of 'theory of mind'. *Trends Cogn Sci* **7**, 77-83 (2003).

BIOGRAPHY

Sergey V. Chernyak graduated from High School #30 (now #1276), Moscow, Russia. He received his Bachelor of Science from Gubkin State University, Moscow in 1984. He was employed as an Industrial Engineer in Moscow for four years and, after arriving to the United States, he worked as a Software Engineer for another 15 years. He received his Master of Arts in Psychology in 2011 and his Ph.D. in Neuroscience in 2016 from George Mason University.