Fall 2022

Physiological Synchrony Predicts Cooperative Behavior with High Stakes

Lena Arai
Claremont Graduate University

Physiological Synchrony Predicts Cooperative Behavior with High Stakes

By

Lena Arai

Claremont Graduate University
2022

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Lena Arai as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Economics and Information Systems & Technology.

Paul J. Zak, Chair

Claremont Graduate University

Professor of Economics

Jorge A. Barraza

University of Southern California

Professor of Psychology

George D. Montañez

Harvey Mudd College

Professor of Computer Science

Abstract

Physiological Synchrony Predicts Cooperative Behavior with High Stakes

By

Lena Arai

Claremont Graduate University: 2022

Can subconscious bodily responses explain our natural tendency to be trusting and trustworthy towards a stranger? I address this question by conducting, to my knowledge, the first study of physiological synchrony (PS) between pairs of partners playing the trust game face-to-face. Participants were given the choice to send $0, $40, $80, or $120 to their partner; these choices were categorized as showing no, low, medium, and high trust, respectively. Participants were endowed with a more considerable sum of money ($120) than many other trust games (Johnson & Mislin, 2011) to encourage participants to perceive their decisions to have significant consequences, i.e., for ecological validity. Most trust game experiments study college students (Johnson & Mislin, 2011); here participants were working-age adults between the ages of 25–50 from diverse cultural backgrounds. Before making their decisions, partners were given two minutes to interact and make promises to each other about their game decisions. Few studies on the trust game allow participant pairs to communicate face-to-face before making their decisions (Ben-Ner et al., 2011; Johnson & Mislin, 2011; Lev-On et al., 2010; Zak et al. 2022). PS between participants' skin conductance levels (SCLs) was measured during the interaction period and analyzed using two methods, intersubject correlation (ISC) and dynamic time warping (DTW). The DTW analyses revealed the no trust participants exhibited greater PS than low trust participants. DTW also indicated that high trust individuals exhibited greater PS than low and medium trust individuals, consistent with my expectation. The second mover ISC

analysis showed untrustworthy participants exhibited greater PS than trustworthy participants. These findings reveal that participants playing the trust game exhibit PS and engage in no trust, high trust, and untrustworthy behavior, indicating PS, trust, and trustworthy behavior are nonlinear. This is the first study of its kind to demonstrate that individuals display PS in the trust game.

Dedication

A mi Esposo, Mamá, Papá, y Abuelita

Acknowledgements

I would like to thank my advisor, Paul Zak, and my committee members, Jorge Barraza and George Montañez, for their expertise, guidance, and support.

I would also like to thank my husband, Kenley, and my mom for their unconditional support.

Table of Contents

List of Figures

I. Introduction

Psychologists and economists study trust because it is an integral part of society and our daily lives. Across all social and economic sciences, trust is one of the most extensively studied constructs because it influences interpersonal and group behavior more than any other single factor (Golembiewski & McConkie, 1975, as cited in Thielmann & Hilbig, 2015; Thielmann & Hilbig, 2015). Human nature is built on trust and reciprocity, both in interpersonal relationships and in economic and societal structures (Fareri, 2019). It is an important construct that underpins the success of businesses and economies, as well as a pillar of close relationships and social networks (Fareri, 2019). Nobel laureate Kenneth Arrow (1974) labeled trust "an important lubricant of a social system" (p. 23). It is essential to personal well-being and economic success: customers trust that purchased goods will work as promised; businesses trust their employees to do honest work; and investors trust corporations to report accurate quarterly numbers (Evans & Krueger, 2010). Several aspects of daily life even require trusting unknown individuals (Thielmann & Hilbig, 2015). Without trust, we could not survive.

Trust is a form of prosocial behavior — "any act designed to increase others' well-being, such as cooperating with, sharing resources with, and helping others (Tomasello, 2009)" (Zaki & Mitchell, 2013). The most widely cited definition of trust is from Mayer, Davis, and Schoorman (1995), who defined it as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712; Thielmann & Hilbig, 2015). This definition captures two necessary aspects of trust, vulnerability and positive expectation (Evans & Krueger, 2009; Thielmann & Hilbig, 2015). Thielmann and Hilbig (2015) sought to update this definition to incorporate the third necessary aspect of trust, risk and uncertainty, and defined trust as “a risky choice of making oneself dependent on the actions of another in a situation of uncertainty, based upon some expectation of whether the other will act in a benevolent fashion despite an opportunity to betray” (p. 251). The act of trusting another creates the possibility of mutual benefit, provided the other person cooperates, but exposes oneself to the risk of loss if they do not (Cox, 2004).

Trust is commonly studied within the context of social exchanges because they involve providing something of value to another with the expectation of future reciprocity. Research on social dilemmas has provided a 40-year-old approach to modeling cooperation that uses social dilemmas as a model for social exchange (Boone & Buck, 2003). Economists study trust in social exchanges and measure interpersonal trust using a set of economic games designed to measure trust and trustworthy behavior. The trust game is used to study the behavior in a controlled, experimental setting, with the dictator game serving as a control for other-regarding behavior, such as altruism.

To play the trust game participants are first randomly paired and assigned the role of first mover (FM) or second mover (SM). Both players are endowed with a small sum of money, usually $10, and given the opportunity to share their endowment with the other player. Participants are informed that any amount sent by the FM is tripled by the experimenter before being transferred to the SM. In the first round of game play, the FM can send some, all, or none of their $10 to their anonymously paired partner, which the experimenter triples. If the FM sends money, the second round provides the SM an opportunity to reciprocate and return some of the tripled amount to their partner. Given that the FM’s behavior in the game affects the SM’s payoff, it incorporates elements of trust as well as cooperation (Ben-Ner & Putterman, 2009; Burks, Carpenter, & Verhoogen, 2003; Cox, 2004; Evans, Athenstaedt, & Krueger, 2013; Holm

& Nystedt, 2008; Kamas & Preston, 2012; McEvily, Radzevick, & Weber, 2012; Vollan, 2011; Vyrastekova & Garikipati, 2005; Yamagishi et al., 2013).

The dictator game is a one-shot task similar to the first round of the trust game. Only one player, the dictator, is endowed with a sum of money; they can divide it between themselves and the other player, but they are not required to share it (Bolton, Katok, & Zwick, 1998; Fehr & Krajbich, 2014; Guala & Mittone, 2010). The game is designed to measure positive concern for the other player's material payoff unconditional of the other player's behavior, since the other player can take no actions (Fehr & Krajbich, 2014). When a FM in the trust game sends more than they do in the dictator game, one can be certain that the individual has deliberately exposed themselves to a risk of loss and thereby demonstrated trust (Cox, 2004).

The combination of economic games enables trusting behaviors to be measured and distinguished from risky behaviors that do not make an individual vulnerable. Together the games produce a measure of trust with construct validity (Cox, 2004). Trustworthiness is measured as the amount the SM returns to the FM, although an experiment design in which all subjects play the dictator game can also be used to derive a more robust measure of trustworthiness (Cox, Kerschbamer, & Neururer, 2016).

According to social exchange theory, individuals make decisions about their behavior based on their assessments of costs and benefits (Blau, 1986). In the trust game, if the FM believes the SM will be trustworthy and return more than the FM sent, they may take the risk for the chance of a larger payoff (Alos-Ferrer & Farolfi, 2019; Cox, 2004; Evans & Krueger, 2011). The trust game is an example of a situation in which we must decide whether to be selfish or strategize for a greater benefit. This often requires negotiating with the other party, but most trust game experiments do not allow participant pairs to communicate prior to decision making

(Boone & Buck, 2003; Lev-On, Chavez, & Bicchieri, 2010); few allow participants to communicate face-to-face beforehand (Ben-Ner, Putterman, & Ren, 2011; Johnson & Mislin, 2011; Lev-On et al., 2010; Zak, Barraza, Hu, Zahedzadeh, & Murray, 2022). This may be why "researchers in social dilemmas have been primarily interested in group effects and paid relatively little attention to individual level interactions or communicative processes" (Boone & Buck, 2003, p. 172). Although recently researchers have studied behavior in the game when computer-mediated (Lev-On et al., 2010) or video communication (Brosig-Koch, Ockenfels, & Weimann, 2003) is allowed, this may not be ecologically valid. Another flaw of many trust game experiments is the reliance on college students as a convenient sample; their behaviors may differ from those in the general population, limiting the generalizability of the results (Henrich, Heine, & Norenzayan, 2010). Finally, considering that the game causes players to weigh the costs and benefits of trust, a significant limitation is the extensive use of small stakes. Participants are often endowed with $10 or less, an amount that may not impact most people's lives (Johnson & Mislin, 2011).

But why are individuals motivated to exhibit trusting and trustworthy behavior towards a stranger? The literature has reached a consensus comprised of two broad answers: "because people have trusting dispositions and because others are trustworthy" (Baer & Colquitt, 2018, p. 163). In other words, individuals trust because they can communicate their trustworthiness through their dispositional qualities (Hardin, 2002, as cited in Boone & Buck, 2003) and because they experience another's trustworthiness in a given situation. Research also suggests that trust can be explained as a strategy for evolutionary stability (Berg, Dickhaut, & McCabe, 1995). "Placing trust in another might reflect an adaptive strategy to ensure positive social outcomes by broadcasting to other people that we have a reputation for being willing and reliable interaction partners (Berg et al., 1995; Jordan et al., 2016)" (Fareri, 2019, p. 2). But these theories do not explain the motivations behind trust and trustworthy behavior. Absent from existing models of trust is an examination of the dynamic process unfolding during trusting (van der Werff, Legood, Buckley, Weibel, & de Cremer, 2019). A recent volume (Searle, Nienaber, & Sitkin, 2018) notes the narrow focus within the trust literature and the absence of process focused theorizing and empirical work. For example, "people's willingness to be vulnerable toward others constitutes an aspect at the heart of the definition of trust that has remained unexplored" (van der Werff et al., 2019, p. 100). In general, we lack a clear understanding of the mechanisms involved in trust motivation (van der Werff, et al. 2019). Despite their importance, there is little consensus in the literature regarding the drivers of trust and trustworthiness (Cox et al., 2016).

Zaki and Mitchell (2013) argued that prosocial behaviors stem from intuitive, reflexive, and even automatic processes. For example, physiological responses have been found to be indicative of social decision making (Damasio, 1994). Physiological responses, such as heart rate (HR) and electrodermal activity (EDA), are examples of commonly measured autonomic processes that are difficult to consciously control (Critchley & Harrison, 2013; Figner & Murphy, 2011). EDA and skin conductance (SC) are commonly used in judgment and decision-making research because they serve as an indicator of arousal and affective processes (Figner & Murphy, 2011). For example, increases in SC are associated with risky behaviors, while trust is associated with decreases in SC (Bechara, Damasio, H., Tranel, & Damasio, A., 1997; Figner, Mackinlay, Wilkening, & Weber, 2009; Kéri & Kiss, 2011). In addition, SC is a *process tracing method*, a research method used to develop and test theories (Figner & Murphy, 2011; Schulte-Mecklenbeck, Kühberger, Ranyard, 2011). While researchers have extensively studied the neurochemical basis of trust (Zak, Borja, Matzner, & Kurzban, 2005; Zak, Kurzban, & Matzner, 2004; Zak, Kurzban, & Matzner, 2005), little research has been conducted on autonomic responses during trust situations (Mitkidis, McGras, Roepstorff, & Wallot, 2015). Physiological responses may be able to capture the unconscious process occurring when individuals choose to trust and be trustworthy. The aim of this research is to explore the role of physiology, specifically in the form of EDA, in trust and trustworthy behavior.

One aspect of physiology that has been studied extensively is physiological synchrony (PS). PS refers to "any interdependent or associated activity identified in the physiological processes of two or more individuals" (Palumbo et al., 2017, p. 100). The study began in the 1950s, when researchers found significant correlations between the EDA and HR of therapists and clients, which was interpreted as evidence of therapeutic rapport and empathy (Coleman, Greenblatt, & Solomon, 1956, as cited in Palumbo et al., 2017; Di Mascio, Boyd, & Greenblatt, 1957, as cited in Palumbo et al., 2017; Di Mascio, Boyd, Greenblatt, & Solomon, 1955, as cited in Palumbo et al., 2017). Thus, the presence PS between individuals is said to represent a physiological component of empathy (Palumbo, 2015). PS has been found to occur between individuals interacting in a variety of settings (Palumbo, 2015). It has been used to study romantic relationships (Levenson & Gottman, 1983; Liu, Palumbo, & Wang, 2016; McAssey, Helm, Hsieh, Sbarra, & Ferrer, 2013), and more recently to examine behavior between strangers playing economic games (Behrens, Snijdewint, et al., 2020; Mitkidis et al., 2015). Only recently have the mechanisms underlying strategic cooperation been explored through the lens of PS (Behrens, Snijdewint, et al., 2020). Behrens, Snijdewint, et al. (2020) theorize that "cooperation flourishes when people synchronize their autonomic responses because they align emotional states based on genuine emotional cues that are perceived by interaction partners" (p. 4).

The present research expands the literature on PS to explore the process underlying trust, trustworthiness, and cooperation, specifically when high monetary stakes are involved. I hypothesize that PS is an underlying mechanism of trust, and that synchronized neurophysiology is a predictor of trust and trustworthiness. The goal is to determine whether PS between two individuals while they discuss trusting behaviors will predict their subsequent monetarily meaningful decisions.

Participants were recruited to play high-stakes versions of the trust and dictator game with an endowment of $120. They were asked to play the games with two other participants in two separate sessions. Before making their decisions, participants talked face-to-face with each other for two minutes and were instructed to make promises to each other about how much money they would transfer in the trust game. Thus, participants had to decide whether to trust that their dyadic partner would keep their promise based on this interaction. EDA was collected during these interaction periods, and I hypothesize that subjects who were subsequently trusting and trustworthy would have higher physiologic synchrony, as measured by their skin conductance levels (SCLs), than individuals who had low or no trust or trustworthiness. SCL data was selected for analysis for two reasons. First, SC is a relatively fast responding measure of arousal, as it changes within 0.5 to 5 seconds of stimulus onset (Boucsein, 1992; Figner & Murphy, 2011; Healey, Seger, & Picard, 1999). Second, in their study on PS between participants playing several rounds of the prisoner's dilemma game, Behrens, Snijdewint, et al. (2020) found that only SCL synchrony, but not HR synchrony, predicted cooperation. The innovation in the present research is to test whether PS during pre-decision conversation would predict trusting behaviors.

A few differences between this experiment and traditional trust games should be noted. First, participants were endowed with a more considerable sum of money than many other trust game experiments (Johnson & Mislin, 2011) to encourage participants to perceive that their decisions have significant consequences, i.e., for ecological validity. Second, in contrast to most trust game experiments that study college students (Johnson & Mislin, 2011), participants were working-age adults between the ages of 25–50 from diverse cultural backgrounds. Finally, few studies on the trust game allow participant pairs to communicate face-to-face before making their decisions (Ben-Ner et al., 2011; Johnson & Mislin, 2011; Lev-On et al., 2010; Zak et al. 2022). However, the lack of communication greatly limits the generalizability of these studies to any out-of-lab interactions in which people interact face-to-face (Zak et al. 2022). Instead, this experiment took place under ecologically valid circumstances, which allowed partners to interact and discuss their decisions beforehand.

Based on van der Werff et al.'s (2019) model on trust motivation (discussed in 2.1.2. Trust Motivation), if two individuals express mutual arousal, exhibited by their SCL, this could lead to the motivation to trust. Within the context of the trust game, synchronous changes in physiological arousal may motivate players to cooperate because they expect a positive outcome, i.e., to benefit from cooperation (Isoni & Sudgen, 2019). In other words, PS may induce the motivation to trust (and be trustworthy) in the game, leading to the joint willingness to engage in reciprocal cooperation (Isoni & Sudgen, 2019). I aim to test whether participant pairs who exhibit trusting and trustworthy behavior, as measured by the dollar amounts transferred in the trust and dictator games, display synchronous physiological signals.

To measure participant pairs' PS, I use two measures of synchrony, intersubject correlation (ISC) and dynamic time warping (DTW). ISC is a windowed and lagged cross-correlation measure used to quantify the commonalities between non-stationary signals such as SC and HR (Golland, Keissar, & Levit-Binnun, 2014; Lahnakoski & Chang, 2021). DTW is an algorithmic technique that uses dynamic programming optimization to compare two sequences by minimizing the cumulative distance between them (Berndet & Clifford, 1994). Both methods are suitable for non-stationary data; together they provide two different means of assessing the concordance of two signals.

This is the first study to analyze the PS of individuals playing the trust game preceded by face-to-face interactions. In addition, to my knowledge, only two papers studying PS in the context of economic games have been published to date. Mitkidis et al. (2015) used a public goods game and Behrens and Snijdewint, et al. (2020) a prisoner's dilemma game; thus, PS in the trust game remains unexplored prior to this study. Using the trust game to measure trust and trustworthy behavior, this study examines the electrodermal response associated with these behaviors. The goal is to determine whether individuals who display trusting or trustworthy behavior in the game also exhibit PS, as measured by their SCL.

The following is a summary of the chapters contained in this dissertation. Chapters two and three provide a summary of the literature on trust, physiological arousal and synchrony, and PS methodologies. Chapter four outlines the methodology, experiment design, and hypotheses. The results of the study are presented in chapter five. Lastly, chapter six summarizes the findings and limitations of the study and provides suggestions for future research examining PS within the trust game.

II. Literature Review

This literature review is comprised of two sections: trust and physiological arousal and synchrony. The first section examines trust and trustworthy behavior through the lens of both psychology and economics, as well as its measurement in economics using the trust game. I review definitions and describe a model of the determinants of trust from Thielmann and Hilbig (2015), and a model of trust motivation by van der Werff et al. (2019). The neoclassical definition of trust and its role in game theory is examined. Within the field of game theory, I explain the trust and dictator game, which is used as a control for other-regarding (social) preferences. Next, I describe social preferences and the importance of controlling for these confounds to elicit a measure of trust with construct validity. Finally, I discuss the importance of the 'communication effect' on outcomes in the trust game.

Section two provides an overview of physiology, the autonomic nervous system (ANS), and PS. I describe the mechanisms behind our automatic, physiological response during decision making. Next, I explain the mechanisms underlying PS. Finally, I review the existing literature on PS in economic games – to my knowledge, only two studies have been published examining PS during economic game play.

1. **Trust**

1. ***Interpersonal Trust***

Interpersonal trust is a behavior, referring to cooperation that occurs between individuals (Evans & Krueger, 2009). Trusting behavior allows one to accept vulnerability based on a positive expectation of the behavior of another (Rousseau, Sitkin, Burt, & Camerer, 1998). The most commonly cited definition of trust is from Mayer et al. (1995), who defined trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712; Thielmann & Hilbig, 2015). Making oneself vulnerable is risky because of this uncertainty and the possibility of an adverse outcome or loss (Hong & Bohnet, 2007; Mayer et al., 1995; Thielmann & Hilbig, 2015). This is especially true when the SM has an incentive to take advantage of the FM's vulnerability and betray them for personal gain, making the FM worse off (Hong & Bohnet, 2007; Thielmann & Hilbig, 2015). In a situation with uncertain outcomes, depending on another means accepting risk and personal vulnerability (Hong & Bohnet, 2007; Thielmann & Hilbig, 2015).

Rousseau et al., (1998) identified two necessary conditions for trust: risk and interdependence. Across disciplines, risk is a necessary condition in the conceptualization of trust since "the need for trust only arises in a risky situation" (Mayer et al., 1995, p. 711; Rousseau et al., 1998). Rousseau et al. (1998) defined risk as "the perceived probability of loss, as interpreted by a decision maker" (p. 395). Trust is inherently risky because of the possibility of loss and regret (Evans & Krueger, 2009). Thus, trust cannot occur without taking a risk and accepting the possibility of betrayal (Thielmann & Hilbig, 2015). Interdependence occurs when the outcome of an interaction depends in part on the action(s) of another person (Balliet & Van Lange, 2013). According to a meta-analysis conducted by Balliet and Van Lange (2013), social interdependence is one of the most agreed upon elements of trust since it is an explicit indicator of the willingness to trust (Thielmann & Hilbig, 2015).

Additionally, we trust because we expect a positive outcome. It is rational to trust only when one believes that another is sufficiently trustworthy, i.e., they will act in a benevolent and favorable manner; otherwise, trusting is self-destructive (Evans & Krueger, 2009; Thielmann &

Hilbig, 2015). This illustrates the vital role of expectations (referred to as *beliefs* in economics) on trust behavior (Thielmann & Hilbig, 2015). For example, according to Cox (2004), "trust is inherently a matter of the beliefs that one agent has about the behavior of another" (p. 263). This is why Evans & Krueger (2009) argued that any definition of trust must contain two core elements: vulnerability and expectation. Together with Rousseau et al.'s (1998) necessary conditions for trust, any definition of trust must contain three critical elements: uncertainty and risk, vulnerability, and expectation (Evans & Krueger, 2009; Thielmann & Hilbig, 2015).

In their review, Thielmann & Hilbig (2015) combined these three conditions and incorporated the diverse perspectives of the behavioral sciences to construct an updated definition of trust. They define trust as "a risky choice of making oneself dependent on the actions of another in a situation of uncertainty, based upon some expectation of whether the other will act in a benevolent fashion despite an opportunity to betray" (Thielmann & Hilbig, 2015, p. 251). According to their definition, trust implies risk and uncertainty, is predicated on the expectation that the SM will behave positively and benevolently, and necessitates accepting personal vulnerability, loss, and betrayal (Hong & Bohnet, 2007; Thielmann & Hilbig, 2015).

> In line with previous conceptualizations of trust, this definition includes all core components identified in earlier research (i.e., uncertainty, risk, expectation, and vulnerability toward betrayal), but extends these definitions by incorporating the diverse perspectives found in different scientific disciplines: First, we explicitly distinguish between uncertainty as a feature of the trust situation (i.e., the absence of conclusive information about the trustee's trustworthiness) and risk as a feature of the trusting action (i.e., accepting the possibility of a loss arising from the other's opportunity to betray). (Thielmann & Hilbig, 2015, p. 251)

This definition is also the first to differentiate between trust cognition (i.e., expectation) and behavior (i.e., making a risky choice). Finally, unlike the definitions of trust above from Mayer et al. (1995) — "the willingness…to be vulnerable" (p. 712) — and Rousseau et al. (1998) — "a psychological state" (p. 395) — Thielmann and Hilbig's (2015) definition allows trust to be easily operationalized (Thielmann & Hilbig, 2015). This is the main advantage of their definition; it conceptualizes trust behavior as quantifiable and directly observable (Thielmann & Hilbig, 2015).

Finally, Coleman (1990) outlined four central features of a trust situation. First, the one's decision to trust must be voluntary. Second, there should be a time delay between the FM's and SM's decisions. Third, the FM exhibiting trust is a necessary condition for the SM to abuse or honor the demonstrated trust. Fourth, the FM is vulnerable to the SM, such that the FM is worse off had no trust been shown. Alos-Ferrer & Farolfi (2019) added a fifth feature with regards to economic efficiency. Namely, that the decision to trust should be an optimal one from the standpoint of maximizing the sum of payoffs.

1. 1. **Determinants of Trust**

Economists and psychologists study trust differently, with psychologists assessing trust in terms of the attributes of the individuals, focusing upon the internal cognitions of these personal attributes (Rousseau et al. 1998). Economists view trust as calculative, and influenced by the situation, context, and institution(s) (Rousseau et al., 1998). For example, in Williamson's (1993) review of the trust literature, he made a distinction between calculative and personal trust (Berg et al., 1995). Calculative trust is the result of the FM making the decision to trust based on a calculation of expected utility (Berg et al., 1995). Personal trust is described as a feeling one has that trusting another is the right decision (Berg et al., 1995). A blending of the two approaches provides the most comprehensive view, as trust is a function of "dispositions, decisions, behaviors, social networks, and institutions" (Rousseau et al., 1998, p. 394).

The determinants of trust are dependent upon the relevant characteristics of the individual, the situation, and their interaction (Evans & Krueger, 2009). Researchers have found consistent and significant individual differences when measuring trusting behavior (Evans & Krueger, 2009). Specifically, the decision to trust is related to personality differences, social identity, and expectations, and these differences are in part predicted (and perhaps caused) by an underlying propensity, or disposition, to trust (Evans & Krueger, 2009; Rotter, 1967). Dispositional trust is "the general expectation that others will behave fairly and responsibly" (Evans & Krueger, 2009). Dispositional theory posits that "trust is motivated by individual differences in a general expectation" (Evans & Krueger, 2009, p. 1014). For example, individuals develop differing amounts of *generalized expectancy* – the learned expectation that promises kept between other social agents will generalize from one social agent to another (Rotter, 1967). "The development of such a generalized attitude may be learned directly from the behavior of parents, teachers, peers, etc., and also from verbal statements regarding others made by significant people or trusted sources of communication such as newspapers and television" (Rotter, 1967, p. 653). Studies suggest that children that experience a higher proportion of promises kept by parents and authority figures in the past have a higher generalized expectancy for trust from other authority figures (Rotter, 1967). Provided that reliable and validated instruments are used, behavior is consistently found to be correlated with disposition (Evans & Krueger, 2009). Rotter's (1967; 1971; 1980) conceptualization of trust "has been referred to by other scholars as dispositional trust, generalized trust, and trust propensity" (Baer & Colquitt, 2018, p. 166).

Dispositional trust is "the general expectation that others will behave fairly and responsibly" (Evans & Krueger, 2009). Dispositional theories explain interactions between strangers, referred to as initial trust, while trust in established relationships is more dependent upon context and experience (Evans & Krueger, 2009; McKnight, Cummings, & Chervany, 1998, as cited in Thielmann & Hilbig, 2015). Economists study trust in social exchanges, typically in situations with undeveloped or short-term relationships (Evans & Krueger, 2009). Thus, economists are more interested in one's propensity to trust since it "is most relevant in ambiguous situations; dispositions matters most when there is little first-hand knowledge or past experience" (Evans & Krueger, 2009, p. 1011). Economists focus on state trust, a measurement of how much someone expects another to cooperate in an experimental social dilemma (Balliet & Van Lange, 2013). Furthermore, economists study trust in one-time settings, which is typically derived by the estimation of gains and losses and weighed by perceived risks, making it easier to study in a controlled setting (Rousseau et al., 1998).

Based on Thielmann and Hilbig's (2015) definition of trust (described in the previous section), they identified three critical determinants of trust: trustworthiness expectations (judgments/assessments), attitudes toward risky prospects (i.e., risk aversion and loss aversion), and betrayal sensitivity, and modeled their relationship and antecedents. The following is a summary of these elements, their components, relationships, and relevant personality traits, as detailed in Thielmann and Hilbig (2015).

In interactions with strangers or in the absence of information about another's trustworthiness, one must form an expectation about another's likely behavior and infer their trustworthiness. Given that trust decisions are conceptually similar to decisions under risk, trust behavior is conditioned on both the probabilities and outcomes associated with trust realization (gain) versus betrayal (loss). Thus, trustworthiness expectations represent an individual's probability estimates of gains versus losses (Thielmann & Hilbig, 2015). Thielmann and Hilbig (2015) suggest that trustworthiness expectations may be inferred from three implicit sources: personal and situational trust cues (depending on their availability), prior trust experiences to the one at hand, and social projection of one's own trustworthiness behavior, which represent the underlying personality trait aspects – fairness and honesty – of trustworthiness expectations (Thielmann & Hilbig, 2015). They theorize that these pieces of information enable someone to derive a probability estimate of trust realization (gains) versus betrayal (losses) (Thielmann & Hilbig, 2015). Nevertheless, depending on the available evidence, one's judgment may not be an accurate reflection of the actual situation (Thielmann & Hilbig, 2015).

1. 1. 1. ***Trustworthiness Expectations***

The ability to detect the trustworthiness of another, and in turn identifying a cooperative partner, using available cues is an important evolutionary skill (Boone & Buck; Krumhuber et al., 2007; Scharlemann, Eckel, Kacelnik, & Wilson, 2001). Trust cues are the implicit and explicit pieces of information an individual uses to draw inferences about another's trustworthiness, related to another (i.e., personal trust cues) or the situation (i.e., situational trust cues) (Funder, 2008, as cited in Thielmann & Hilbig, 2015; Thielmann & Hilbig, 2015). Personal trust cues describe the individual characteristics that may be used to predict another's trustworthiness, such as appearance or social category (Thielmann & Hilbig, 2015). Personal trust cues are especially relevant in interactions with strangers, where one has little first-hand knowledge. These refer to individuals' outward appearance, including facial features (Todorov, 2008; Todorov, Baron, & Oosterhof, 2008; Todorov, Pakrashi, & Oosterhof, 2009), facial expressions (Krumhuber et al., 2007; Scharlemann et al., 2001), facial expressivity (Boone & Buck, 2003), and body language (Meeren, van Heijnsbergen, & de Gelder, 2005; Naumann, Vazire, Rentfrow, & Gosling, 2009) (Thielmann & Hilbig, 2015).

An individual's face plays a significant role in signaling social intentions, and people immediately form strong and reliable impressions of someone on the basis of their facial appearance (van 't Wout & Sanfey, 2008). Research shows that trustworthiness judgments based on facial appearance are automatic, general valence assessments (Oosterhof & Todorov, 2008) that can be formed in as little as 100ms, with additional exposure increasing individuals' confidence in their impressions (Willis & Todorov, 2006). Todorov, Mende-Siedlecki, and Dotsch (2013) used principal component analyses to determine that valence evaluations of faces account for about 60% of the variance of participants' trustworthiness judgments. "Such valence evaluation permeates social judgments (Kim & Rosenberg 1980; Rosenberg et al. 1968; cf. Osgood et al. 1957) and is one of the organizing principles of person impressions (Wyer & Srull 1989)" (Todorov, 2008, p. 210). These valence (i.e., trustworthiness) evaluations are based on the perception of facial features expressing emotions, such as happiness and anger, indicating individuals' inferences of behavioral intentions and signaling approach-avoidance behavior (Todorov, 2008). The second principal component, dominance, accounted for 18.3% of the variance of participants' trustworthiness perceptions. Oosterhof & Todorov (2008) used 300 faces to model trustworthy and dominant looking faces, which they found were highly correlated with trustworthiness and dominance assessments, respectively. Dominance evaluation is reflected by facial features signaling physical strength/weakness and aggression (Oosterhof & Todorov, 2008). Threatening faces are perceived as untrustworthy; faces with extreme dominance are often labeled as angry, and vice versa (Oosterhof & Todorov, 2008; Todorov, 2008). The use of valence and dominance related facial cues can signal an individual's intentions, whether they are harmful or harmless, and whether the individual is capable of causing harm (Oosterhof & Todorov, 2008). Together, valence and dominance are sufficient to describe evaluations of neutral faces that resemble judgments of trustworthiness and dominance (Oosterhof & Todorov, 2008).

Research shows that emotionally neutral faces are evaluated using evolutionary cues, and by the same mechanisms underlying the perception of emotional expressions (Todorov, 2008; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). A person's facial expressions communicate information about their mental and emotional state, signaling behavioral intentions, and in turn appropriate response behavior (Ekman, 1982, as cited in Krumhuber et al., 2007; Todorov & Oosterhof, 2011; Todorov, 2008). Todorov (2008) argued that signals of approach-avoidance behaviors might be derived from emotional expressions inferred by face evaluations. This may be because faces that are perceived negatively reflect subtle features resembling angry facial expressions, while faces that are perceived positively reflect subtle features resembling happy facial expressions (Todorov, 2008). In experiments where participants evaluated emotionally neutral faces, perceptions of trustworthiness were positively correlated with perceptions of happiness and negatively correlated with perceptions of anger (Todorov, 2008). When there are no clear emotional cues to illuminate the intentions of the individual, Oosterhof and Todorov

(2008) reasoned that faces are evaluated based on their similarities to expressions of anger and happiness. To assess these intentions, faces are evaluated based on their facial maturity in the absence of other cues (Oosterhof & Todorov, 2008). For example, a person with a facial expression resembling anger may be perceived as having a dominant or aggressive personality (Todorov & Oosterhof, 2011). In fact, studies show that faces that appear angry can trigger automatic avoidance responses and are thus perceived as unapproachable and untrustworthy (Adams, Ambady, Macrae, & Kleck, 2006, as cited in Todorov, 2008; Marsh, Ambady, & Kleck, 2005, as cited in Todorov, 2008; Oosterhof & Todorov, 2008; Rule, Krendl, Ivcevic, & Ambady, 2013). Alternatively, a person with a facial expression resembling happiness, i.e., smiling, is perceived as friendly and approachable, and therefore more likely to cooperate (Oosterhof & Todorov, 2008; Todorov, 2008). Todorov (2008) reasoned that if faces are evaluated by the same mechanisms underlying the perception of emotional expressions, then constructing model faces with exaggerated facial features should generate faces with emotional expressions. Using these face models, he found that exaggerating the facial features of a neutral face to look more trustworthy generated expressions of happiness (Oosterhof & Todorov, 2008; Todorov, 2008). Correspondingly, exaggerating the facial features to make a neutral face look untrustworthy generated expressions of anger (Oosterhof & Todorov, 2008; Todorov, 2008).

Regarding facial features, Todorov et al. (2008) identified four features highly correlated with trustworthiness perceptions. Their experiment/study found that “faces with high inner eyebrows, pronounced cheekbones, wide chins and shallow nose sellion looked more trustworthy than faces with low inner eyebrows, shallow cheekbones, thin chins and deep nose sellion” (p. 121). These features accounted for 29.4% of the variance in participants’ trustworthiness assessments. Faces that are perceived as trustworthy appear more feminine and baby-faced, which can be manipulated using model faces by increasing the distance between the eyes and the eyebrows (Keating, Mazur, & Segall, 1981, as cited in Oosterhof & Todorov, 2008; Zebrowitz, 1999, as cited in Oosterhof & Todorov, 2008). When faces were manipulated to appear more attractive, they became thinner, with higher cheekbones; when they were manipulated to appear more extroverted, they became wider and looked happier (Todorov & Oosterhof, 2011). On the other hand, faces manipulated to look more dominant appeared more masculine, darker, and mature, with facial shape serving as a cue for maturity (Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011). These structural properties of the face are assessed by the brain for cues regarding (un)trustworthiness (Todorov, 2008). We make social inferences from facial appearance based on their resemblance to features with adaptive significance, and we detect emotional expressions to assist in successful social exchange (Krumhuber et al., 2007; Scharlemann et al., 2001; Todorov & Oosterhof, 2011; Zebrowitz & Montepare, 2008).

Kret (2015) emphasized that the eye region is particularly salient and powerful since it provides implicit cues used to assess someone’s trustworthiness. Since changes in one’s pupils occur unconsciously, it provides more accurate information than facial expressions and eye gaze, which can be controlled or faked (Kret, 2015). Hess (1975, as cited in Kret, 2015) was the first to introduce the study of pupil dilation and mimicry anecdotally, hypothesizing that changes in an interacting partner’s pupil size are contagious (Kret, Fischer, & De Dreu, 2015). His research showed that one’s pupil size was positively correlated to their level of interest in the object or person they were observing (Hess, 1975, as cited in Kret & De Dreu, 2019). Additional research has shown/demonstrated that humans are sensitive to pupil size changes, i.e., dilation and constriction, observed in another individual serve a communicative function and pupil mimicry is commonly observed (Kret, 2015; Kret et al., 2015; Kret, Tomonaga, & Matsuzawa, 2014). For example, Kret et al. (2015) found that pupil dilation mimicry predicted trust between in-group partners in the trust game, especially when their eyes displayed happiness. “The finding that humans process another’s pupil size may imply not only that humans attend to their companion’s pupils, but also that they automatically synchronize their own pupil movements with them and – via pupil mimicry – quickly and automatically infer whether or not their partner is trustworthy” (Kret et al., 2015, p. 2). Finally, Prochazkova et al. (2018) found that pupil dilation mimicry between interacting partners of the same in-group activated key theory-of-mind network areas of the brain, known to be involved in social cognition, including prosocial behavior and trust formation.

Research shows that individuals also infer personality traits from different components of physical appearance (Naumann et al., 2009). Todorov and Oosterhof (2011) found that emotionally neutral faces that were judged as attractive, extroverted, competent, or likable were perceived as more trustworthy than those judged/rated as dominant, threatening, frightening, or mean. Todorov (2008) argued that even neutral facial expressions contain subtle cues resembling/reflecting emotional expressions, which we interpret as signaling personality dispositions. In another study Naumann et al. (2009) collected personality trait information and took full-body photographs of one set of participants. They asked another set of participants to rate the individuals in the photos to determine which appearance-based cues were associated with the photographed participants’ (actual) personalities and which were only associated with the observers’ judgments (Naumann et al., 2009). The characteristics were classified as either static, referring to physical grooming such as clothing style and hairstyle, or dynamic, referring to nonverbal expressive behaviors such as facial expressions and posture (Naumann et al., 2009; Riggio, Widaman, Tucker, & Salinas, 1991, as cited in Naumann et al., 2009). Naumann et al.

(2009) found that observers perceived the photographed participants who were smiling and standing in a more energetic and relaxed way (vs having a tense or tired stance) as more likable (Naumann et al., 2009). Observers formed accurate impressions of the extroverted individuals (based on the personality questionnaire), identified as those smiling, having an energetic stance, and looking stylish and healthy (Naumann et al., 2009). Individuals with higher self-esteem stood with their arms behind their backs and were also judged as looking healthier, smiling more, and having a more energetic stance (Naumann et al., 2009). In addition, individuals who were highly agreeable – a potential predictor of cooperativeness – were more likely to smile and stand in a more relaxed manner (Denissen & Penke, 2008, as cited in Thielmann & Hilbig, 2015; Naumann et al., 2009). Naumann et al.'s (2009) findings demonstrate that we use both static and dynamic aspects of one's physical appearance to form impressions, and that these characteristics are reflected in certain personality traits (Naumann et al., 2009).

Humans evaluate faces to understand the communicative meaning of emotional expressions (Todorov, 2008). Boone & Buck (2003) argued that the communication of one's trustworthiness is an affective process and is thus governed by the communication of emotion (Boone & Buck, 2003). They theorized that emotional expressivity, defined as displaying or communicating one's emotions accurately, serves an evolutionary purpose as a marker for cooperative behavior and trustworthiness since emotionally expressive individuals are more likely to reveal their motivational intentions (Boone & Buck, 2003). It makes sense that those who are easy to "read" emotionally will be preferred by individuals trying to coordinate their actions (Boone & Buck, 2003). Furthermore, emotionally expressive individuals have a lower probability of successfully deceiving others in a social exchange (Boone & Buck, 2003). Thus, "the interpersonal communication of emotion, based upon emotional expressivity, is a critical component to establishing conditions under which two people, even strangers, can feel each other out as trustworthy or untrustworthy" (Boone & Buck, 2003, p. 174). Confirming this theory, Schug, Matsumoto, Horita, Yamagishi, and Bonnet (2010) found that cooperative participants displayed a greater number of emotional expressions, both positive and negative, suggesting cooperators express more emotions than non-cooperators. Based on these findings, they hypothesized that the expression of positive emotion alone is not as revealing of one's cooperativeness as general emotional expressivity (Schug et al., 2010). Although emotions can be faked and exaggerated, researchers have found that we are able to distinguish between real and fake smiles (Krumhuber et al., 2007). Smiling is one of the most effective communication signals, but it is also one of the easiest to fake (Ekman & Friesen, 1982, as cited in Krumhuber et al., 2007; Ekman, Friesen, & O'Sullivan, 1988, as cited in Krumhuber et al., 2007; Krumhuber et al., 2007). Krumhuber et al. (2007) showed participants a video clips of smiling potential trust game partners and found that those displaying authentic smiles were rated as more likable, attractive, cooperative, and trustworthy than fake-smiling or unexpressive individuals. Importantly, participants were more likely to elect to play the game with an authentically smiling counterpart (Krumhuber et al., 2007).

One's social category serves as another personal trust cue, as information about a person's social category, such as their age, gender, and race, can be quickly inferred from their facial characteristics (Todorov, 2008). The field of social vision has proven that face and body cues drive assessments of social category membership (Freeman, Johnson, Adams, & Ambady, 2012, as cited in Lick & Johnson, 2014; Johnson & Adams, 2013, as cited in Lick & Johnson, 2014). But social categorization leads to stereotypical and prejudiced beliefs that can shape perceivers' impressions of others (Allport, 1954, as cites in Lick & Johnson, 2014). For example, individuals often display a behavior known as group-based trust, whereby individuals show a bias towards trusting strangers who are in-group members more than out-group members (Foddy, Platow, & Yamagishi, 2009, as cited in Thielmann & Hilbig, 2015; Platow, Foddy, Yamagishi, Lim, & Chow, 2012, as cited in Thielmann & Hilbig, 2015). Social categorizations can trigger negative stereotypes that guide decisions and interpersonal behaviors accordingly (Devine, 1989, as cited in Lick & Johnson, 2014; Sinclair & Kunda, 1999, as cited in Lick & Johnson, 2014). Knowing another's social category can also elicit stereotypes related to trustworthiness expectations (Brewer, 2008, as cited in Thielmann & Hilbig, 2015). Research shows that individuals exhibit more optimism regarding the trustworthiness of in-group members (Balliet, Wu, & De Dreu, 2014, as cited in Thielmann & Hilbig, 2015). Thus, perceptions shaped by social categorization and stereotypes may be positive or negative (Brewer, 2008, as cited in Thielmann & Hilbig, 2015; Lick & Johnson, 2014).

We automatically infer meaning, such as intentions, personality traits, and social characteristics (e.g., gender, age, race) from another's facial features and appearance and use these when making decisions, such as whether to cooperate in a risky decision-making game (Naumann et al., 2009; Behrens, Snijdewint, et al., 2020; Todorov et al., 2008; van 't Wout & Sanfey, 2008). For example, van 't Wout and Sanfey (2008) found that participants playing the trust game sent more money to partners whose faces were rated as more trustworthy, thus demonstrating greater trust in them. In trust-related contexts, facial expressions provide relevant behavioral and situational information regarding who and when to trust (Boone & Buck, 2003; Krumhuber et al., 2007). "Participants were indeed influenced by the perceived subjective trustworthiness of the partner in their judgment of whether to trust, and how much trust to place" (van 't Wout & Sanfey, 2008, p. 799). Even when no context is provided, trustworthiness judgments of faces are based on inferences about the positivity or negativity of the face (Todorov, 2008). Facial signals influence how we judge and behave towards others, and in an interactive social context, facial features influence strategic decision-making by providing non-explicit, but automatic social information (van 't Wout & Sanfey, 2008).

The accurate assessment of another's trustworthiness is an important facet of a successful social exchange (Fridlund, 1994 as cited in Oosterhof & Todorov, 2008; Ghanzafar & Santos, 2004, as cited in Oosterhof & Todorov, 2008). But research has shown that trustworthiness assessments based on facial cues are consistently inaccurate (Thielmann & Hilbig, 2015). Some studies have found that individuals accurately predict the trustworthiness of others based on facial features (Stirrat & Perrett, 2010), while others have found that perceptions of trustworthiness do not correspond to objective measures (Rule et al., 2013). For example, Bond, Berry, and Omar (1994; as cited in Oosterhof & Todorov, 2008) measured the trustworthiness of actual behavior and found a modest positive correlation – ratings of honesty from faces only explained 4% of the variance of behavior. Thus, the "reliance on trust cues has no immediate implication for whether or not the trustworthiness judgment ultimately formed is, in fact, correct or accurate" (Thielmann & Hilbig, 2015, p. 257). To explain this discrepancy, Rule et al. (2013) suggested that trustworthiness judgments are simply widely agreed upon subjective impressions. For example, as discussed previously, researchers have found that faces expressing positive emotions, e.g., happiness, are perceived as approachable and trustworthy, whereas faces expressing negative emotions, e.g., anger, are perceived as unapproachable and untrustworthy (Oosterhof & Todorov, 2008; Rule et al., 2013; Todorov, 2008). These judgments reflect one's impressions of subtle emotional cues, which are misattributed to stable personality traits and trustworthy behavior (Oosterhof & Todorov, 2008). This is explained by the emotion overgeneralization hypothesis (Zebrowitz & Montepare, 2008), which states that trustworthiness judgments of both emotionally neutral and emotionally expressive faces are, in fact, over-generalizations; they do not accurately reflect one's personality traits or behavior (Oosterhof & Todorov, 2008; Rule et al., 2013; Todorov, 2008; Zebrowitz, Kikuchi, & Fellous, 2010). These over-generalizations are the consequence of adaptive mechanisms which lead us to infer meaning from emotional expressions (Oosterhof & Todorov, 2008; Todorov, 2008). Therefore, while researchers have found that about 80% of the variance of participants' trustworthiness judgments are valence and dominance evaluations (Todorov et al., 2013), these are not accurate reflections of participants' actual traits or behavior (Bond et al., 1994, as cited in Oosterhof & Todorov, 2008; Knutson, 1996; Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Rule et al., 2013; Todorov, 2008).

The second type of trust cue, situational cues, refer to aspects of a given trust situation, such as the cost, benefit, temptation to betray, and availability of sanctions or enforcement, that may affect one's probability estimate of trust realization versus betrayal (Thielmann & Hilbig, 2015). An important situational trust cue is one's temptation to betray, i.e., the difference in payoff between honoring and betraying the trust, also known as their conflict of interest (Evans & Krueger, 2011; Kelley, Holmes, Kerr, Reis, Rusbult, & Van Lange, 2003; Thielmann & Hilbig, 2015). Researchers have found that temptation is the strongest predictor of strangers' reciprocity (Snijders & Keren, 1999, as cited in Evans & Krueger, 2011). Empirically, increased temptation causes trust behavior to decline (Evans & Krueger, 2011, 2014; Snijders & Keren, 1999, 2001, as cited in Thielmann & Hilbig, 2015). Interestingly, FMs tend to focus primarily on their own potential gains and underestimate the risk of betrayal when their interaction partner's temptation to betray is high (Evans & Krueger, 2011). Evans & Krueger (2011) found that FMs assessed the temptation of a SM only when their own potential gains were favorable, i.e., egocentrism exceeded perspective-taking. Lastly, "trusting is arguably a 'learned behavior'" (Blair & Stout, 2001, p. 1742; as cited in Thielmann & Hilbig, 2015, p. 258). Thus, prior trust experiences serve as another important heuristic that guides individuals' decisions to trust, especially in situations with few or ambiguous trust cues (Thielmann & Hilbig, 2015). If a recent prior trust situation had a favorable outcome, one might consider this and decide to trust in the current situation, expecting a similar positive outcome.

Some scholars argue that one's trustworthiness characteristics are the primary determinants of trust (Baer & Colquitt, 2018). These characteristics encompass three aspects: ability (competence), benevolence, and integrity (Mayer et al., 1995). Baer and Colquitt (2018) proposed an additional antecedent of trust influenced by less conscious processes, heuristics. One such heuristic is projecting one's own trustworthiness behavior onto another, expecting another to behave as they themselves would (Hill & O'Hara 2006; Krueger & Acevedo, 2005, as cited in Thielmann & Hilbig, 2015; Thielmann & Hilbig, 2015). Social projection exemplifies an individual's underlying personality traits, reflecting their trait fairness and honesty (Thielmann & Hilbig, 2015). Thus, individuals form trustworthiness expectations by projecting their own characteristics onto another, imagining that another person possesses their same personality traits and cooperative preferences (Hill & O'Hara 2006; Thielmann & Hilbig, 2015). Considering this, the personality traits that determine an individual's trustworthiness – cooperativeness and honestly – should be predictive of trust behavior (Thielmann & Hilbig, 2015). In fact, Thielmann and Hilbig (2014) identified that cooperative traits, which are reducible to fairness traits, are predictors of trustworthiness expectations (Thielmann & Hilbig, 2015). Additionally, cooperative individuals expressed more optimism regarding a stranger's fairness, i.e., trustworthiness

(Thielmann & Hilbig, 2014). In their review, Thielmann and Hilbig (2015) clarified the relationship between trust and trustworthiness, which had previously been lacking from the literature (Colquitt, Scott, & LePine, 2007, as cited in Thielmann & Hilbig, 2015). This relationship can be explained as individuals projecting their own trustworthiness traits onto others (Thielmann & Hilbig, 2014, 2015).

Trust cues, situational cues, prior trust experiences, and social projection together form one's trustworthiness expectations, the first of three core components of trust behavior outlined by Thielmann and Hilbig's (2015) model on the determinants of trust. Trust cues, situational cues, and prior experiences pertain to a given trust situation, whereas social projection reflects the personality traits underlying trustworthiness expectations (Thielmann & Hilbig, 2015). The second determinants, attitudes toward risky prospects (i.e., risk aversion and loss aversion), are discussed next.

1. 1. 2. ***Risk Aversion and Loss Aversion***

Since the decision to trust involves risk and uncertainty, Thielmann and Hilbig's (2015) model of the underlying determinants of trust behavior incorporates two prominent attitudes towards risky prospects: risk aversion and loss aversion. Risk aversion refers to an individual's attitude towards uncertainty and is defined as someone who generally prefers outcomes with low uncertainty over high uncertainty, even if the return of the latter is greater (Werner, 2008). Within the field of decision-making research, uncertainty is defined as the decision maker's unfamiliarity with the probability of a gain versus a loss (Thielmann & Hilbig, 2015). Uncertainty is relevant in a trust situation because the decision maker is uncertain about another's trustworthiness and thus faces an unknown probability of gain versus loss (Thielmann

& Hilbig, 2015). Thielmann and Hilbig (2015) define risk aversion as an individual's "willingness to take a risk as a function of the probabilities of gain vs. loss" (p. 254). Studies have shown that dispositional risk aversion is a meaningful predictor of FM transfer amounts in the trust game (Altmann, Dohmen, & Wibral, 2008; Bigoni, Bortolotti, Casari, & Gambetta, 2013; Fetchenhauer & Dunning, 2012; Karlan, 2005; Lönnqvist, Verkasalo, Walkowitz, & Wichardt, 2010; Sapienza, Toldra-Simats, & Zingales, 2013; Schechter, 2007).

Loss aversion was first identified by psychologists Daniel Kahneman and Amos Tversky in their seminal paper introducing prospect theory, a framework for decision-making under risk (Kahneman & Tversky, 1979). Loss aversion influences one's preferences (i.e., utilities) towards potential gains and losses and is defined as someone who would prefer avoiding a loss over the prospect of acquiring an equivalent gain (Kahneman & Tversky, 1979). Thielmann and Hilbig (2015) define loss aversion as an individual's "willingness to take a risk as a function of the relation between positive and negative outcomes" (p. 254). Risk is defined as exposure to a potential loss, or the presence of both a gain or a loss (*The Definition of Risk*, 2019); Thielmann & Hilbig, 2015). Risk is present in a trust situation because the SM has an opportunity to betray the FM, resulting in the FM experiencing a loss (Hong & Bohnet, 2007; Thielmann & Hilbig, 2015). When deciding whether to trust someone, loss aversion may impact how an individual weighs the potential gains from honored trust and the potential losses from betrayed trust (Thielmann & Hilbig, 2015). Although researchers agree that loss aversion is an underlying determinant of the decision to trust (Aimone & Houser, 2012; Bohnet, Herrmann, & Zeckhauser, 2010; Bohnet & Meier, 2005), there is a lack of empirical evidence connecting dispositional loss aversion and trust behavior (Thielmann & Hilbig, 2015).

These attitudes each relate to particular risk-taking elements of trust behavior – risk aversion to the associated probabilities of gain versus loss, and loss aversion to the possible positive and negative outcomes (Thielmann & Hilbig, 2015). Either aspect may drive trust behavior more or less strongly, depending on the relative importance one assigns to each in a given trust situation (Thielmann & Hilbig, 2015). Finally, both elements are affected by one's trait fear and anxiety, and thus individual variation regarding decisions to trust may be due to the personality traits that drive risk and loss aversion, anxiety-and fear-related traits (Thielmann & Hilbig, 2015).

1. 1. 3. ***Betrayal Sensitivity***

The last determinant of trust behavior, betrayal sensitivity, influences one's evaluation of a potential loss resulting from trust betrayal, defined as an individual's propensity to ascribe greater severity to a loss due to human selfishness or untrustworthiness than to an equivalent loss caused by nature (Thielmann & Hilbig, 2015). Research shows that people differentiate between risks resulting from a social decision (i.e., another person's betrayal) and those resulting from chance (i.e., nature) (Bohnet et al., 2010; Bohnet & Zeckhauser, 2004; Thielmann & Hilbig, 2015). Betrayal sensitivity impacts an individual's general willingness to trust since one must accept the possibility of a loss resulting from another person's betrayal (Thielmann & Hilbig, 2015). More specifically, individuals who are sensitive to betrayal will, on average, be less willing to trust since they perceive the severity of betrayal more intensely than others (Thielmann & Hilbig, 2015). In addition to the material loss, individuals may feel that a loss resulting from betrayal, rather than nature, inflicts an additional utility cost – a psychological loss – based on the way the loss came about (Bohnet et al., 2010; Thielmann & Hilbig, 2015). This concept is known as betrayal aversion, where an individual shows an aversion to social betrayal because they experience “disutility from the experience, anticipation or observation of non-reciprocated trust” (Aimone, Ball, & King-Casas, 2015, p. 2; Bohnet & Zeckhauser, 2004; Fehr, 2009; discussed further in section 1.4.2. The Trust Game). For example, studies have shown that individuals are, in general, less inclined to take a risk in a social trust situation compared to a nonsocial risk situation (Aimone & Houser, 2012; Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Bohnet & Zeckhauser, 2004; Hong & Bohnet, 2007; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). Finally, Thielmann and Hilbig (2015) suggest that one’s tendency to forgive is the personality trait responsible for individual differences in betrayal sensitivity.

1. 1. 4. ***Summary: Determinants of Trust***

Thielmann and Hilbig’s (2015) model on the determinants of trust – trustworthiness expectations, risk and loss aversion, and betrayal sensitivity – has significantly advanced our understanding of trust and its antecedents. The three components are connected in a manner where some serve as informational inputs for others. Risk aversion is primarily driven by the subjective probabilities of trustworthiness expectations and an individual’s loss aversion is driven by their subjective perception of betrayal, together accounting for how potential outcomes are processed and integrated (Thielmann and Hilbig 2015).

Their model details the different aspects of a trust situation (e.g., trust cues, another’s trustworthiness, the temptation to betray) that could impact trust behavior independent of (or in interaction with) the related personality characteristics (Thielmann & Hilbig, 2015). They specified three personality traits – trustworthiness, anxiety and fear, and forgiveness – as the main sources of interindividual trust variation. Through social projection, trustworthiness expectations are reflected by fairness and honesty traits, trait anxiety and fear explain individual differences in risk and loss aversion, and individual differences in betrayal sensitivity can be explained by trait forgiveness (Thielmann & Hilbig, 2015). Prior models and definitions relied predominantly on the dispositional theory of trust, which considered only one personality trait, trust propensity (Mayer et al., 1995; Schoorman, Mayer, & Davis, 2007; Thielmann & Hilbig, 2015). In addition, the field lacked an understanding of the link between trust and trustworthiness (Thielmann & Hilbig, 2015).

While trust behavior may be mainly driven by expectations, once an individual expects a minimum chance of trust appreciation, other factors such as risk aversion, loss aversion, and betrayal sensitivity may influence their decision to trust (Thielmann and Hilbig 2015). Individuals may be particularly influenced by these aspects when they do not have strong negative or positive expectations about their interaction partner's trustworthiness (Thielmann & Hilbig, 2015). For example, Eckel and Wilson's (2004) experiment increased the "subjective confidence in trustworthiness expectations by providing specific information about an interaction partner (i.e., sex, preferences, appearance, ethnicity)" which they found "eliminated the otherwise observed relation between risk aversion and trust behavior" (Thielmann & Hilbig, 2015, p. 262). Thus, there is a direct correlation between the degree of uncertainty and risk expressed in trustworthiness expectations, such that when they are vague, risk aversion plays a larger role in trust behavior (Thielmann and Hilbig 2015). Finally, trust decisions may not be influenced by all trust determinants equally and universally (Thielmann and Hilbig 2015).

2. ***Trust Motivation***

In their extension of Mayer et al.'s (1995) model of organizational trust, van der Werff et al. (2019) conceptualize the willingness to be vulnerable as a volitional act (model shown in Figures 1 and 2). They argue that motivation is an overlooked aspect involved in the initiation of trust. Specifically, traditional models of trust do not account for the seemingly irrational or habitual range of trust decisions, nor the ongoing changes in trust regulation (van der Werff et al., 2019). They define trust motivation as "an intraindividual psychological state that represents a desire to be vulnerable to another in order to build or maintain a trusting relationship" (van der Werff et al., 2019, p. 101). Their model integrates the existing theory of dispositional (trust propensity) and interpersonal (trustworthiness characteristics) drivers of trust with a process of trust motivation as an additional, direct antecedent of trust (van der Werff et al., 2019). They propose that trust motivation has an indirect impact on the decision to trust based on its influence on trustworthiness perceptions and cognition, in line with Weber (2004; van der Werff et al., 2019). Through a process of trust goal setting, trust is driven by motivational forces derived from the perceptions and experiences of the relationship in question (van der Werff et al., 2019).

Importantly, van der Werff et al.'s (2019) theory of trust motivation accounts for trusting decisions for which rational models would suggest individuals should not cooperate.

Figure 1: Trust Motivation Processes, from van der Werff et al. (2019)

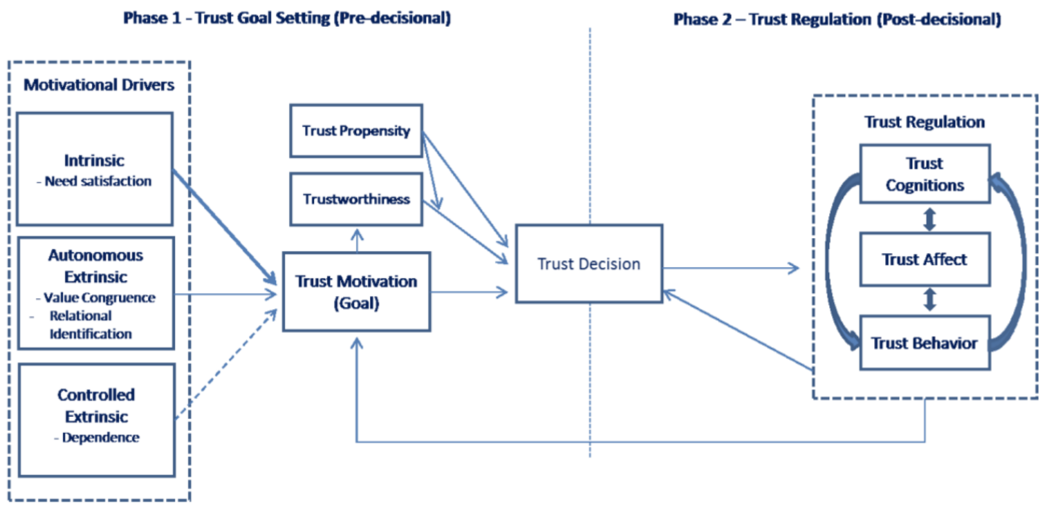
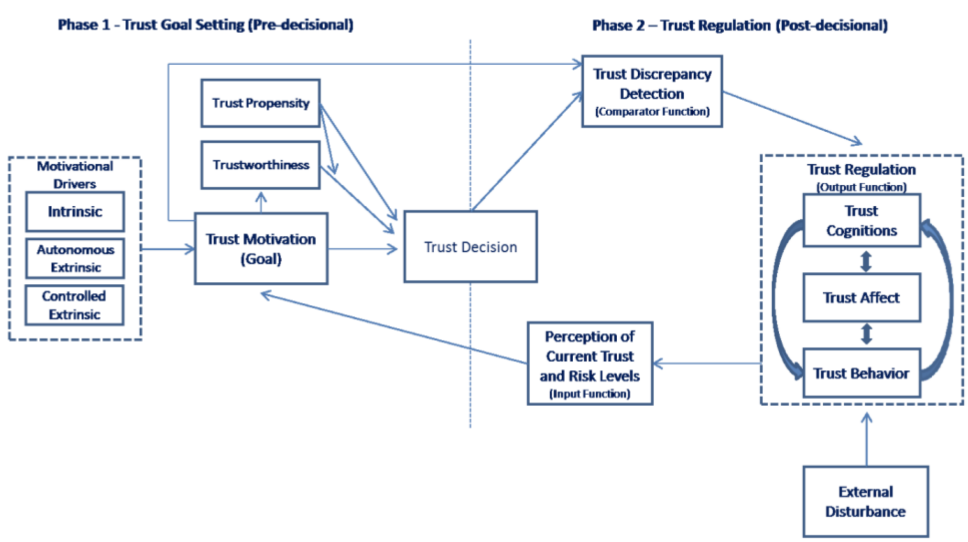


Figure 2: Trust Goal Setting and Regulation Processes, from van der Werff et al. (2019)



3. ***The Neoclassical Theory of Trust***

The standard, neoclassical economic model of decision making depicts a rational, purely self-interested man, "homo economicus." Behavior that deviates from self-interest is viewed as irrational, but this view of rationality fails to predict how people act (Berg et al., 1995). The orthodox neoclassical concept of trust is a prime example of economic theory that does not represent observed behavior. “Trust is an important example of a generalized human tendency towards unselfish behavior” (Evans & Krueger, 2009, p. 1004). However, “neoclassical economic theory considers trust in strangers to be irrational” (Evans & Krueger, 2009, p. 1003). Neoclassical economic theory defines rational individuals as purely self-interested, and trusting and trustworthy behavior would be counter to one’s own self-interest. In a world with only rational actors “there is no reason to take the risk of trusting someone else. In a world of only self-interested individuals, every transaction would need to be enforced by legal contract or even brute force” (Evans & Krueger, 2009, p. 1004). Behavior in experiments conducted by behavioral economists has shown that “trust occurs even when it opposes self-interest” (Evans & Krueger, 2009, p. 1004; Smith, 2003). Research on trust demonstrates that the neoclassical model of rationality does not sufficiently describe human behavior (Evans & Krueger, 2009). Fortunately, behavioral economists have proposed theories of rationality that are more consistent with behavior (Evans & Krueger, 2009).

4. *Game Theory and Experimental Economics*

Game theory is the study of one or more strategic interactions and the ways in which incentives affect the decisions made by the players of the game and the group level outcomes (Houser & McCabe, 2014). Economic interactions are considered strategic when a decision made by one person influences the opportunities and payoffs available to another person, and all parties know this is the case (Houser & McCabe, 2014). Economists investigate the conditions relating to interpersonal trust using game theory in laboratory experiments, because games are designed to reflect important elements of real-life situations (Evans & Krueger, 2009; Rotter, 1967). Laboratory game theory experiments are particularly useful for uncovering critical aspects of the human decision-making process that might be difficult to detect outside of a controlled environment (Houser & McCabe, 2014). For example, behavioral game theory attempts to explain the behavior of individuals in experimental bargaining games.

An important aspect of laboratory game theory experiments is that participants' decisions can be influenced by the experiment design & implementation. For example, the framing of instructions, randomization, anonymity, and salient rewards (real money incentives) are key experimental economics procedures that can have profound impacts on the internal and external validity of an experiment (Houser & McCabe, 2014). Games are represented visually as mathematical models by a game tree, which specifies which player (or players) can move when, what moves can be made at a given stage, what information players possess when they make their move, and how moves (i.e., decisions) made by different players can impact further decisions and determine the outcome(s) of other players (Houser & McCabe, 2014).

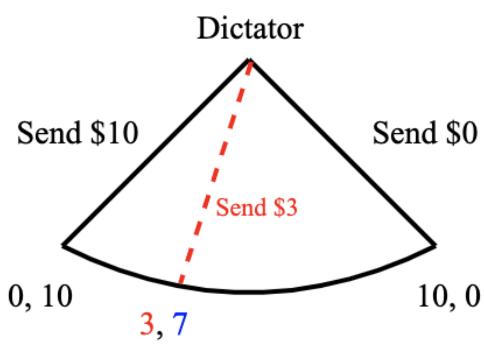
4. 1. **The Dictator Game.**

In a dictator game, one of two anonymously paired individuals is randomly selected to be the *dictator* and allocated a sum of money (Guala & Mittone, 2010). The dictator can divide the money between themselves and the other player, but they are not required to share any of their endowment (Bolton et al., 1998; Fehr & Krajbich, 2014). The paired individual is passive and does not have a decision to make (Fehr & Krajbich, 2014). Rather, as the name implies, the dictator has the final say.

The game is designed to measure positive concern for the other player's material payoff that is independent of the other player's behavior, since the other player has no decision to make (Fehr & Krajbich, 2014). Neoclassical economic theory predicts that the dictator will not share any of their endowment because it is in their self-interest to keep the entire amount for themselves (Bolton et al., 1998). Further, there is nothing compelling them to share. In a striking deviation from selfishness, experiments show approximately 60% of dictators give money to the other player, typically 20-30% of the pie, with some giving away as much as 50% (Alos-Ferrer & Farolfi, 2019; Bolton et al., 1998; Guala & Mittone, 2010). Studies also show that allocations in the dictator game can be influenced by how the game is described (Camerer, 2003), whether the players are anonymous (Eckel & Grossman, 1996), and whether the recipient knows that they are playing a dictator game (Dana, Cain, & Dawes, 2006) (Fehr & Krajbich, 2014).

An example of the dictator game is depicted in Figure 3. The dictator is endowed with $10 and can send any integer amount to the recipient. The dotted line represents one example choice, in which the dictator chooses to send $3 and keep $7 for themselves.

Figure 3: Example Dictator Game, Adapted from Fehr and Krajbich (2014)



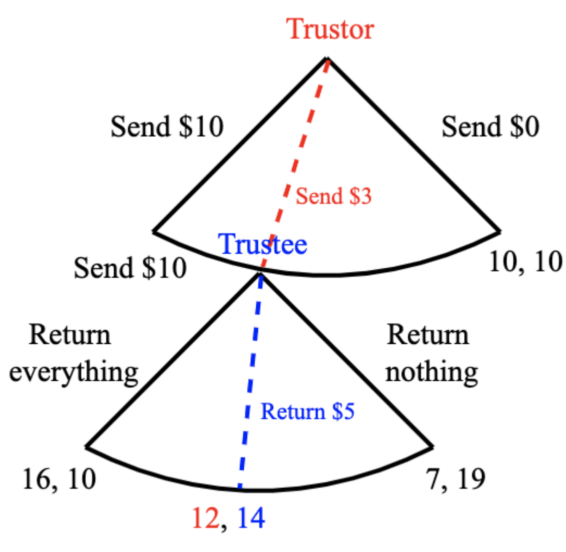
4. 2. **The Trust Game.**

The trust game was introduced into the experimental literature in 1995 by Berg et al. to control for alternative explanations of trusting behavior. The game is unique because the experiment design incorporated trust into game theory by enabling trust to be used for mutual gain (Berg et al., 1995; van Witteloostuijn, 2003). At the time, it was referred to as the investment game, but it is now universally known as the trust game (Evans & Krueger, 2009). The game has become an influential experimental paradigm in psychology, behavioral and experimental economics as an incentivized measure of trust, establishing itself as the standard instrument to measure trust in the laboratory (Alos-Ferrer & Farolfi, 2019; Evans & Krueger, 2011). It is widely used in neuroeconomics experiments and has significantly shaped how economists study trust and reciprocity (Houser & McCabe, 2014).

The Berg et al. (1995) investment game was played as follows. Participants were randomly assigned to one of two rooms, room A and room B, then randomly and anonymously matched to play a one-shot game – one as the investor or FM, and one as the SM. All participants received an initial endowment of $10. The participants in room A, the investors or FMs, were informed that they could send some, all, or none of their $10 to an anonymously paired individual in room B, a SM. They were also told that every dollar sent would be tripled by the experimenter before being transferred to the room B participant. Lastly, the participants in room B observe the amount sent and decide whether to keep the tripled amount they received or return some of it to the same anonymously paired participant. A double-blind procedure implemented with unmarked envelopes passed between 3 rooms and mailboxes guaranteed anonymity and ruled out reputation, collusion, or the threat of punishment as explanations for participants’ behavior.

The game requires the FM to choose between 'trust' – taking a risk with the expectation of earning more than $10 – and the 'status quo,' $10 (Evans & Krueger, 2011). With the status quo, the game ends and both players receive their initial endowments (Evans & Krueger, 2011). If the FM chooses to take the risk, the game enters a second stage in which the SM chooses between reciprocity and betrayal (Evans & Krueger, 2011). The second stage is a dictator game, in which the SMs must decide "how much of the tripled money to keep and how much to send back to their respective counterparts" (Berg et al., 1995, p. 123). It is because of this two-stage nature that the game is regarded as the standard dilemma of trust and reciprocity in behavioral economics (Evans & Krueger, 2009). Figure 4 depicts an example of the game, in which the FM sends $3 and the SM returns $5.

Figure 4: Example Trust Game, Adapted from Fehr and Krajbich (2014)



Berg at al. (1995) found that FMs send about 50% of the endowment on average, and SMs generally return the same amount sent. The most variance was seen in the amounts returned, with 50% of SMs returning $1 or less. Their results presented a strong rejection of the neoclassical prediction that participants would be neither trusting nor trustworthy, choosing not to send any money to their counterparts. Instead, they provided empirical evidence that people are willing to engage in such behaviors when faced with this type of social dilemma. Results from replication experiments have shown that, on average, FMs send about half of their endowment, and SMs send back about as much as the FMs (Camerer, 2003; Fehr & Krajbich, 2014).

The trust game is an example of a non-cooperative game – players are unable to enforce cooperation nor punish non-cooperation, and any cooperation must be self-enforcing (Shor, 2005). It is also an extensive form game, which incorporates non-simultaneous play, allowing researchers to study participants' reactions to their partner's previous actions (Alos-Ferrer & Farolfi, 2019). Within the study of economic games, expected utility (EU) theory assumes that all players have perfect information, at least some individuals are EU-rational, and that an equilibrium, such as Nash, exists (Blume & Easley, 2018). Although the Nash equilibrium assumes perfect information, the trust game represents an economic situation with imperfect information – participants do not know what their partner has chosen (Houser & McCabe, 2014). In addition, research shows participants perceive their behavior as rational, even when it does not fit the economic definition of rationality. Malhotra (2004) found that subjects interpret the behavior of others in the trust game in moral terms, yet they perceive their own behavior as rational. These results suggest that rational behavior is not defined exclusively by utility maximization (Evans & Krueger, 2009).

Experimental evidence demonstrates that players often choose to trust the stranger they have been randomly paired with, contradicting the Nash equilibrium prediction of selfish behavior (Evans & Krueger, 2011; McCabe, Rigdon, & Smith, 2003). Standard non-cooperative game theory fails to describe real-world behavior in the game because it predicts players will reason by backward induction (Evans & Krueger, 2011). This means players must view the payoffs in the final stages of the game and then reason backward to the game's beginning (Evans & Krueger, 2011). Rational choice theory assumes that individuals have self-regarding preferences, predicting the SM will keep all of any tripled amount sent by the FM (Cox, 2004). Knowing this, the FM will choose not to send any money, and instead keep their initial endowment, ending the game after one round (Cox, 2004). In other words, players reasoning by backward induction would conclude that trust is irrational and thus make decisions consistent with the subgame perfect equilibrium (SPE) (Berg et al., 1995; Evans & Krueger 2011). But, due to the tripling of the FM's transfer, the SPE allocation is Pareto-inferior to alternative feasible allocations since participants can be made better off by earning a combined $40 instead of $20 (Berg et al., 1995; Cox, 2004). Both players are better off collectively if more money is transferred, profiting from friendly behavior (Fehr & Krajbich, 2014). Ultimately, mutual cooperation is the socially optimal outcome of the game (Evans & Krueger, 2009). Instead, classic game theory predicts an inefficient equilibrium: mutual defection (Evans & Krueger, 2009). In practice, backward induction is counterintuitive, and the assumptions of expected utility theory prove false, which is why it fails to predict behavior in the game (Evans & Krueger, 2011). Instead, players tend to focus on the payoffs in the first round and reason forward (Evans & Krueger, 2011). This is true even after individuals gain experience in the game (Evans & Krueger, 2011).

The trust game is designed to capture the three critical features of trust: uncertainty and risk, vulnerability, and expectation (Evans & Krueger, 2009; Rousseau et al., 1998; Thielmann & Hilbig, 2015). A FM who chooses to trust is at risk of betrayal and being made worse off had they chosen not to trust (Evans & Krueger, 2011; Hong & Bohnet, 2007). Although reciprocated trust may make the FM better off, they must give up the certain payoff for a potentially higher one, with no guarantee of reciprocity (McCabe et al., 2003). The FM prefers to trust if trust is reciprocated, but not if the SM decides to keep the tripled amount for themselves (Alos-Ferrer & Farolfi, 2019). Of course, receiving the future benefit is dependent upon the SM's behavior (McCabe et al., 2003). Thus, the FM assumes the risk in anticipation of the SM reciprocating as expected (Evans & Krueger, 2009; Mayer et al. 1995). They must decide if the benefit is worth the risk (Evans & Krueger, 2011).

Similarly, Chaudhuri and Gangadharan (2007) suggested that trust can be decomposed into two elements. The first is similar to the concept of vulnerability, where the FM takes a calculated social risk (Alos-Ferrer & Farolfi, 2019; Bohnet & Zeckhauser, 2004; Hong & Bohnet, 2007). The second is related to humans' general social orientation and natural tendency towards prosocial behavior, which can be quantified as the trade-off between the outcome for the self and the outcome for another, i.e., social value orientation (Bohnet & Zeckhauser, 2004; McClintock and Allison, 1989, as cited in Fareri, 2019; Van Lange, 1999, as cited in Fareri, 2019). To study the concept of social risk, Bohnet and Zeckhauser (2004) compared behavior in the trust game to a risky dictator game, in which the dictator chose between a sure payoff and a gamble with uncertain probabilities. Unbeknownst to the participants, the probabilities were set to be the same as those in their trust game results (Bohnet & Zeckhauser, 2004). They wanted to find whether the decision to trust is equivalent to taking a risky bet or if it includes a risk premium to offset the potential cost of betrayal (Bohnet & Zeckhauser, 2004). Bohnet and Zeckhauser (2004) found that participants distinguished between the payoffs of the gamble and the trust game, even though they resulted in the same outcomes. Participants showed an aversion to experiencing betrayal in the trust game that is independent of the non-social component of risk sentiments (Alos-Ferrer & Farolfi, 2019; Bohnet & Zeckhauser, 2004). To describe this social aspect of risk in the game, Bohnet and Zeckhauser (2004) coined the term 'betrayal aversion' (Alos-Ferrer & Farolfi, 2019). The other side of this phenomenon, known as "guilt aversion" (Battigalli and Dufwenberg, 2007, as cited in Fareri, 2019; Dufwenberg and Gneezy, 2000, as cited in Fareri, 2019), occurs when the SM considers the assumed expectations of the FM (Fareri, 2019). These can be understood as second-order beliefs, whereby the SM estimates the likelihood that the FM believes they will reciprocate (Fareri, 2019). Cox et al. (2016) tested this theory as an explanation for the SM's motivation to reciprocate. His experiments revealed that the SM is driven by "vulnerability-responsiveness," where they send money back to avoid feeling guilty about the FM's negative feelings if betrayed.

The psychological situations and choices faced by the two players are notably different. The FM faces a strategic dilemma that requires making assumptions about the trustworthiness of their partner (Evans & Krueger, 2011). While the FM must accept the risk and vulnerability, the SM faces a choice between fairness and selfishness: they can reciprocate and distribute the money equally or betray trust by taking more money for themselves (Evans & Krueger, 2011). Because the SM's payoffs are maximized by betraying trust, reciprocating means choosing to part with some of the money to make the FM better off. Therefore, both players incur an opportunity cost to reach the socially optimal outcome. Ultimately, the outcome of the game depends on the behavior of the SM (Evans & Krueger, 2009).

The economic discipline typically defines trust and trustworthiness in terms of behavior in the trust game (Cox et al., 2016). The consensus within experimental economics is that the FM transfer is a measure of trust, and the SM transfer is a measure of trustworthiness (Smith, 2010). But within this context, numerous definitions have been used. One definition characterizes trust as the FM sending more than the minimum possible amount (Ben-Ner & Putterman, 2009; Berg et al., 1995; Burks et al., 2003). Another definition states that the FM trusts if they send an amount that makes them vulnerable to a loss of utility (Cox, 2004). For example, a FM with self-regarding preferences who sends a positive amount exhibits trusting behavior because they will lose utility if their partner does not return at least the same amount they sent (Cox, 2004). This is evidenced by sending more as the FM in the trust game than as the dictator in a dictator game with the same feasible decision set. The dictator game serves as a control for other-regarding preferences, explained in detail in sections 1.4.3. and 1.4.4. This is because "a trust act is an identification of behavior not motivation for that behavior" (Cox et al., 2016, p. 209). A FM's act of trust may be motivated by altruistic preferences or the expectation of reciprocity. Cox (2004) argues that a distinction should be made between the two to derive an accurate measure of trust. Burks et al. (2003) and Ben-Ner & Putterman (2009) define SM behavior as trustworthy if they return more than the minimum possible amount to their partner. Schotter and Sopher (2006) and Chaudhuri and Gangadharan (2007) use a stronger definition, where the SM must return an amount that exceeds the amount send to be considered trustworthy.

Finally, trust and trustworthiness are different concepts, explained by different elements and motivations within the trust game. For example, the desire for a larger payoff may partially explain the decision to trust, but this motivation directly contradicts and would not lead to trustworthy behavior. (Alos-Ferrer & Farolfi, 2019). Another difference between trust and trustworthy behavior discovered in the lab is that their prevalence is not symmetric across the same individual. Chaudhuri and Gangadharan (2007) found evidence suggesting trustworthy participants are more trusting, but not vice versa. Thus, the same considerations and motivations cannot explain why a FM chooses to be vulnerable and a SM chooses not to be selfish.

4. 2. 1. ***Trust and Reciprocity Hypothesis.***

Findings from experiments where participants play the dictator and trust game show that players send more money in the trust game than the dictator game (Cox, 2004; McCabe et al., 2003). The FM's decision to trust may create goodwill between the players, leading to reciprocity (Evans & Krueger, 2009). McCabe et al. (2003) proposed the Trust and Reciprocity (TR) Hypothesis to explain the observed behaviors of trust and reciprocity in the game. It describes players as being in a "reciprocal-trust relationship" if the FM takes a risk by trusting their partner, the SM forfeits something to reciprocate their partner's trust, and there are mutual gains from their decisions (McCabe et al., 2003). The increased gain from the exchange is measured against the subgame perfect equilibrium since players in a reciprocal-trust relationship will realize a Pareto superior outcome relative to the one prescribed by non-cooperative game theory (McCabe et al., 2003). The mutual gain from cooperation can only be realized if neither individual plays their dominant strategy, i.e., they both deviate from the subgame perfect strategy (McCabe et al., 2003).

The TR hypothesis asserts that the SM must consider the motives of and be able to attribute intention to the FM's behavior, i.e., they must possess Theory of Mind (Camerer & Hare, 2014; McCabe et al., 2003). This suggests that the SM can read the action of the FM as signaling trust that the SM will reciprocate if given the chance (McCabe et al., 2003). Thus, it follows that the SM understands that the FM incurs an opportunity cost (McCabe et al., 2003). Further, the SM's behavior can be characterized as reciprocal only if they interpret the FM's actions as trusting (McCabe et al., 2003).

McCabe et al. (2003) introduced the concept of positive reciprocity to explain the SM's tendency to reciprocate. They define positive reciprocity as "the costly behavior of a SM that rewards a FM based on both the gains from exchange to the SM as well as the SM's beliefs about the intentions motivating the action of the first mover" (McCabe et al., 2003, p. 269). Cox (2004) offers a similar definition, which emphasizes the SM's motivation. He defines positive reciprocity as the motivation to repay a generous or helpful action by taking actions that are generous or helpful towards the other person (Cox, 2004). Positive reciprocity is conditional kindness and should be distinguished from unconditional kindness that may be motivated by altruism (Cox, 2004). A SM's return transfer can be explained by positive reciprocity because their generous action may be taken in response to the FM's generous action (Cox, 2004).

4. 2. 2. ***Cost, Benefit, and Temptation Model.***

Evans and Krueger (2011) model the payoff structure of the trust game along a number line, depicted in Figure 5. It is composed of three distinct elements: cost, benefit, and temptation (Evans & Krueger, 2011). Initial game endowments are represented by the letter P. The amount transferred by the FM (to the SM) is labeled S. The difference between the two, P-S, represents the FM's potential cost. The tripled amount received by the SM is labeled T. R represents the amount returned to the FM by the SM. The potential benefit to the FM is the difference between the amount returned by the SM and the initial endowment, R-P. The SM's temptation to defect is the difference between the tripled amount the SM receives and the amount the SM sends back (returns) to the FM, T-R. "Empirically, temptation is the best predictor of reciprocity among strangers" (Snijders & Keren, 1999, as cited in Evans & Krueger, 2011, p. 171). The two of the three critical elements of trust, vulnerability and expectation, are represented in the model (Evans & Krueger, 2009; Thielmann & Hilbig, 2015). The concept of vulnerability or risk is represented by the cost and benefit, while the element of temptation is relevant to the concept of expectation

(Evans & Krueger, 2011). If the FM believes the SM will be trustworthy and return more than the FM sent, they may take the risk for the chance of a larger payoff (Alos-Ferrer & Farolfi, 2019; Cox, 2004; Evans & Krueger, 2011).

Figure 5: Cost, Benefit, and Temptation in the Trust Game, from Evans and Krueger (2011)

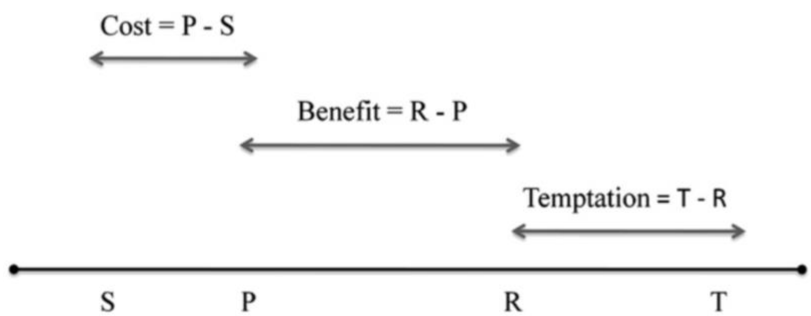


Fig. 2. Cost, benefit, and temptation in the trust game. By definition, S<P<R<T. Cost and benefit are factors related to the trustor's payoffs. Temptation describes the trustee's payoffs. Temptation is critical to the trustor because it implies the probability of reciprocity.

4. 2. 3. ***Trust World and The Paradox of Trust.***

Isoni & Sugden (2019) designed a simple dictator and trust game to describe *Trust World*, where the outcome of the game is mutually beneficial trust and trustworthiness. This is the scenario where the FM is fully trusting, and the SM is fully trustworthy, resulting in the FM choosing *send* and the SM choosing *return* in the trust game tree depicted in Figure 6 (Isoni & Sugden, 2019). They designed their game with the dictator's choice set identical to the SM's to analyze whether the SM's behavior is contingent on the FM choosing *send.* They do this to highlight a problem with Rabin's (1993) model of reciprocal kindness — it does not explain the outcome of mutually beneficial trust. Rabin's (1993) model of "fairness" was the starting point for the study of preferences for reciprocity and defines an individual's "kindness" or "unkindness" in relation to the consequences of their actions for another (Isoni & Sugden, 2019). But Isoni & Sugden (2019) argue that Trust World describes "the workings of some kind of reciprocity," which cannot be explained by Rabin's (1993) model (p. 221). They call this property of Rabin's (1993) model the Paradox of Trust. "The Paradox of Trust is that expecting to benefit from one's own act of kindness can undermine the kindness of the act" (Isoni & Sugden, 2019, p. 225). Although the relationship between trust and trustworthiness is typically described in terms of reciprocity, players are expecting to benefit from their acts of cooperation (Isoni & Sugden, 2019). The paradox is that in a simple trust game there is not an equilibrium in which FM trust and SM trustworthiness are mutually beneficial (Isoni & Sugden, 2019). In the absence of a trust equilibrium reflecting cooperation, the model of reciprocal kindness falls victim to the Paradox of Trust (Isoni & Sugden, 2019; van Witteloostuijn, 2003). Thus, mutually beneficial trust and trustworthiness does not arise from reciprocal kindness (Isoni & Sugden, 2019).

Figure 6: The Trust and Dictator Games in Isoni and Sugden (2019).

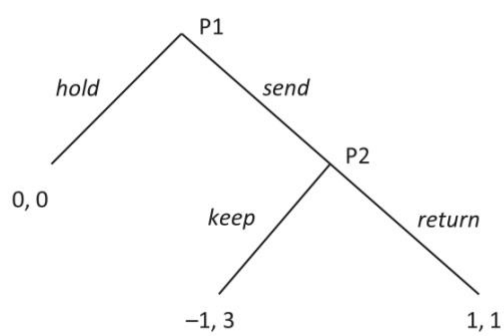


Fig. 1. The Trust Game (G_1).

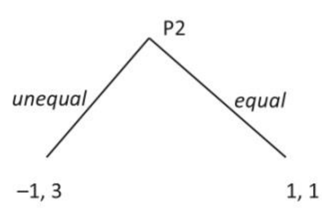


Fig. 2. P2's Dictator Game (G_1').

To resolve this paradox, Isoni and Sugden (2019) distinguish between two forms of reciprocity, Rabin's (1993) concept of reciprocal kindness and their theory of reciprocal cooperation. They argue that the relationship between the FM's trust and SM's trustworthiness is better understood as reciprocal cooperation, defined as "the reciprocity of playing one's part in a cooperative practice when one believes that the other party to that joint action will play (or has played) hers" (Isoni & Sugden, 2019, p. 225). Acts of trust and trustworthiness should be understood as complementary components of cooperative actions in which individuals are motivated to play their part, conditional on the other person playing theirs (Isoni & Sugden, 2019). If the FM chooses to trust, they must be thinking of themselves as playing their part in a joint action, the other half of which is the SM reciprocating. Similarly, by returning money the SM does not think of themselves as rewarding the FM for their gratuitous kindness (Isoni & Sugden, 2019). The SM's behavior is contrary to their self-interest, but given the FM's behavior, the SM returns money to fulfill their half of the combination of actions that benefits them both (Isoni & Sugden, 2019). Thus, the SM's motivation should not be characterized as rewarding their partner's meritorious behavior (Isoni & Sugden, 2019). The SM is simply playing their part in the cooperative behavior that the FM initiated (Isoni & Sugden, 2019).

Isoni and Sugden (2019) argue that while the relationship between the players actions is defined as reciprocity, it is not based on kindness, i.e., it is not reciprocal kindness. Since both players are motivated to cooperate because they are expecting a benefit from cooperation, behavior exhibited by the Paradox of Trust should be characterized as reciprocal cooperation (Isoni & Sugden, 2019). A theory of reciprocal cooperation does not fall prey to The Paradox of Trust since "expecting to benefit from an act of cooperation does not undermine the cooperativeness of the act. To the contrary: the whole point of cooperation is that both parties benefit" (Isoni & Sugden, 2019, p. 225). Further, Isoni and Sugden (2019) suggest that in Berg et al.'s (1995) description of their experiment as a study of reciprocity, they were conjuring a scenario of reciprocal cooperation. Finally, Isoni and Sugden's (2019) theory of reciprocal cooperation is consistent with the Trust World scenario of the game, and thus able to explain mutually beneficial trust and trustworthy behavior (Isoni & Sugden, 2019).

4. 3. **Social Preferences.**

Until recently, experimental game theorists neglected to incorporate utility theory into their models and instead assumed individuals' utilities were "affine transformations of (only) their own monetary payoffs in the games" (Cox, 2004, p. 260). Classic game theory assumed that individuals do not care about another's (relative or absolute) material payoffs or intentions (Cox, 2004). But results of public goods and trust and reciprocity experiments consistently conflicted with the prediction of a subgame perfect equilibria made by the theory of self-regarding preferences (Cox, 2004).

A theory of social preferences was developed to explain observed behavior in laboratory games and incorporate preferences that could be deemed social or other-regarding. In one-shot games, an individual displays social preferences if they are "willing to forgo their own material payoff for the sake of increasing or decreasing another individual's material payoff" (Fehr & Krajbich, 2014, p. 196). In other words, an individual cares enough about helping or harming another to bear the associated cost. Social preference theories assume that people are self-interested, but they are also concerned about and can derive utility from the payoffs of others (Charness & Rabin, 2002; Cox, 2004). Theories of social preferences posit that individuals' utility functions depend not only on their own material payoff, but also on non-monetary payoff elements such as altruism, concerns for fairness, reciprocity, equality (inequality-averse), and efficiency (Cox, 2004; Fehr & Krajbich, 2014), i.e., at least some people regard the payoffs of others as positive utilities (Evans & Krueger, 2009).

With regards to the trust game, both an altruistic and efficiency motivated FM may decide to transfer some money without expecting a return (Alos-Ferrer & Farolfi, 2019; Bolton and Ockenfels, 2000, as cited in Alos-Ferrer & Farolfi, 2019; Charness and Rabin, 2002). One reason for this could be benevolent preferences, in which an individual derives utility from being altruistic — they feel good when they choose to trust, cooperate, and reciprocate (Andreoni & Miller, 2002, as cited in Cox, 2004; Cox et al., 2002, as cited in Cox, 2004; Evans & Krueger, 2009). Due to the tripling of the amount transferred, this behavior could also be explained by preferences for efficiency since the amount sacrificed by the FM is far less than the amount the SM will receive (Alos-Ferrer & Farolfi, 2019). In either case, participants consider self-interest and the satisfaction of costly other-regarding behavior (Evans & Krueger, 2009).

Cox (2004) points out that the SM may be motivated by reciprocity, i.e., conditional self-regarding preferences, or unconditional other-regarding (social) preferences. Conditional self-regarding preferences means a player sends money in the trust game because they expect a greater payoff based on reciprocity and cooperation (Cox, 2004). Cox (2004) found that return transfers in the trust game are approximately one-third higher than those in the dictator game, concluding that return transfers in the trust game are partly motivated by conditional other-regarding preferences. Further, since a selfish FM is solely motivated by the belief that the SM will reciprocate, a FM with self-regarding preferences may trust the SM because they believe the SM has other-regarding preferences (Alos-Ferrer & Farolfi, 2019; Cox, 2004). On the other hand, a FM with unconditional other-regarding preferences transfers money without expecting the SM to reciprocate (Cox, 2004).

Another example of social preferences includes inequality aversion, wherein an individual prefers their payoff to be neither lower nor higher than the payoff of another (Bolton & Ockenfels, 2000, as cited in Evans & Krueger, 2009; Fehr & Schmitt, 1999, as cited in Evans & Krueger, 2009). Other theories emphasize fairness and reciprocity, such that individuals prefer to punish others who hurt them and help those who are kind (Evans & Krueger, 2009; Rabin, 1993). In addition, there is empirical evidence demonstrating that people often prefer equality (Evans & Krueger, 2009).

The majority of individuals display social preferences – prosocial behavior is consistently demonstrated in empirical studies (Evans & Krueger, 2009; Fehr & Krajbich, 2014). Thus, it is safe to say that most people care about others' material payoff or well-being and take the welfare of other individuals into account (Fehr & Krajbich, 2014). Social preferences are also considered an important component of individuals' utility functions (Andreoni & Miller, 2002, as cited in Fehr & Krajbich, 2014). Research shows that they "play a decisive role for aggregate social and economic outcomes in strategic settings" (Fehr & Krajbich, 2014, p. 193). Also, there is evidence of considerable individual heterogeneity in social preferences: some people display strong, positive social preferences, while others display little or no concern for the other person (Fehr & Krajbich, 2014).

The research stimulated by social preferences and norm-based theories has enabled behavioral economics to overcome the limitations of the neoclassical, self-interest hypothesis (Evans & Krueger, 2009). "Trust and reciprocity are not necessarily irrational when they are understood in these terms" (Evans & Krueger, 2009, p. 1009). Berg et al. (1995) noted that participants' decision to reciprocate in the trust game may be influenced by social norms. Now, social norms have been incorporated into utility theory by introducing the possibility that individuals' preferences over outcomes depend on the observed behavior of another person (Cox, 2004). Finally, the experimental data demonstrate that other-regarding behavior often leads to socially desirable results (Evans & Krueger, 2009).

4. 4. **Measuring Trust with Construct Validity.**

A core issue in the experimental design of the trust game is whether the game design produces a measure of trust with construct validity. Cox (2004) first emphasized the importance of distinguishing between decisions motivated by reciprocity and behavior motivated by other-regarding preferences, which are not conditional on the intentions or behavior of another, "because they have different implications for game-theoretic modeling" (p. 262). He explained that, by itself, the trust game does not support the conclusion that participants' behavior is characterized by trust (Cox, 2004). A positive transfer by the FM cannot be validly interpreted as evidence of trust, because although it reveals that the individual has exposed themselves to a risk of monetary loss, it does not reveal a risk to a loss of utility (Cox, 2004). Because the amount transferred in the trust game is tripled, a FM with altruistic preferences may prefer to transfer some money without expecting any amount transferred back, since the amount sacrificed by the FM is far less than the amount the SM will receive (Alos-Ferrer & Farolfi, 2019; Cox, 2004). On the other hand, if a FM sends the same amount to their counterpart in a dictator game (in which their partner cannot make a return transfer) as they do in the trust game, then the FM's choice could be attributed to altruism. In this example, the FM did not risk a loss of utility from the positive amount transferred in the trust game. Unless it is known that the FM has self-regarding preferences, the mere act of sending money cannot be interpreted as evidence of trusting behavior (Cox, 2004). This is due to the presence of motivational confounds in the form of other-regarding preferences (Alos-Ferrer & Farolfi, 2019; Cox, 2004).

Mathematically, a single trust game is insufficient to discriminate between monetary transfers motivated by trust (or reciprocity) and those motivated by other-regarding preferences (Cox, 2004). To validly measure trust, a second dictator game with the same decision set as the FM trust game should serve as a control (Cox, 2004). Based on Cox (2004), trust should be measured mathematically as the difference between the amount sent by the FM in the trust game and the amount the same individual sent in the (control) dictator game (Alos-Ferrer and Farolfi, 2019). Together, the games distinguish trust from observationally similar behavior characterized by other-regarding preferences, such as altruism, that do not expose individuals to risk of loss (Cox, 2004). Only if the amount of money that a participant sends as FM in the trust game is greater than the amount sent as the dictator in the dictator game can one conclude that the individual has exhibited trust (Cox, 2004).

One confirmed mechanism of trust is oxytocin (OT), a neuropeptide associated with trustworthy behaviors (Alos-Ferrer and Farolfi, 2019; Zak, 2011; Zak, Kurzban, & Matzner, 2005). Zak, Kurzban, & Matzner (2005) measured the change in of OT levels after participants made decisions in a trust game and found increases in OT were correlated with participants' trustworthiness, measured as reciprocity. Their research did not find evidence that OT predicted FM transfers in the game, suggesting changes in OT are related to trustworthy SM behavior. Confirming these findings, Kosfeld et al. (2005) found that intranasal OT administered to participants before playing the trust game resulted in FMs transferring more money, but there was no difference in reciprocity between the control and OT group since high trust in this study likely caused SMs to experience the endogenous release of OT. Thus, the brain's release of OT captures the neurophysiologic effect of a positive social experience and motivates reciprocation (Zak, 2012). In the Kosfeld et al (2005) study, exogenous OT gave FMs the physiologic effect of a positive social interaction thus influencing their behavior while SMs, who mostly received the positive effect of a transfer from FMs, did not need additional (exogenous) OT to cooperate. Researchers have found OT influences behavior by increasing individuals' empathy and reducing their betrayal aversion in social interactions (Alos-Ferrer and Farolfi, 2019; Kosfeld et al., 2005; Zak, 2011).

4. 5. **The 'Communication Effect'**

In most studies in which participants face an economic social dilemma, communication is not permitted prior to decision making (Lev-On et al., 2010). Further, few studies on the trust game allow participant pairs to communicate face-to-face before making their decisions (Ben-Ner et al., 2011; Johnson and Mislin, 2011; Lev-On et al., 2010; Zak et al. 2022). However, this lack of communication greatly limits the generalizability of these studies to any out-of-lab interactions in which people interact face-to-face (Zak et al. 2022). Enabling partners to speak to each other before making decisions in the trust game has been shown to increase cooperation, likely due to promise-keeping social norms, the perceived rules that prescribe appropriate behavior (Ben-Ner et al., 2011; Bicchieri, 2002; Lev-On et al., 2010).

Bicchieri (2002) coined the term 'communication effect' to describe the finding that "in one-shot social dilemma experiments, cooperation rates dramatically increase if participants are allowed to communicate before making a choice" (p. 192). Studies show that without prior communication and promise-making, cooperation in one-shot games is about 50% (Bicchieri, 2002; Sally, 1995). Sally's (1995) meta-analysis of social dilemma experiments run between

1958 and 1992 revealed that promise-making increased cooperation by 30% and communication increased cooperation by 40%. If the primary content of the discussion is making commitments to solve the dilemma, allowing participants to discuss the dilemma prior to decision making increases cooperation rates (Mackie, 1997, as cited in Bicchieri, 2002). "The cooperation-enhancing effect of communication can be to a large extent attributed to the opportunity of coordinating behavior [during] communication" (Brosig-Koch et al., 2003, p. 2). Individuals may perceive cooperation as less risky because they are able to judge the trustworthiness of their partner through behavioral signals such as facial expression, eye gaze, and body language (discussed in 2.1.1.1. Determinants of Trust) (Behrens, Snijdewint, et al., 2020; Bicchieri, 2002).

The prevailing explanation for the high rates of cooperation observed after allowing participants to communicate prior to decision making is that promise-making elicits social norms, presumably through the availability heuristic (Bicchieri, 2002). "The effect of discussion on cooperation rates might precisely be due to the fact that discussing the dilemma often involves an exchange of pledges and promises, and the very act of promising focuses subjects on a norm of promise-keeping," evoking both a descriptive and an injunctive norm (Bicchieri, 2002, p. 198). Discussing the dilemma, and in turn making commitments, initiates the descriptive norm of promise-keeping (Bicchieri, 2002). Typically, we keep promises when we make them, leading participants to deduce that the appropriate course of action is to keep their promise (Bicchieri, 2002). This descriptive norm may also be perceived as injunctive — keeping a promise is 'the right thing to do' (Bicchieri, 2002). Additional evidence in favor of promise-keeping norms as the best explanation of behavior come from experiments in which participants interact and make promises with both a human and computer. Participants make and follow through on commitments to computers, demonstrating the power of salient promise-keeping norms.

Communicating prior to decision-making has the potential to introduce bias. Participants' impressions may be negatively influenced by stereotypical and prejudiced beliefs, and unrestricted face-to-face communication may influence their judgments of others and decisions regarding trust beyond their conscious knowledge (Allport, 1954, as cited in Lick & Johnson, 2014; Bicchieri, Lev-On, Chavez, 2010; Fareri, 2019). Biases may include those based on gender, race, ethnicity, tattoos, apparel, attractiveness, and any other characteristics that may trigger stereotypes (Kurzban, Tooby, & Cosmides, 2001, as cited in Zak et al., 2022). In at least one trust game experiment, the researchers forbade participants from sharing personally identifying information, such as race and gender, to limit their potential impacts on decisions (Lev-On et al., 2010). It is possible that negative effects are negated by the power of promise-making social norms, since studies show that participants cooperate most frequently when communication is face-to-face and unrestricted (Bicchieri et al., 2010; Brosig-Koch et al., 2003; Dawes, McTavish, & Shaklee, 1977, as cited in Boone & Buck, 2003).

2. **Physiological Arousal & Synchrony**

1. ***Physiology and the Autonomic Nervous System (ANS)***

Physiological arousal is our bodies subconscious response to some stimuli as expressed through our physiological responses (Appelhans & Luecken, 2006). The Autonomic Nervous System (ANS) is responsible for the generation of physiological arousal by coordinating and managing a complex, highly differentiated network of nerves, organs, and biological sensors distributed throughout the body (Appelhans & Luecken, 2006; Levenson, 2014). "Functionally, the ANS plays a number of different roles, serving as regulator, activator, coordinator, and communicator" (Levenson, 2014, p. 101). It regulates vital bodily functions including heart rate, respiration, and digestion, and internal processes such as the cardiac, respiratory, and glandular systems (Palumbo et al., 2017). The ANS is also the primary mechanism of our fight-flight-or-freeze response. It is composed of two systems, the excitatory sympathetic nervous system (SNS) and the inhibitory parasympathetic nervous system (PNS). Together, they dynamically regulate physiological arousal. The SNS is activated during physical or psychological stress, and is characterized by an increase in pulse, sweat secretion, and alertness. During periods of SNS activation, the body chooses to divert energy away from secondary processes such as digestion. The immediate threat requires our full attention. The PNS is most active in non-stressful situations. The effects of the PNS include a slowed and steady heart rate, pupil constriction, and increased digestion (Marieb, 2002). Both systems work in tandem, dynamically changing as they regulate the body in preparation for and response to endogenous and exogenous environmental conditions (Palumbo et al., 2017).

There are a variety of techniques and measures used to study the complex interaction between the SNS and PNS. For example, the human skin displays many forms of bioelectric phenomena, particularly in the hands, fingers, and soles of the feet (Pflanzer & McMullen, 2012). The skin's ability to conduct electricity is known as electrodermal activity (EDA). This specifically refers to the changes in the electrical conductance of the skin, which depends on sweat secretion. EDA "is an indirect measure of eccrine sweat glands, which are uniquely innervated by the SNS" (Palumbo et al., 2017). During arousal the excitatory SNS is active, causing subtle changes in sweat production, which can be measured and analyzed in a laboratory setting. The most commonly studied property of EDA is SC, which is quantified "by applying an electrical potential between two points of skin contact and measuring the resulting current flow between them" (Braithwaite, Watson, Jones, & Rowe, 2015). Increased sweat gland activity in response to SNS activation results in an increase in SC.

Because EDA has been closely linked to implicit emotional responses and cognitive processing that may occur without conscious awareness, it is widely used as an objective index of emotional processing and sympathetic activity (Braithwaite et al., 2015). For example, changes in one's SCL reflect general changes in autonomic arousal (Braithwaite et al., 2015). According to Braithwaite et al. (2015), it is "the most useful index of changes in sympathetic arousal that are tractable to emotional and cognitive states as it is the only autonomic psychophysiological variable that is not contaminated by parasympathetic activity" (p. 3). It is difficult to overstate the importance of ANS measures to the study of psychology and psychological research; "psychophysiological findings have contributed to almost every aspect of psychology, including cognition, emotion, and behavior" (Palumbo et al., 2017, p. 100)

2. ***Physiological Response to Decision Making***

Bechara et al. (1997) were the first to examine physiological response during decision making and uncovered that individuals generate skin conductance responses (SCRs) in anticipation of a large loss. Their study compared healthy individuals and those with ventromedial prefrontal cortex (vmPFC) and decision-making defects as they performed a gambling task in which participants are asked to draw cards from four decks, two "risky" decks in which repeated drawings lead to an overall loss, and two "safe" decks in which repeated drawings lead to an overall gain. After encountering a few losses, participants without prefrontal cortex damage began to generate anticipatory SCRs before selecting a card from the "risky" decks and began avoiding these decks (Bechara et al., 1997). Those with vmPFC damage did neither. To determine whether participants “choose correctly only after or before conceptualizing the nature of the game and reasoning over the pertinent knowledge” the researchers briefly interrupted the game to ask questions that would allow them to assess the participants’ “account of how they conceptualized the game and of the strategy they were using” (Bechara et al. 1997, p. 1293).

They found that although the individuals with vmPFC damage were able to accurately conceptualize the game and the correct strategy, they failed to generate autonomic responses and continued to select cards from the bad decks. The individuals with frontal lobe damage failed to generate any SCRs whatsoever (Damasio, 1994). In other words, they failed to make decisions based on their accurate conceptual knowledge. Damasio (1994) described this as if the individuals’ “entire scope of knowledge is available except for the dispositional knowledge paring a particular fact with the mechanism to reenact an emotional response” (1994, p. 211).

To explain this novel finding, Bechara et al. (1997) proposed a “sensory representation of a situation that requires a decision” (Figure 7) and showed how it “leads to two largely parallel but interacting chains of events” (1997, p. 1294). When an individual is presented with a situation, neural systems are activated that “hold nondeclarative [sic] dispositional knowledge related to the individual’s previous emotional experience of similar situations” (Bechara et al., 1997, p. 1294). The ventromedial frontal cortices are among the brain structures that store dispositional knowledge, the activation of which activates the autonomic nervous system and neurotransmitter production (Bechara et al., 1997). These non-conscious signals act as “covert biases on the circuits that support processes of cognitive evaluation and reasoning” (Bechara et al., 1997, p. 1294). In the second chain of events, a situation triggers the overt recall of pertinent facts, e.g., possible responses and potential outcomes given a particular course of action, and the application of reasoning strategies to facts and options. Participants with vmPFC damage also complete this chain of events while making decisions during the gambling task.

Figure 7: The Proposed Steps Involved in Decision-Making, from Bechara et al. (1997)

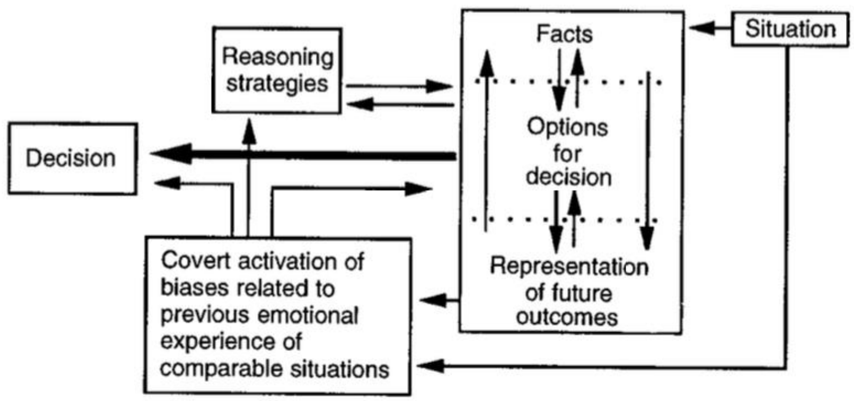


Fig. 2. Diagram of the proposed steps involved in decision-making.

Bechara et al.'s (1997) experiment demonstrates that brains of healthy individuals activate covert biases that precede overt reasoning on the available facts, and that these biases may assist the reasoning process by facilitating "the efficient processing of knowledge and logic necessary for conscious decisions" (p. 1294). The researchers theorized that their results are evidence of "a complex process of nonconscious signaling, which reflects access to records of previous individual experience — specifically, of records shaped by reward, punishment, and the emotional state that attends them" (Bechara et al., 1997, p. 1294). Conversely, individuals with damaged vmPFCs are prevented from accessing these particular kinds of previous and related experiences; they are unable to produce nonconscious signals and covert biases that precede overt reasoning.

Bechara et al.'s (1997) theory aligns with Damasio's (1994) Somatic Marker Hypothesis, which suggests that these bodily signals are an anticipatory arousal response corresponding to the value of the choice (Glimcher & Fehr, 2013). When faced with a choice "either the sensory representation of the situation or of the facts evoked by it activate neural systems that hold nondeclarative dispositional knowledge related to the individual's previous emotional experience of similar situations" (Bechara et al., 1997, p. 1294). The sensation that arises can be described as a *somatic marker* because the experience has 'marked' an image – a visceral and non-visceral sensation – on the body (Damasio, 1994). "Somatic markers depend on learning within a system that can connect certain categories of entity or event with the enactment of a body state, pleasant or unpleasant" (Damasio, 1994, p. 180). It encompasses an integral change of body state, including modifications in the viscera and musculoskeletal system, which are induced by both neural signals and chemical changes (Damasio, 1994). Further, these bodily signals serve to deter participants away from "risky" choices (Glimcher & Fehr, 2013). The sensation may lead you to immediately reject the negative course of action. "The automated signal protects you against future losses." (Damasio, 1994, p. 173). Damasio (1994) argued that somatic markers increase the accuracy and efficiency of the decision process. After this automated step reduces the number of options, individuals can utilize cost/benefit analysis and deductive reasoning.

3. ***Mechanisms of Physiological Synchrony***

The study of PS began in the 1950s with the study of patient-therapist relationships and analyses of physiological responses of therapists and their clients during interviews (Coleman, Greenblatt, & Solomon, 1956, as cited in Palumbo et al., 2017; Di Mascio, Boyd, & Greenblatt, 1957, as cited in Palumbo et al., 2017; Di Mascio et al., 1955, as cited in Palumbo et al., 2017). Several studies "found significant positive and negative correlations in the EDA and HR of therapists and clients during therapy, interpreted as evidence of therapeutic rapport and empathy (Coleman, Greenblatt, & Solomon, 1956; Di Mascio, Boyd, & Greenblatt, 1957; Di Mascio et al., 1955)" (Palumbo et al., 2017, p. 110). Recent studies have found that "clients' reports of therapist empathy were positively correlated with the magnitude of PS" (Palumbo et al., 2017, p. 125). Clients with therapist-patient relationships displaying high PS reported feeling more understood. Further, contextual data suggested that sessions with higher levels of PS were experience as more empathetic, which the researchers concluded indicate there exists a physiological component of empathy (Palumbo, 2015). Conversely, Levenson and Gottman (1983) found that the autonomic response signals of couples engaged in a heated argument were significantly more synchronized than those of non-distressed couples. This shows that PS is not only present during positive social interactions; it can occur during negative interactions as well (Danyluck & Page-Gould, 2018). PS is best interpreted by combining self-reports with behavioral measures to triangulate the interpersonal meaning of this internally shared state (Palumbo et al., 2017).

From the early studies, researchers have considered the possibility that experiential connections that define emotional empathy are reflected in physiology (Palumbo et al., 2017). As discussed in 2.1.4.4. Measuring Trust with Construct Validity, one known mechanism of trust is the neuropeptide hormone oxytocin (OT) (Kosfeld et al., 2005; Zak, 2011; Zak, Kurzban, & Matzner, 2005). Zak, Kurzban, & Matzner (2005). Research on OT and the trust game have found that increases in OT are associated with participants' subjective experience of empathy (Zak, 2011). The pioneering research found that natural increases in OT were correlated with participants' trustworthiness, measured as reciprocity as the SM (Zak, Kurzban, & Matzner 2005). In addition, participant administration of intranasal OT resulted in increased FM transfers in the game (Kosfeld et al., 2005). Further experiments found that participants administered intranasal OT who experienced previous betrayals in the game did not reduce their FM transfer amounts (trust) in subsequent rounds (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr 2008). Together, the research suggests that OT influences behavior by increasing individuals' empathy and reducing their betrayal aversion (Alos-Ferrer and Farolfi, 2019; Zak, 2011).

Contemporary research on PS has provided evidence of several psychosocial constructs, including empathy, attachment, conflict, and emotional co-regulation, with empathy and attention considered the most common explanations (Palumbo et al., 2017). This suggests that the autonomic nervous system "reflects a component of shared experience, so PS may be an objective measure of internal processes accompanying an empathic interaction" (Palumbo et al., 2017, p. 123). Although PS has been shown to be correlated with emotional empathy, other findings suggest these constructs are independent (Palumbo et al., 2017).

In general, it is unclear how one individual's autonomic activity resembles or synchronizes with another individual or individuals (Palumbo et al., 2017). While there is evidence that PS is not dependent on environment or behavior, it is unclear how one individual's autonomic activity resembles or synchronizes with another individual or individuals (Palumbo et al., 2017). Researchers have found that the presence of PS between individuals "is not dependent on (a) shared conditions such as behavior or environment, (b) a specific sensory mode of communication, or (c) psychosocial conditions such as valence or relationship type" (Palumbo et al., 2017, p. 124). It is unknown the degree to which "given variables are driving interpersonal physiological interactions during a given condition, as multiple modalities may be simultaneously involved" (Palumbo et al., 2017, p. 124). One possibility is that these complex interactions develop through different mechanisms, including shared environment, coordinated behaviors, and interpersonal processes, however, more research is needed to determine the variables and conditions that contribute to PS (Palumbo et al., 2017).

4. ***Physiological Synchrony in Economic Games***

Few studies have been conducted examining PS between partners in economic games. To my knowledge, only two papers studying PS in the context of economic games have been published to date. Mitkidis et al. (2015) used a public goods game and Behrens and Snijdewint, et al. (2020) a prisoner's dilemma game, thus PS in the trust game remains unexplored prior to this study. The following is a summary of their findings.

Behrens and Snijdewint, et al. (2020) had participants play a modified iterated prisoner's dilemma games in a dyadic interaction setting, with participants able to see each other in the treatment group and not in the control group, to verify the connection between synchrony and cooperation. They measured HR and EDA during game play and found that participants exhibited higher PS when they could see each other, and thus exchange nonverbal signals, before making their game decisions, as measured by windowed time lagged cross-correlation analysis (Behrens, Snijdewint, et al., 2020; Boker, Xu, Rotondo, & King, 2002). They also ran a multilevel linear regression and found that only SCL synchrony, but not HR synchrony, predicted cooperative success in the game. They applied/used Boker et al.'s (2012)

Mitkidis et al. (2015) compared HR synchrony between participant groups that both completed several rounds of a joint action task, and a treatment group that also played a public goods game after each session. The joint action task consisted of building model cars with LEGOs, while the public goods game entailed contributing to a common investment pool that was split equally between participants. HR synchrony was measured using Multivariate Recurrence Quantification Analysis (MVRQA), a time series analysis method used to quantify the degree of synchrony between three or more time series (Thomasson, Webber, & Zbilut, 2002, as cited in Mitkidis et al. 2015). They found that participants in the treatment condition had significantly higher HRs and synchrony, and that HR synchrony was a significant predictor of participants' expectations of their partners in the public goods game. Together, these findings show that PS might be a predictor of interpersonal trust in economic games.

III. Physiological Synchrony Methodology

Several analytical methods have been applied to measure PS and the development of an accurate measure is an ongoing endeavor within interpersonal autonomic physiology (IAP) research (Palumbo et al., 2017). The first section defines the six parameters identified by Palumbo et al. (2017) that have been used to define PS. Section two provides a brief overview of the various statistics of PS used since the inception of IAP research in the 1950s. Finally, the two statistical methods used to measure PS are discussed: intersubject correlation (ISC) and dynamic time warping (DTW).

1. **Components of Physiological Synchrony**

IAP is defined as the interconnection between individuals' physiological dynamics, as indexed by continuously measuring the autonomic nervous system (ANS) (Palumbo et al., 2017). There is evidence that the ANS is responsive to, and in some instances, dependent on and/or shaped by, the nervous system of others between-participants (Palumbo et al., 2017). Findings from IAP research indicate that physiological activity between two or more people can become associated or interdependent: PS refers to "any interdependent or associated activity identified in the physiological processes of two or more individuals" (Palumbo et al., 2017, p. 100). PS is also referred to as interpersonal physiology, physiological linkage, physiological coherence, and physiological covariation (Palumbo et al., 2017, p. 101). For example, Reed, Randall, Post, and Butler (2013) used the term physiological linkage "in reference to both concurrent and lagged interdependencies between participants' cardiac and electrodermal measures" (Palumbo et al., 2017, p. 102).

Following a review of the IAP literature, Palumbo et al. (2017) identified six parameters that have been used to define PS: magnitude, sign, direction, lag, timing, and arousal. Magnitude is a measure of "the strength of synchrony, such as a regression or correlation coefficient. This typically represents the effect size of a given measure of synchrony" (Palumbo et al., 2017, p. 102). Sign indicates the direction in which individuals' arousal levels move, either in the same or opposite directions. This is also referred to as concordant and discordant synchrony (Palumbo et al., 2017). The direction parameter "refers to the predictability of one person's physiology from another's" (Palumbo et al., 2017, p. 102). Tests of predictability may reveal a unidirectional or multi-directional relationship (Palumbo et al., 2017). Lag represents the shift in the temporal alignment of data. "The difference in time alignment from one data set to another is the specific lag, which may indicate millisecond differences, or much longer time offsets" (Palumbo et al., 2017, p. 102). According to Palumbo et al. (2017), "although lagged PS implied a unidirectional relationship, there is some evidence that the length of a lag (e.g., 1s vs. 10s-time offset) can reflect psychosocial properties that are independent of direction and worthy of exploration" (102). Lagged PS will be utilized in the intersubject correlation (ISC) analyses, explained below. Timing refers to the length of time that an interaction is assessed or analyzed. The time scale selected is a key element of the research question (Palumbo et al., 2017). Arousal can be assessed as a covariate of synchrony. For example, mean arousal level can be tested as a covariate or moderator of PS, such as examining whether increasing PS correlates with decreasing arousal (Palumbo et al., 2017).

Asynchrony is the opposite construct of PS; it describes a lack of observable PS (Palumbo et al., 2017). Asynchrony is "difficult to substantiate without the use of multiple models to test for PS," but identification of interactions that lack PS can provide valuable insight

(Palumbo et al., 2017, p. 103). For instance, periods of asynchrony may exist during an interaction of either person is ignoring their partner (Palumbo et al., 2017).

Palumbo et al.'s (2017) review of the IAP literature characterizes synchrony as a nonspecific construct, since its detection is dependent on the procedure used to test it. In general, analyses of synchrony identify both concurrent and/or lagged interdependencies between individuals' cardiac and/or electrodermal measures (Palumbo et al., 2017). Different analyses of synchrony address different components of data; therefore, the results can differ substantially (Palumbo et al., 2017). Thus, it is necessary that an appropriate PS analyses correspond to the research question(s) at hand because different approaches can alter study findings and implications (Palumbo et al., 2017). Additionally, most studies on the ANS have been performed at the intrapersonal level, or within-subjects (Palumbo et al., 2017). Instead, temporal changes are explored here at the interpersonal level, or between-subjects.

2. **Statistics of Physiological Synchrony**

The critical component of predicting interpersonal autonomic physiology (IAP) is the type of statistical analysis used. Determining an accurate measure of PS is an ongoing process within IAP research. Various analytical methods have been developed to create an accurate statistic of PS. The ideal methods are those that best match the research question(s) regarding the dyadic interactions (Gates & Liu, 2016). Statistical procedures used to measure PS have evolved since the first studies relied on Pearson correlations and cross-correlations. Although, cross-correlations have remained the most common measure of PS, with more advanced techniques applying moving windows and lags (Boker et al., 2002; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; Schoenherr et al., 2019). With advances in computing power, researchers have been applying new statistical techniques, with the goal of finding a measure that accurately measures PS without over-fitting.

Gates and Liu (2016) and Schoenherr et al. (2019) discussed several statistical techniques used for quantifying PS, reviewing dynamic systems approaches and linear time series analysis methods (TSAMs), respectively. Gates and Liu (2016) review the methods for dyad-level modeling, including: cross-correlation, vector autoregression (VAR), state-space models (SSMs), unified structural equation models (uSEMs), hidden Markov models (HMMs), state-space grids (SSGs) (GridWare), recurrence quantification analysis (RQA), differential equations, and spectral analysis. The linear time series analysis methods covered by Schoenherr et al. (2019) include: cross-lagged correlation (CLC), cross-lagged regression (CLR), windowed cross-correlation (WCC), various versions of windowed cross-lagged correlations (WCLCs), and windowed cross-lagged regression (WCLR). It should be noted, “the analysis of multivariate, non-stationary, intensive time series of physiology is wrought with complexities, as these data violate a number of assumptions of parametric statistics (e.g., stationarity, independence of measures)” (Palumbo et al., 2017, p. 126). Thus, behavioral science researchers have utilized a variety of statistical methods to measure PS.

3. **Intersubject Correlation (ISC)**

The first analysis conducted is intersubject correlation (ISC), which “provides a measure of the across-subject reliability of physiological responses by quantifying the commonalities of the response time courses among individuals” (Golland et al., 2014, p. 3). ISC analysis is an increasingly popular method to measure synchrony between individuals or groups. It is commonly used in physiological and neurological research, including ECG, EEG, and fMRI.

For this research, it will be used to determine whether continuous ANS signals, measured by EDA, exhibit time-locked response patterns that are consistent across individuals (Golland, et al., 2014).

ISC analysis calculates cross-correlations between participant pairs and derives an ISC measure from these correlations to determine the level of PS (Lahnakoski & Chang, 2021). Cross-correlations are Pearson correlations calculated between vectors of sequential points of the same length from two different time series (Boker et al., 2012). A time lag, which shifts the temporal alignment of the vectors, is present when the vectors do not begin at the same epoch (Boker et al., 2012). When cross-correlations are calculated using both windows and lags the analysis is referred to a windowed cross-lagged correlation (WCLC) (Schoenherr et al., 2019). ISC is a version of a windowed cross-lagged correlation (WCLC) and requires four parameters to be specified: window size, window increment, maximum lag, and lag increment. Since global stationarity is not assumed, the time series are broken into short, overlapping windows, which are used to find a moving estimate of association (Boker et al., 2012). This accounts for non-stationarity and the strength of concordance can vary between windows (Behrens, Moulder, & Kret, 2020). In addition, overlapping windows prevent missing points of synchronization that may occur at the edges of adjacent segments (Behrens, Moulder, & Kret, 2020). For each window, the segments are lagged up to a given maximum lag, which accounts for the potential varying delay between the two responses (Behrens, Moulder, & Kret, 2020).

Given two SCL time series, X and Y, of n observations:

$$X = x_1, x_2, \cdots, x_i, \cdots, x_n$$

$$Y = y_1, y_2, \cdots, y_i, \cdots, y_n.$$

Suppose a window size, w_{max}, and a time lag, τ, on the integer interval $\tau_{max} \geq \tau \geq -\tau_{max}$. For every i in w_{max}, a pair of window segments, W_x and W_y, are selected from the two series X and Y, respectively:

$$W_x = \left\{ \begin{matrix} \{x_i, x_{i+1}, x_{i+2}, \cdots, x_{i+w_{max}}\} \; if \; \tau \leq 0 \\ \{x_i, x_{i+1}, x_{i+2}, \cdots, x_{i+w_{max}}\} \; if \; \tau > 0 \end{matrix} \right\} \text{ and}$$

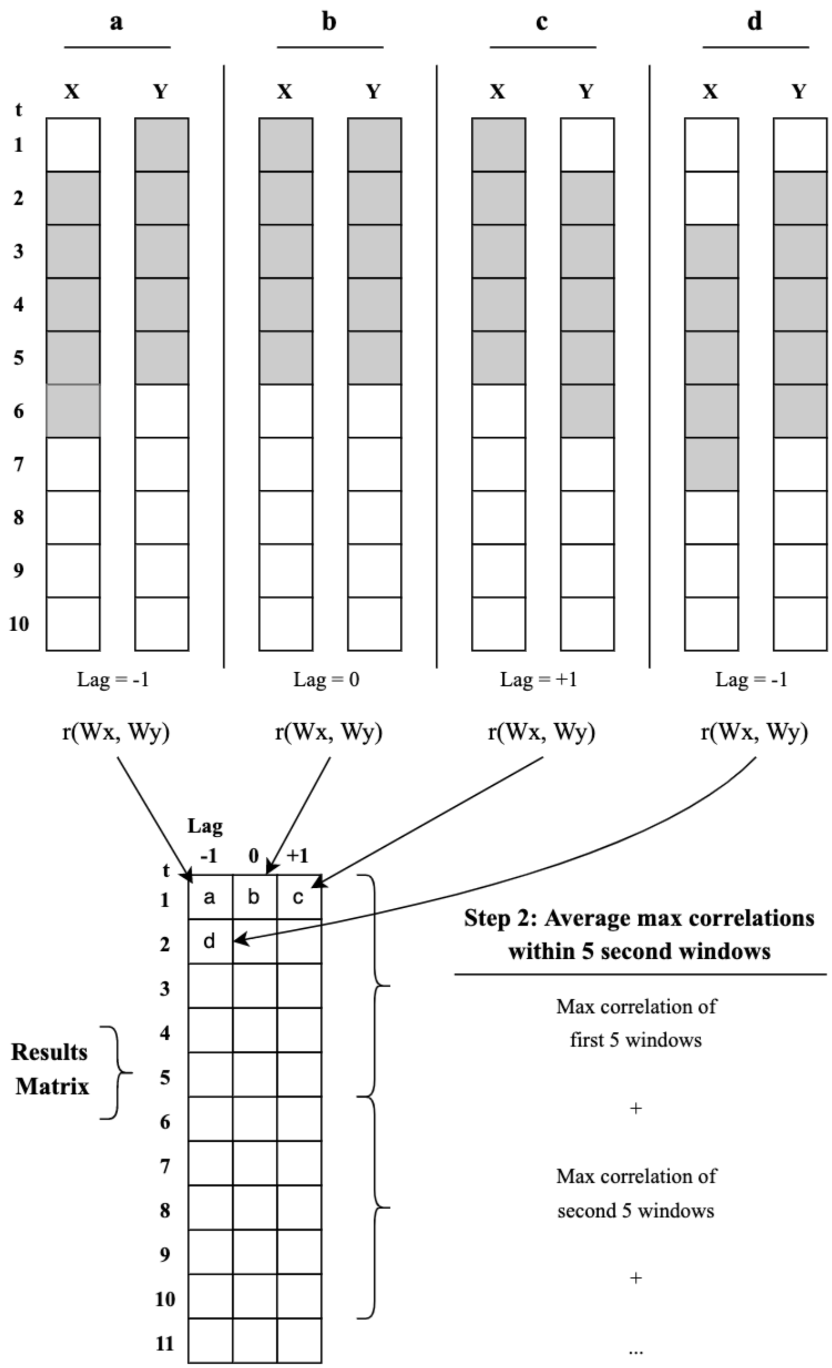
$$W_y = \left\{ \begin{matrix} \{y_i, y_{i+1}, y_{i+2}, \cdots, y_{i+w_{max}}\} \; if \; \tau \leq 0 \\ \{y_i, y_{i+1}, y_{i+2}, \cdots, y_{i+w_{max}}\} \; if \; \tau > 0 \end{matrix} \right\}.$$

Selecting the windows in the above manner results in a mirror symmetry such that if X and Y are swapped, the set of correlations for a given lag contain the same values in reverse order (Boker et al., 2012). The distinction between the two equations when τ is less than or greater than 0 prevents biased correlations if the series are non-stationary (Boker et al., 2012). The cross-correlations of each lag for each window segment generate a matrix, r, with the elements defined as

$$r(W_x, W_y) = \frac{1}{w_{max}} \sum_{i=1}^{w_{max}} \frac{(W_{xi} - \mu(W_x))\left(W_{yi} - \mu(W_y)\right)}{\sigma(W_x)\sigma(W_y)}.$$

Each row in the matrix represents one window segment and each column a lag (Behrens, Moulder, & Kret, 2020). The first and last columns of the matrix are the cross-correlations for the given maximum lag and the middle column contains the cross-correlations for a lag of 0. Therefore, the number of columns in the matrix is equal to $2\frac{\tau_{max}}{\tau_{inc}} + 1$ and the number of rows is equal to $\frac{N - w_{max} - \tau_{max}}{w_{inc}}$ (Behrens, Moulder, & Kret, 2020). To capture changes in peak cross-correlations between window segments, the maximum correlations within the specified widow size are averaged. The mean of the maximum cross-correlations of all window segments represents a participant's ISC. An example result matrix with a window size of 5 and a lag of 1 is depicted in Figure 8.

Figure 8: ISC Process Example with Window Size 5 and Lag 1, Adapted from Boker et al. (2012)



4. **Dynamic Time Warping (DTW)**

Dynamic time warping (DTW) is an algorithmic technique that uses dynamic programming optimization for time series data to compare two sequences by minimizing the cumulative distance between them (Berndet & Clifford, 1994). DTW has been used extensively for speech recognition since the 1970s, but it has a wide range of applications, including: handwriting and signature matching, sign language and gesture recognition, data mining and time series clustering (time series database searches), computer vision and animation, surveillance, protein sequence alignment and chemical engineering, music and signal processing, robotics, manufacturing, and physiology (e.g., electrocardiogram (ECG)) (Keogh & Pazzani, 2002; Keogh & Ratanamahatana, 2005; Senin, 2008). One advantage of the algorithm is that it does not require stationary data (Dilmi, Barthès, Mallet, & Chazottes, 2020). It will be used here as a synchrony measure to assess the similarity between dyads' SCLs. A classic version of the algorithm will be applied as follows:

Given two SCL time series, X and Y, of length n and m, respectively:

$$X = x_1, x_2, \cdots, x_i, \cdots, x_n$$

$$Y = y_1, y_2, \cdots, y_i, \cdots, y_n.$$

The series are aligned by constructing an $n \times m$ matrix where the (i^{th}, j^{th}) element is the distance $d(x_i, y_j)$ between the two points, x_i and y_j (Keogh & Pazzani, 2002). Euclidean distance is commonly used (Keogh & Pazzani, 2002), where:

$$d(x_i, y_j) = (x_i - y_j)^2.$$

The algorithm builds a cost matrix $C \in \mathbb{R}^{n \times m}$ representing all pairwise distances between X and Y (Müller, 2007; Senin, 2008). Each matrix element (i, j) corresponds to the alignment between points x_i and y_j (Keogh & Pazzani, 2002). A warping path, W, is calculated to align the elements of X and Y, such that the cumulative distance between them is minimized (Berndet & Clifford, 1994). The warping path is a sequence of grid points where the k^{th} element of W is defined as $(i,j)_k$ (Berndet & Clifford, 1994; Keogh & Pazzani, 2002):

$$W = w_1, w_2, \cdots, w_k, \cdots, w_K \qquad max(n,m) \leq K < n + m - 1.$$

The search space for possible warping paths was restricted by four conditions (Berndet & Clifford, 1994; Senin, 2008):

1. Monotonicity: The points are monotonically ordered with respect to time, $n_1 < n_2 \leq \cdots \leq n_K$ and $m_1 < m_2 \leq \cdots \leq m_K$.

2. Continuity: The allowable steps to calculate the warping path are confined to adjacent points, $i_k - i_{k-1} \leq 1$ and $j_k - j_{k-1} \leq 1$.

3. Warping window: A warping window constrains the allowable points to fall within a given positive integer window width, w, where $|i_k, j_k| \leq w$.

4. Boundary conditions: The start and end points of the warping path are the first and last points of the aligned sequences, $w_1 = (1,1)$ and $w_K = (n,m)$.This condition can be relaxed by an "offset" in the warping window. For example, a starting point is specified with subsequent path constraints replacing a fixed end point.

There are an exponential number of warping paths that satisfy these conditions; however, the DTW problem is defined as the path that minimizes the warping cost based on the cumulative distance of each path, where d is the Euclidean distance between the two time series elements (Berndet & Clifford, 1994; Keogh & Pazzani, 2002):

$$DTW(X,Y) = min_w \left[\sum_{k=1}^{p} d(w_k) \right].$$

Therefore, the optimal warping path has the lowest cumulative distance.

The optimal warping path is found using dynamic programming to calculate the following recurrence, which defines the cumulative distance $\gamma(i,j)$ as the sum of the distance $d(x,y)$ between the current element and the minimum of the cumulative distances of the adjacent points (Berndet & Clifford, 1994; Keogh & Pazzani, 2002):

$$\gamma(i,j) = d(x_i, y_j) + min[\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)].$$

The dynamic programming algorithm computes a table of cumulative distances and finds the optimal warping path by tracing backward in the table, choosing the previous points with the smallest cumulative distance (Berndet & Clifford, 1994). After the optimal warping path is found, a distance defining the "fit" of the two series is calculated, quantifying the degree of alignment achieved by stretching or compressing the series along the time axis (Berndet & Clifford, 1994).

Figure 9 is a graphical representation of the difference between calculating the distance between two time series using the simple Euclidean distance and DTW with the Euclidean distance. The DTW algorithm finds the minimum distance along the warping path within a given window width. This "warps" the time axis of one or both series, achieving a better alignment than the simple Euclidean distance measure (Keogh & Pazzani, 2002). Note that a DTW window of 0 is equivalent to the simple Euclidean distance measure.

Figure 9: Euclidean Distance versus Dynamic Time Warping (DTW) Distance, from Keogh and Ratanamahatana (2005).

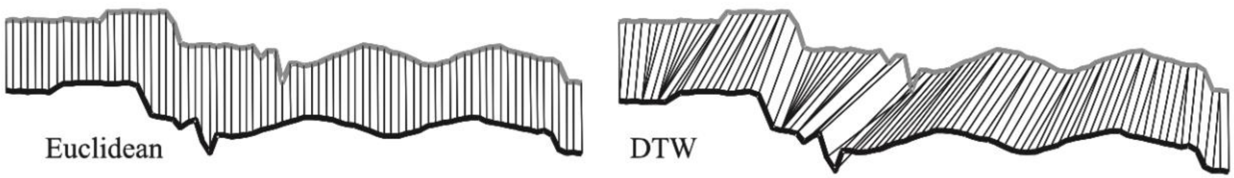


Fig. 1. Note that, while the two time series have an overall similar shape, they are not aligned in the time axis. Euclidean distance, which assumes the i^{th} point in one sequence is aligned with the i^{th} point in the other, will produce a pessimistic dissimilarity measure. The nonlinear dynamic time warped alignment allows a more intuitive distance measure to be calculated

IV. Methods

Models of trust and trustworthy behavior have recently been extended to examine the dynamics of the underlying processes and physiological correlates (Kret, 2015; Schoorman et al., 2007; van der Werff et al., 2019). Given the limited research regarding the motivational processes underlying trust (van der Werff et al., 2019), in this experimental study, I aim to determine whether participant pairs who communicate exhibit PS in the trust game, a strategic game with high monetary stakes. Thus, the goal of this research is to expand the existing literature on PS by examining a physiological correlate underlying trust and trustworthiness and high monetary stakes are involved. In addition, I aim to determine whether participant pairs who exhibit PS as measured by their relative electrodermal SCLs, a component of EDA, demonstrate trust and trustworthy behavior in the trust game.

Each participant was asked to play the dictator and the trust game while paired successively with two other participants; overall, each participant is part of two pairs, i.e., two dyads. The pairs were given 2 minutes to interact with each partner before making their trust game decisions and were instructed to make promises to each other about how much money they would choose to transfer. Participants' EDA, a physiological measure of the autonomic changes in the electrical conductance of the skin, was collected during these interaction periods. SCLs indicate arousal and thus were used to assess whether participants displayed PS (Braithwaite et al., 2015). When using these economic games as a measure of trust, several questions were pertinent for the study; does PS play a role in the motivation to engage in mutually beneficial behavior? Can the decision to trust and be trustworthy, thus the game outcome, be predicted by participants' PS? To address these questions, the experimental design, procedures, interaction period, participant sample, game models, and hypothesis testing are outlined below, followed by the data collection and cleaning process.

1. **Design**

In this quantitative, experimental, interpersonal trust and trustworthiness were measured using the trust game, with the dictator game as a control (Berg et al., 1995; Cox, 2004; Cox et al., 2016). The dictator game was played by assigning one of two paired individuals to be the dictator and endowing them with $120. The decisions in the game are unilateral; only the dictator can decide to share some of the endowment with their partner (receiver) or keep all of it (Bolton et al., 1998; Fehr & Krajbich, 2014). Participants were told that any amount shared would be tripled by the experimenter. The trust game was played by endowing two matched participants with $120, assigning one person the role of FM and the other as SM (Berg et al., 1995). The FM had a choice of sending $0, $40, $80, or $120 to the SM, which the participants were told would be tripled in value. Next, the SM chose whether to reciprocate and return a fraction of the money or keep it all. A dictator game with exactly the same decision set for the FM in the trust game was a control for participants' other-regarding preferences, such as altruism, a potential confound (Cox, 2004). As described in chapter 2, section 1.4.4., Measuring Trust with Construct Validity, the monetary amount sent in a dictator game with the same decision set as the FM in the trust game should be deducted from the amount sent as the FM in the trust game to differentiate trust from other-regarding preferences, thereby creating a measure of trust with construct validity (Cox, 2004). The experiment was designed for participants to play these games with a $120 endowment given to all participants at the beginning of each session.

For this study, groups of four participants completed two four-hour experimental sessions one week apart. Each session included two sets of participant pairs (dyads), in which each participant interacted separately (one-on-one) with two others in a group of four. The design included pairing participants to interact during a 2-minute communication period, where they were instructed to make promises regarding the decisions they intended to make during the trust games. The potential outcomes of these games were modeled to determine whether participant pairs that exhibited PS during the interaction period also realized the socially optimal outcome of the game, i.e., cooperation, where participant pairs could maximize their joint payoff by being trusting and trustworthy.

1. ***Procedure***

The experiment was conducted at the Center for Neuroeconomics Studies at Claremont Graduate University in Claremont, CA, between April and June 2012 and approved by the Institutional Review Board of Claremont Graduate University (#1006) and the United States Department of the Air Force (#FWR20110168X). To begin session one, phase one (Figure 10), participants arrived at the lab and signed a written informed consent form. Participants were then led into an experiment room separated by a curtain partition, where they completed the Positive Affect Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) to assess baseline affect and fitted with the electrophysiological instrumentation for EDA data collection (phase one, center boxes in Figure 10). This fitting consisted of two disposable Ag-AgCl electrodes placed on the palmar surface of the middle phalanges of the index and middle fingers of the participants' non-dominant hands. These were wirelessly connected to four BIOPAC MP150 data acquisition systems (BIOPAC Systems, Inc.; Goleta, CA). After fitting, research assistants performed an impedance check to ensure signal quality, and additional impedance checks were conducted during the experiment. Continuous EDA data were collected for the duration of the experiment, including several baseline quiet periods (blue boxes in phases 2-6 in Figure 10 and phases 1-5 in Figure 11). Participants were fitted with electrophysiological instrumentation for EDA data collection at the beginning of each session (phase one, purple boxes in Figures 10 and 11). Figures 10 and 11 illustrate the experimental procedures in sessions one and two, respectively, with the horizontal bars representing the sequence (phases) of events.

Figure 10: Session/Day 1

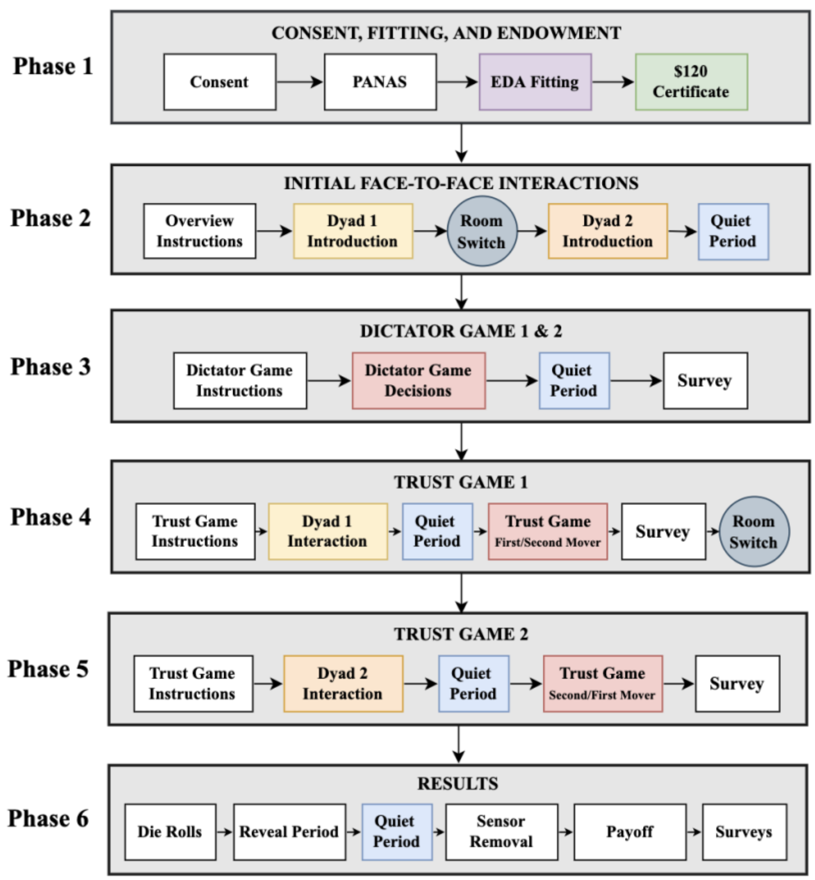
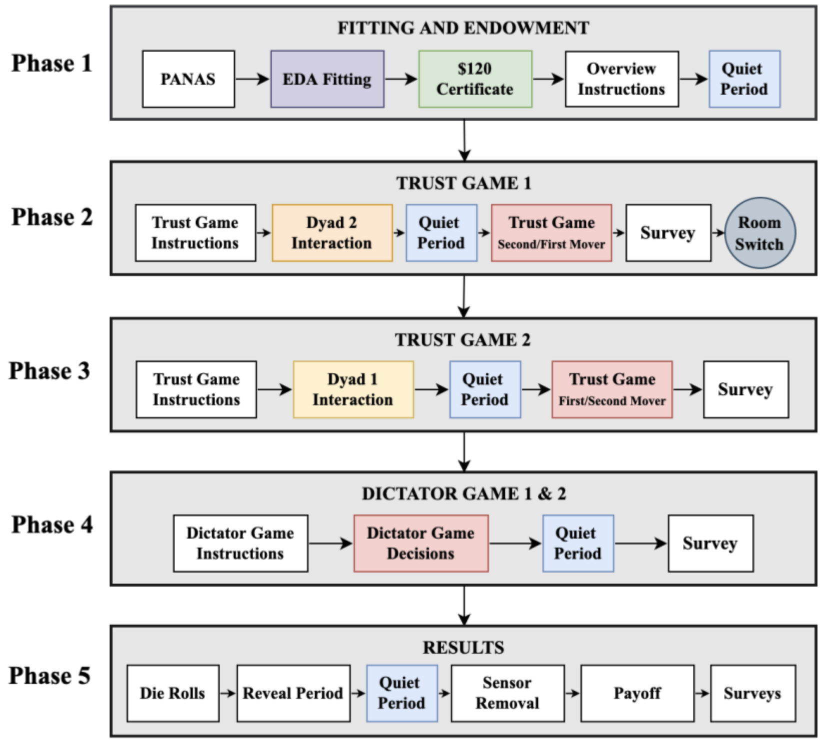


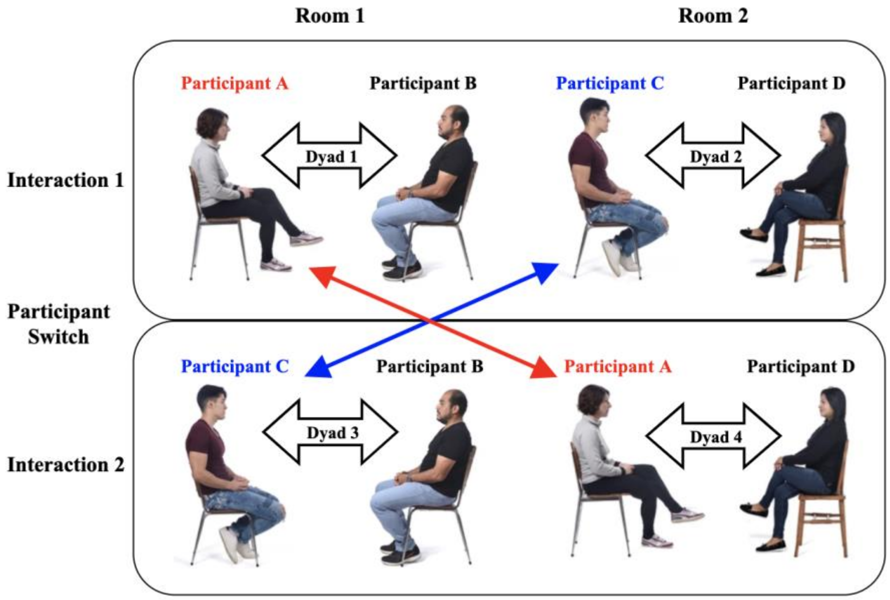
Figure 11: Session/Day 2



Each participant was stationed in front of a designated laptop, where they watched the instruction videos and made game decisions. The instructions were written in neutral language (shown in the Appendix). For instance, the dictator game was called the decision task, while the trust game was referred to as the investment task. To begin the games (i.e., tasks), participants received a $120 endowment as an earnings certificate, shown in the green box (phase 1 Figure 10). At the start of phase two, participants watched an instructional video on their designated laptops and listened to audio instructions using headphones to receive an overview of the experiment. Afterward, research assistants opened the curtain partition and instructed participants to introduce themselves for the initial face-to-face interaction (phase two, yellow box Figure 10).

The experiment was a within-subject design with individuals participating in four interactions: session/day one, interactions one and two, shown in phases 4 and 5 in Figure 10; session/day two, interactions one and two, shown in phases 2 and 3 in Figure 11. After the initial interactions in phase 2 (Figure 10), the curtain was drawn, and a baseline EDA measure was collected during an eight-minute quiet period (end of phase two, first blue box in Figure 10). In each experiment room, the two participants were seated next to each other, approximately three feet apart. Figure 12 illustrates the dyad pairings during the interaction periods, where each participant interacted successively with two others. In the figure, dyads one (participants A and B) and two (participants C and D) represent the first set of participant pair interactions in two separate rooms: participant A partnered with B in room one (dyad one), and participant C partnered with D in room two (dyad two). After the first initial interaction in phase two (Figure 10), one participant from each room switched rooms for the second initial face-to-face interaction, creating the second participant pairs, i.e., forming dyads three (participants B and C) and four (participants A and D), illustrated as interaction two (Figure 12). These initial face-to-face introductions did not take place in session two since participants were already acquainted.

Figure 12: Dyad Pairings



At the beginning of phase three (Figure 10), participants watched an instruction video on the dictator game, which was presented as a set of viewgraphs with voiceover instructions on their designated laptops. A handout was distributed after every instruction video to test participants' understanding of the games. Research assistants reviewed the completed handouts and answered any questions. Finally, participants made their game decisions using the laptops. All games were presented using software written in Python by the researchers (Zak et al., 2022), which prompted participants to choose how much of their endowment, if any, to transfer.

After a second quiet period in phase three (second blue box in Figure 10), research assistants informed participants they would receive instructions on the face-to-face interaction and investment task to follow in phases four and five (Figure 10). To begin phase four, participants watched an instruction video on the trust game. Again, a handout was used, and responses were verified by the research assistants to check participants' understanding. Next, the research assistants opened the curtain partition, reminded participants of the interaction instructions, and left the room for two minutes. Again, dyads were seated facing each other during the interactions (Figure 12), and video recordings captured their communication. Afterward, the partition was closed, participants turned to face their laptops, and a third quiet period ensued (phase four, third blue box in Figure 10). The software then prompted participants to make their trust game decisions in private (phase four, red box in Figure 10). At the end of phase four, one participant from each room switched rooms (phase four, grey circle in Figure 10), and this process was repeated for the second dyad pairing (phase five in Figure 10).

Participants played the trust game twice, once with each partner, as both the FM and SM, such that all participants played both roles. The FM's decision set is depicted in the upper/top portion of Figure 13; the SM's decisions are depicted in the lower half of Figure 13. The FM could send \$0, \$40, \$80, or \$120 to the SM, with the amount selected tripled by the experimenter before reaching the SM. The SM could choose to return \$0, return the amount they received, or split the total equally. Participants made decisions for each possible first-mover choice, i.e., participants chose how much to return as the SM if the FM were to send \$40, \$80, or \$120, known as the "strategy method" (Brandts & Charness, 2000; Brandts & Charness, 2011). They played the control dictator game twice, once with each partner, who chose how much of the \$120 endowment – \$0, \$40, \$80, or \$120 – to give their partners unconditionally (Figure 14). The design required participants to make 10 decisions in each session – five decisions per dyad (one as the dictator, one as the FM, and three as the SM), with two dyads per session (illustrated by the red boxes in phases 3-5 in Figure 10 and phases 2-4 in Figure 11). To minimize order effects, the sequence of interactions and game decisions (FM/SM) was counterbalanced across the two sessions, with the dictator games played last in session two (phase four in Figure 11). To eliminate the effects of repeated decisions, participants did not learn the impacts of their choices and those of their partners until the reveal period at the end of each session (depicted in phase six in Figure 10 and phase five in Figure 11).

Participants' earnings were based on only one game, determined by three die rolls at the end of each session, depicted in Figure 10 beginning phase 6 and phase 5 in Figure 11. The first die roll was performed to select the participant pair; the second determined which game would be selected for payment, i.e., the dictator or trust games, and the third determined the participants' roles. If the outcome of the second die roll was the dictator game, the third die roll determined whether the dictator or receiver role was paid. If the outcome was the trust game, the third die roll determined whether the first or SM role was paid. All participant pairs witnessed their corresponding die rolls. At the end of each session, participants only learned what their dyadic partner chose for the one game randomly selected to determine payouts (illustrated in phase six in Figure 10 and phase five in Figure 11). Selecting one decision for compensation gave participants an incentive to truthfully reveal their preferences (Cox, Sadiraj, Schmidt, 2011). Participants could earn a minimum of $50 (the participation amount) or as much as $530 per session, and earnings were paid privately in cash. More considerable sums of money were used in this experiment than in most other trust game experiments (Johnson and Mislin, 2011) so that participants' decisions had significance.

Participants returned to the lab one week later for the second session and were paired with the same persons they participated with in the first session. The procedure for the second session was similar to the first (Figure 11).

2. ***Interaction Period***

Every interaction period was conducted with the dyads seated facing each other, approximately three feet apart, as shown in Figure 12. The trust game instructions stated that before making any decisions, participants would have an opportunity to speak with their partner. Further, they would have two minutes to discuss any topics they choose, including a strategy for making decisions. Specifically, participants were instructed that they could make promises to each other about the decisions they would make if they so desired.

After checking their understanding of the investment game, a research assistant read from a script reminding participants that during the interaction period, they could strategize and make a promise regarding the decisions they intended to make. The objective was to focus their interaction on these decisions. Two minutes gave participants enough time to discuss their choices and potentially make pre-decision commitments about how much they would each send as the sender (FM) and return as the receiver (SM).

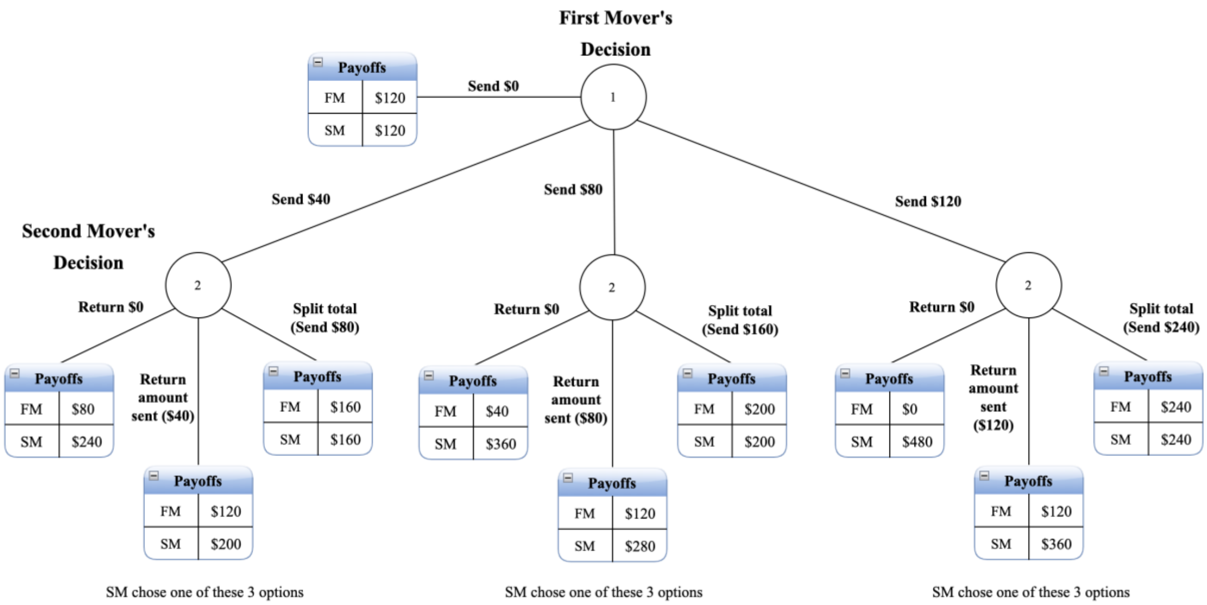
3. ***Participants & Sample***

The data were collected as part of a broader neurophysiological experiment (Willoughby et al., 2012), in which the participant pool consisted of 76 non-student adults between the ages of 25-50 (32 females, $M_{Age} = 30$, $SD_{Age} = 5.73$). In contrast to most trust game experiments that study college students exclusively (Johnson and Mislin, 2011), the objective was to recruit working-age adults from diverse cultural backgrounds. Participants self-identified as White (47%), Asian (20%), Latino/Hispanic (16%), and African American/Black (17%). Participants were a mix of native-English speakers born and raised in the United States and non-native English speakers born and raised outside the United States. Thirty-five participants were born and raised in the United States; 41 were born and raised outside the United States. Participants who moved to the U.S. after 16 years old were considered as raised abroad. All participants were fluent in English. Participants were recruited via the Center for Neuroeconomics Studies website and by direct outreach to targeted community groups. Participants were pseudo-randomly paired, resulting in 47% mixed-gender pairings.

4. ***Trust Game Model***

Each participant played the trust game twice (Figure 13), once as a FM or SM, then again with a role reversal such that all participants played both roles. If, as the FM in the upper left of Figure 13, a participant chose to send $0, no exchange occurred, and players retained their $120 endowment. The other three branches of the left side represent the SMs' decision set after receiving $40 from the FM: keep all the money (return $0), return the $40 the FM sent, or split the total equally. These three second-mover choices were used to limit decision options. Similarly, the center branch shows a FM send of $80 with three subsequent SM options, and the right side shows a FM send of $120 to the SM and the subsequent options. For each FM choice, the experimenter tripled the transferred amounts to the SM. Trustworthy behavior was defined as a participant choosing to split the total as the SM, making the FM better off than if they had kept their initial endowment. Participants played the trust game using the strategy method, whereby SMs made hypothetical decisions for each possible first-mover decision. In other words, participants chose how much to return as the SM if the FM were to send $40, $80, or $120. Generally, decisions made in trust games using the strategy method are comparable to those made in direct response to FM decisions (Brandts & Charness, 2000; Brandts & Charness, 2011).

Figure 13: Trust Game



The trust game can be modeled as follows. As a FM, participants chose $s_{TG} \in S$, where the feasible set S was defined by

$$S = 0, 40, 80, 120. \tag{1}$$

Usually, the choice of s_{TG} by the FM selects the $\Gamma(s_{TG})$ subgame, in which the SM chooses $r_{TG} \in R(s_{TG})$. Based on the design of the experiment, I assessed the $\Gamma(120)$ subgame, in which participants chose $r_{TG} \in R(120)$, where

$$R(120) = 0, 120, 240. \tag{2}$$

I assessed the $\Gamma(120)$ subgame as a measure of trustworthiness because this reflects the situation where the FM appears to show maximum trust by choosing $s_{TG} = 120$. However, the details of the subgame reveal that by sending $120, the FM had a 33% chance of a higher payoff than if they had not trusted their partner(s) and sent nothing.

Because a FM does not know the choice the SM will make, a FM's beliefs are modeled by a probability function. Let the probability distribution function $\Omega(\tilde{r}|s_{TG})$, defined on $R(s_{TG})$, represent a FM's beliefs about the amount of money, $\tilde{r}$, that the SM will send in the subgame $\Gamma(s_{TG})$. Also, assume that $\Omega(\tilde{r}|s_{TG})$ assigns a positive probability to the outcome $r = 0$ for all $s_{TG} \in S$. As the FM, a participant has expected utility from choosing s_{TG}^{o} in the trust game, is

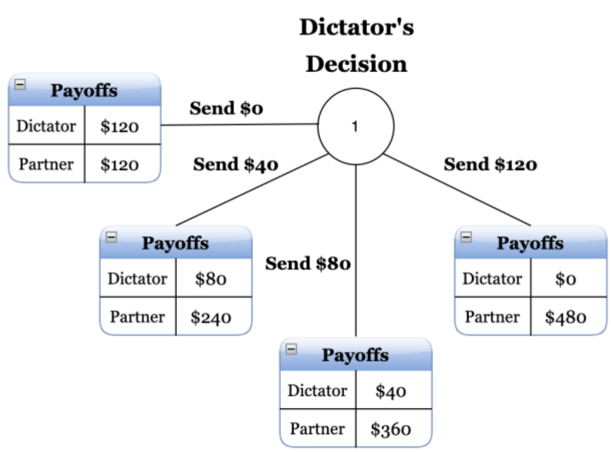
$$EP_{TG}^{1} = E_{\Omega(\tilde{r}|s_{TG})}[u^1(120 - s_{TG} + \tilde{r}, 120 + 3s_{TG} - \tilde{r})]. \quad (3)$$

If, as the FM, a participant's preferences are self-regarding, u^1 is a constant function of the SM's payoff.

5. ***Dictator Game Model***

Participants played a triple dictator game, shown in Figure 14, with two other participants. If the dictator's choice, as shown in the upper left of Figure 14, is to send \$0, no exchange occurs, and players retain their \$120 endowment. The next branch on the center-left side, bottom center, and center-right represent the dictators' decision to send money, as shown by sending \$40, \$80, or \$120, respectively.

Figure 14: Dictator Game



The dictator game can be modeled as follows. As the dictator (sender), each participant chose an amount to transfer to their matched partner, $s_{DG} \in S$, where the feasible set S was the same as in the trust game, defined in equation (1) (Cox, 2004). A utility function can represent the preferences of each participant's payoff and the paired participants' payoffs. The optimal choice, s_{DG}^{o}, is

$$s_{DG}^{o} = argmax_{s_{DG} \in S}\, u^1(120 - s_{DG}, 120 + 3s_{DG}). \quad (4)$$

The utility payoff to the dictator, P_{DG}^{1}, is

$$P_{DG}^{1} = u^1(120 - s_{DG}, 120 + 3s_{DG}). \quad (5)$$

Each participant's choice in the dictator game, s_{DG}^{o}, and strict quasi-concavity of u^1 implies

$$u^1(120 - s_{DG}, 120 + 3s_{DG}) \geq u^1(120 - s, 120 + 3s), \text{ for all } s \in S. \quad (6)$$

6. ***A Measure of Trust with Construct Validity***

According to Cox (2004), a measure of trust with construct validity is defined as the amount sent by the FM in the trust game subtracted by the amount, they sent in a control dictator game (Alos-Ferrer and Farolfi, 2019). Furthermore, based on the experimental design and economic definition of trust as a behavioral measure, a FM displays trust if they send more money in the trust game than in the dictator game, $s_{TG} > s_{DG}$ (Cox, 2004). In this case, we know a FM has intentionally exposed themselves to the risk of loss from defection. This estimation of trust is applied to the models in this study, and more specific definitions corresponding to the decision sets, amounts sent, and implicit outcomes are outlined in Table 1.

The first column of Table 1 categorizes the level of FM trust: high trust, medium trust, low trust, no trust, and no trust with other-regarding preferences. The second and third columns represent participants' dictator game decisions, s_{DG}, and FM trust game decision, s_{TG}, respectively. The last column lists the difference between these amounts, $s_{TG} - s_{DG}$, characterizing the implicit outcome of trust. High trust is defined by participants who send $0 in the dictator game and $120 as the FM in the trust game, where the difference is equal to $120. Medium trust is represented by the difference between the amounts sent in the two games equaling $80. A difference of $40 between the amounts sent is designated low trust. The label no trust identifies participants who send no money in either game. Finally, no trust - other-regarding preferences represent participants who send the identical amounts in both games, where $s_{TG} = s_{DG} > \$0$.

Definition	s_{DG}	s_{TG}	$s_{TG} - s_{DG}$
High trust	$0	$120	$120
Medium trust	$0	$80	$80
	$40	$120	$80
Low trust	$0	$40	$40
	$40	$80	$40
	$80	$120	$40
No trust	$0	$0	$0
No trust - other-regarding preferences	$40	$40	$0
	$80	$80	$0
	$120	$120	$0

Table 1: First Mover Trust Definitions.

To measure trustworthiness, I assess the SMs' decision, r_{TG}, only in response to the FM showing maximum trust by sending the full amount, $s_{TG} = 120$, i.e., the $\Gamma(120)$ subgame (depicted by the bottom right branch in Figure 13 and the first column of Table 2). Since the SM had three options, keep all the money (return $0), return the $120 the FM sent, or split the total equally, I define trustworthy behavior as splitting the total equally by returning $240 to the FM (shown in the second row of Table 2). The decisions to keep all the money and return $120 are defined as untrustworthy, i.e., betrayal, as these decisions either make the FM worse off or no better off than had they not extended trust, respectively (Coleman, 1990; Evans & Krueger, 2011; Hong & Bohnet, 2007; Thielmann & Hilbig, 2015).

Table 2: Second Mover Trustworthy and Untrustworthy Definitions

Definition	s_{TG}	r_{TG}
Trustworthy	$120	$240
Untrustworthy (Betrayal)	$120	$120
	$120	$0

Thus, trustworthy behavior is defined as a participant choosing to split the total as the SM, making the FM better off than if they had kept their initial endowment. Mutual trust can be defined as participants transferring more as a FM in the trust game than in the dictator game, $s_{TG} > s_{DG}$, *a*nd choosing to split the total as the SM. These decisions reflect the socially optimal outcome of the game, cooperation. Participant pairs who chose this strategy maximized their joint payoff by being trusting and trustworthy, with no guarantee that their partner would fulfill their promise.

2. **Hypotheses**

The intuition behind this trust game as a model of trust and trustworthiness depends on the assumption that the FM participants choose to send their total endowment and think of themselves as playing their part in joint action, while the SMs choose to split the total (i.e., share the maxed amount). I hypothesize that two individuals experiencing mutual arousal, exhibited by their SCL, may be motivated to trust and be trustworthy, measured by the dollar amounts transferred in the trust game.

PS is implied by a higher ISC and a smaller DTW distance. I hypothesize that participants who exhibit trust by sending more money as the FM in the trust game than in the dictator game, $s_{TG} > s_{DG}$, have positive average ISCs. Specifically, participants who demonstrate low trust, medium trust, or high trust, as defined in Table 1, have positive average

ISCs. Further, participants who trust more by taking a more considerable risk and sending more money exhibit greater PS with their partner, i.e., have higher average ISCs, than participants who take a more negligible risk. As the SM, I hypothesize that participants exhibiting trustworthy behavior by choosing to split the total have positive average ISCs. Moreover, participants who exhibit trustworthy behavior by choosing to split the total have greater average ISCs than participants who do not exhibit trustworthy behavior by either choosing to return $0 or return the amount sent.

Additionally, I hypothesize that participants who demonstrate trust have smaller DTW distances than participants who do not demonstrate trust, some of whom exhibit other-regarding preferences, $s_{TG} = s_{DG} > 0$, and some of whom do not transfer any money, $s_{TG} = s_{DG} = 0$. More specifically, participants demonstrating low, medium, or high trust may have smaller DTW distances than participants demonstrating trust or no trust - other-regarding preferences. Further, I hypothesize participants who trust more by sending more money exhibit greater PS with their partner, i.e., have smaller DTW distances than participants who spend less money. Finally, as the SM, participants who exhibit trustworthy behavior by splitting the total might have smaller DTW distances than participants who do not exhibit trustworthy behavior by either choosing to return $0 or return the amount sent.

Hypothesis 1: Participants who engage in trusting behavior may exhibit PS with their partner, as measured by their ISCs. The null hypothesis is that the ISCs of the trusting individuals are not significantly different from those who did not trust their partner. A Kruskal-Wallis test will be performed to test the following hypothesis, as the initial analysis shows the data are not normally distributed:

H_0: $ISC_0 = ISC_{40} = ISC_{80} = ISC_{120}$

H_A: Not H_0

Hypothesis 2: Participants who demonstrate more trust by sending *significantly* more money as the FM in the trust game than in the dictator game may have greater ISCs than participants who exhibit less trust by sending *slightly* more money in the trust game than in the dictator game. Participants who exhibit high trust might have higher ISCs than those who demonstrate medium trust. Participants who show medium trust should have higher ISCs than individuals who demonstrate low trust. In turn, participants who demonstrate low trust may have the lowest ISCs of the individuals who exhibit trust. Finally, participants who demonstrate low trust should have higher ISCs than those who do not trust. The null hypothesis is that there is no difference in the ISCs between participants, i.e., participants who exhibit greater trust by sending more money do not exhibit greater PS with their partner than participants who demonstrate less or no trust. Mann-Whitney U tests will be performed to test the following hypothesis:

H_0: $ISC_{120} = ISC_{80}$ & $ISC_{80} = ISC_{40}$ & $ISC_{40} = ISC_0$

H_A: $ISC_{120} > ISC_{80}$ & $ISC_{80} > ISC_{40}$ & $ISC_{40} > ISC_0$

Hypothesis 3: Participants who demonstrate trust may exhibit PS with their partner, as measured by their DTW distances. The null hypothesis is that all participants have similar DTW distances.

H_0: $DTW_{120} = DTW_{80} = DTW_{40} = DTW_0$

H_A: Not H_0

Hypothesis 4: Participants who demonstrate more trust by sending *significantly* more money as the FM in the trust game than in the dictator game may have smaller DTW distances than participants who demonstrate less trust by sending *slightly* more money in the trust game than in the dictator game. Participants who exhibit high trust may have smaller DTW distances than those who demonstrate medium trust. Participants with medium trust should have smaller DTW distances than those with low trust. In turn, participants who demonstrate low trust may have the greatest DTW distances of those who exhibit trust. Finally, participants who demonstrate low trust might have smaller DTW distances than those who do not trust. The null hypothesis is that there is no difference in DTW distances between participants.

$H_0\text{:}\ DTW_{120} = DTW_{80}\ \&\ DTW_{80} = DTW_{40}\ \&\ DTW_{40} = DTW_0$

$H_A\text{:}\ DTW_{120} < DTW_{80}\ \&\ DTW_{80} < DTW_{40}\ \&\ DTW_{40} < DTW_0$

Hypothesis 5: As the SM, participants who exhibit trustworthy behavior by splitting the total should have higher ISCs on average than untrustworthy participants who either choose to return $0 or the amount sent. The null hypothesis is that there is no difference in ISCs between the participant groups.

$H_0\text{:}\ ISC_{Trustworthy} = ISC_{Untrustworthy}$

$H_A\text{:}\ ISC_{Trustworthy} > ISC_{Untrustworthy}$

Hypothesis 6: As the SM, participants who exhibit trustworthy behavior by splitting the total might have smaller DTW distances than participants who do not exhibit trustworthy behavior by either choosing to return $0 or the amount sent. The null hypothesis is that all participants have similar DTW distances.

$$H_0: DTW_{Trustworthy} = DTW_{Untrustwrothy}$$

$$H_A: DTW_{Trustworthy} < DTW_{Untrustwrothy}$$

3. **Data Collection**

EDA data were collected by the original researchers (Willoughby et al., 2012; Zak et al., 2022) at a sampling rate of 1 kHz using four BIOPAC MP150 data acquisition systems and Acknowledge software (BIOPAC Systems, Inc.; Goleta, CA). Two disposable Ag-AgCl electrodes were placed on the palmar surface of the middle phalanges of the index and middle fingers of the participant's non-dominant hand. An eight-minute EDA baseline measurement was collected while the participant looked straight ahead at a blank screen, wore headphones to mask any background noise, and did not engage in any task. The SCL data were collected during the dyadic interactions when participants were given two minutes to talk and strategize.

4. **Data Cleaning**

A visual inspection of the EDA waveforms was conducted to detect temporary periods of signal loss. Signal losses of five seconds or less were reconstructed using average values from an adjacent part of the waveform. Signal losses greater than five seconds were not interpolated. The data were converted to microsiemens (μS), and a 10-Hz low-pass filter was applied to remove high-frequency noise (Norris, Larsen, & Cacioppo, 2007). Next, a square root transformation was applied to correct the positive skew inherent to SC data (Dawson, Filion, & Schell, 1989; Figner et al., 2011). Average SCL (i.e., the tonic component) was extracted for the eight-minute baseline and two-minute interaction periods. These series were exported to CSV files and processed using Python. Finally, the series of interest – the interaction period – was baseline corrected, normalized, and then standardized between participant pairs. After removing missing and incomplete data, 222 individual participant interactions were analyzed.

V. Results

Seventy-six participants completed at least one game play and interaction round, for a total of 222 interactions. The results of the FM and SM decisions are summarized in sections 1 and 2, below. ISC and DTW were used to assess PS between participant pairs; summary statistics and hypothesis tests results are reported in the following sections.

1. First Mover (FM) Results

The FM results of the dictator and trust game are presented in Table 1. Table 2 shows the FM decisions by trust level. Between all sessions, participants trusted their partners with their $120 endowment as FMs in 62 out of 222 interactions, approximately 28% of the time. In 89 interactions participants chose not to send any of their endowment (sent $0) in approximately 40% of interactions. Figure 15 shows FM trust levels in all sessions. Figures 16 and 17 show the distributions of amounts FMs entrusted to their partners by session and by interaction, respectively. Figure 18 shows a plot of one participant pairs' baseline corrected, normalized, and standardized SCLs during an interaction period.

Table 3: Distribution of Dictator and Trust Game Decision

		Dictator Game (DG) Decision				
		0	**40**	**80**	**120**	**Total**
Trust Game (TG) Decision	**0**	51	–	–	–	51
	40	12	6	–	–	18
	80	8	13	5	–	26
	120	62	26	12	27	127
	Total	133	45	17	27	222

TG > DG	TG = DG	TG < DG

Table 2: First Mover Decisions by Trust Level

Trust Level	s_{DG}	s_{TG}	Count
High Trust, n = 62	$0	$120	62
Medium Trust, n = 34	$0	$80	8
	$40	$120	26
Low Trust, n = 37	$0	$40	12
	$40	$80	13
	$80	$120	12
No Trust, n = 51	$0	$0	51
No Trust, Other-Regarding Preferences, n = 38	$40	$40	6
	$80	$80	5
	$120	$120	27

Figure 15: Distribution of Amounts First Movers Entrusted to Second Movers

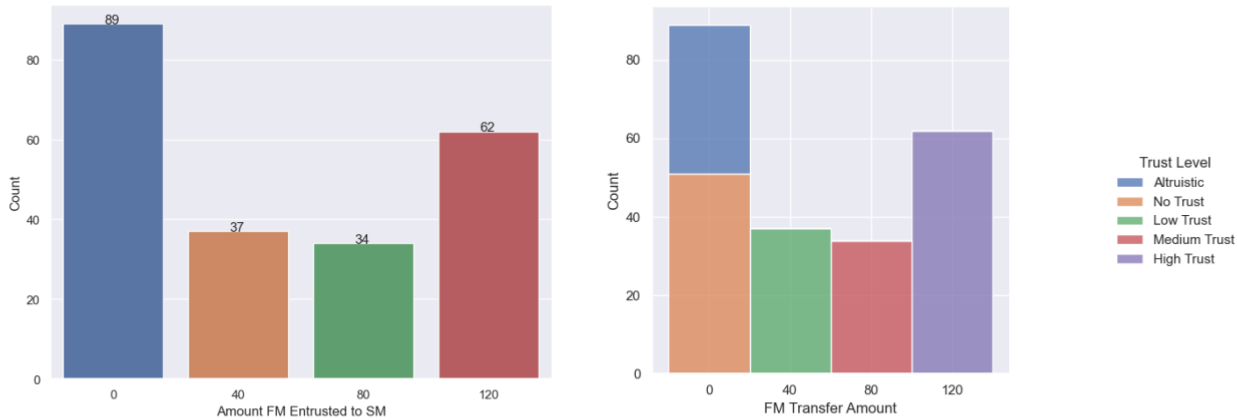


Figure 16: Distribution of Amounts First Movers Entrusted to Second Movers by Session

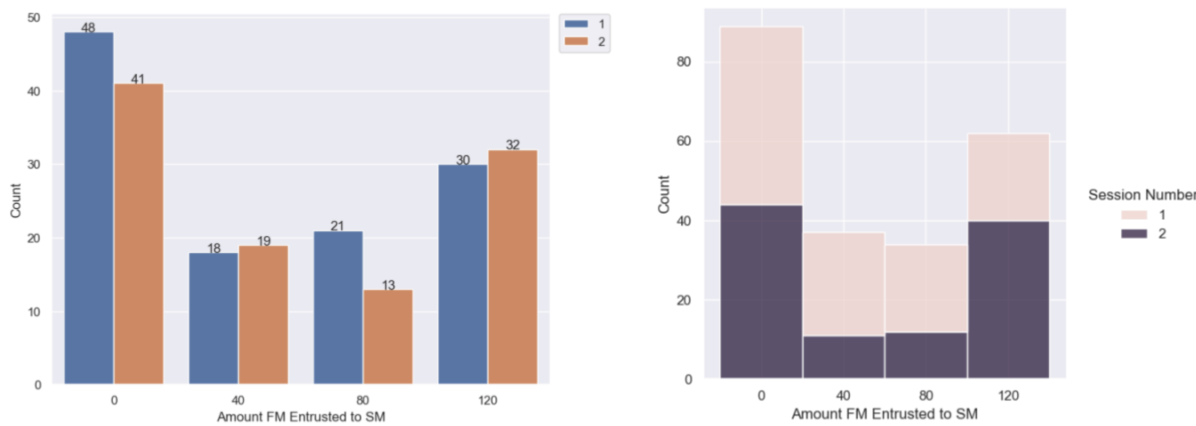


Figure 17: Distribution of Amount First Movers Entrusted to Second Movers by Interaction

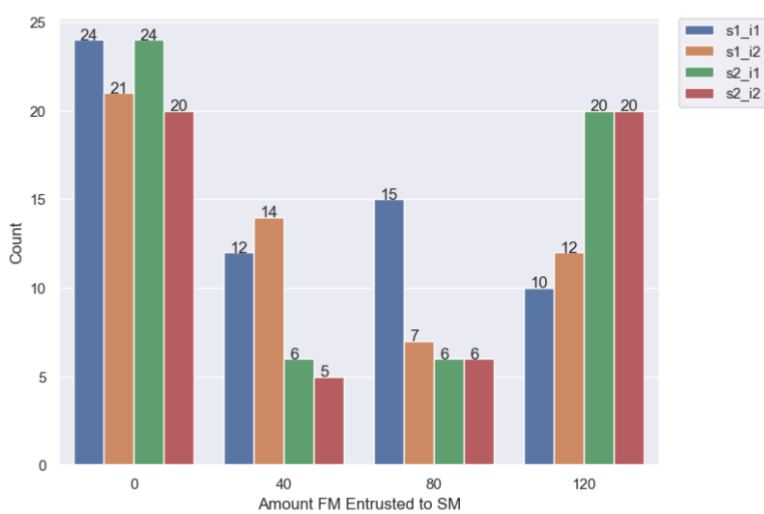
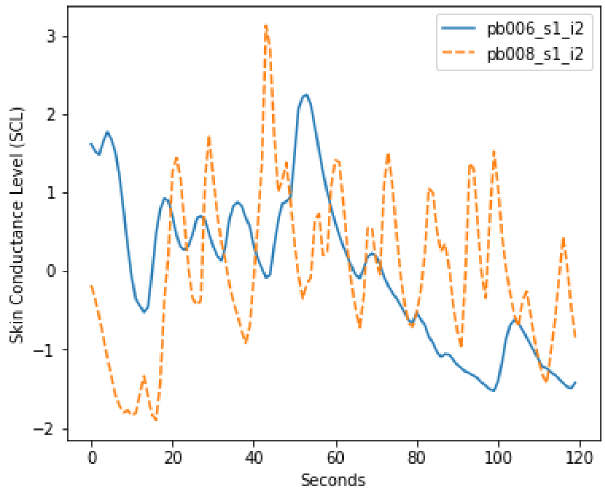


Figure 18: Example Participant Pairs' SCLs During One Interaction Period



1. ***FM ISC Results***

ISCs were calculated to measure PS during the interaction period. ISCs were calculated with window sizes 2, 3, 5, and 8, each with a lag of 1, for a total of four ISCs per participant: window size 2, lag 1; window size 3, lag 1; window size 5, lag 1; window size 8, lag 1. Summary statistics of the ISCs with window sizes 2, 3, 5, and 8 a lag of 1 for each trust level: no, low, medium, and high trust, are reported in Tables 4, 5, 6, and 7, respectively. The following sections report the ISC results of hypotheses 1 and 2 with corresponding post hoc tests.

Table 4: Summary Statistics of First Mover ISCs with Window Size 2, Lag 1

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	0.10	0.08	0.08	0.10
Mean	0.48	0.40	0.48	0.45
Median	0.47	0.43	0.47	0.45
Max	0.77	0.73	0.77	0.77
SD	0.14	0.17	0.17	0.14
SE	0.01	0.03	0.03	0.02
n	89	37	34	62

Table 5: Summary Statistics of First Mover ISCs with Window Size 3, Lag 1

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	0.51	0.40	0.51	0.32
Mean	0.81	0.79	0.82	0.79
Median	0.82	0.81	0.83	0.82
Max	0.99	0.98	0.98	1.00
SD	0.10	0.14	0.11	0.13
SE	0.01	0.02	0.02	0.02
n	89	37	34	62

Table 6: Summary Statistics of First Mover ISCs with Window Size 5, Lag 1

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	0.35	0.41	0.48	0.41
Mean	0.74	0.71	0.76	0.74
Median	0.75	0.72	0.77	0.76
Max	0.93	0.90	0.91	0.94
SD	0.11	0.11	0.09	0.11
SE	0.01	0.02	0.02	0.01
n	89	37	34	62

Table 7: Summary Statistics of First Mover ISCs with Window Size 8, Lag 1

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	0.17	0.24	0.39	0.34
Mean	0.69	0.67	0.70	0.72
Median	0.72	0.71	0.70	0.74
Max	0.88	0.87	0.94	0.95
SD	0.12	0.15	0.12	0.12
SE	0.01	0.02	0.02	0.02
n	89	37	34	62

1. 1. **Hypothesis 1**

Hypothesis 1 stated that participants who engaged in trusting behavior will exhibit PS with their partner, as measured by their ISCs. Testing this hypothesis required comparing the ISCs of the four trust groups: no, low, medium, and high trust. Shapiro-Wilk tests were performed on each window size and showed the data are not normally distributed: window size 2, lag 1, $W = .98,\ p = .02$; window size 3, lag 1, $W = .93,\ p < .001$; window size 5, lag 1, $W = .96,\ p < .001$; window size 8, lag 1, $W = .95,\ p < .001$. Therefore, the nonparametric Kruskal-Wallis test was used to compare the groups. The results of the Kruskal-Wallis test of ranked ISCs indicated that none of the window sizes were statistically significant: window size 2, lag 1 ($H(3) = 6.66,\ p = .08$); window size 3, lag 1 ($H(3) = 1.00,\ p = .80$); window size 5, lag 1 ($H(3) = 3.41,\ p = .33$); window size 8, lag 1 ($H(3) = 3.00,\ p = .39$).

1. 2. **Hypothesis 2**

Mann-Whitney U tests were conducted to determine whether FMs who demonstrated more trust by sending *significantly* more money in the trust game than in the dictator game have greater ISCs than those who exhibited less trust by sending *slightly* more money. Three one-tailed comparisons were run: median ISCs of high versus medium, medium versus low, and low versus no trust participant groups. Because three groups were compared, target p-values were adjusted accordingly; alpha levels .05 and .10 were Bonferroni corrected to .0167 (.05/3) and .033 (.1/3), respectively. The results for window sizes 2, 3, 5, and 8 with lags of 1 are reported in Tables 8, 9, 10, and 11, respectively.

Table 8: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – ISC Window Size 2, Lag 1

Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	2086.5		
p-value	0.99		
Medium Trust			
MWU statistic		789.0	
p-value		0.033	
High Trust			
MWU statistic			918.0
p-value			0.85

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 9: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – ISC Window Size 3, Lag 1

Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	1705.5		
p-value	0.62		
Medium Trust			
MWU statistic		699.0	
p-value		0.21	
High Trust			
MWU statistic			932.0
p-value			0.83

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 10: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – ISC Window Size 5, Lag 1.

Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	1906.0		
p-value	0.92		
Medium Trust			
MWU statistic		777.0	
p-value		0.04	
High Trust			
MWU statistic			1013.0
p-value			0.62

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 11: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – ISC Window Size 8, Lag 1

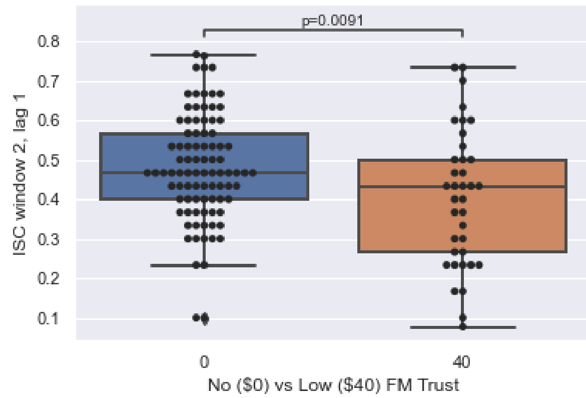
Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	1792.0		
p-value	0.78		
Medium Trust			
MWU statistic		668.0	
p-value		0.33	
High Trust			
MWU statistic			1183.0
p-value			0.16

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

1. 2. 1. ***ISC Results: Window Size 2, Lag 1***

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (.0167 for alpha level .05) indicated that median ISCs of the no trust participants were higher than those of the low trust participants, $U = 2086.5,\ p = .009$. Therefore, the individuals who did not trust their partner exhibited greater PS than those who showed low trust. This result was unexpected given the hypothesis that individuals who trusted their partner would have higher ISCs than those who did not trust. Box plots of the distributions of ISCs with window size 2, lag 1 of no trust participants compared to low trust participants and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 19.

Figure 19: Box Plots of Distributions and Comparison of No ($0) vs Low ($40) Trust Group – ISC Window Size 2, Lag 1



2. ***FM DTW Results***

DTW distances were calculated to measure PS during the interaction period. Distances were calculated using the Euclidean distance and four windows widths: 0, 5, 10, and 15 to assess a broad range of window sizes. Summary statistics of the DTW distances with window widths 0, 5, 10, and 15 for each trust level: no, low, medium, and high trust, are reported in Tables 12, 13, 14, and 15, respectively. The following sections report the DTW results of hypotheses 3 and 4 with corresponding post hoc tests.

Table 12: Summary Statistics of First Mover DTW Distances with Window Width 0

Decision	**No Trust**	**Low Trust**	**Medium Trust**	**High Trust**
Min	16.47	34.51	16.47	20.50
Mean	69.42	76.47	71.98	61.80
Median	65.35	63.58	66.17	59.35
Max	200.55	144.15	129.57	117.19
SD	30.15	29.73	26.96	25.22
SE	3.20	4.89	4.62	3.20
n	89	37	34	62

Table 13: Summary Statistics of First Mover DTW Distances with Window Width 5

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	16.34	58.24	16.47	24.68
Mean	100.90	115.28	109.83	95.84
Median	102.17	112.50	112.84	97.84
Max	200.55	173.59	164.63	174.53
SD	30.15	31.98	32.03	35.63
SE	3.75	5.26	5.49	4.52
n	89	37	34	62

Table 14: Summary Statistics of First Mover DTW Distances with Window Width 10

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	16.34	47.83	16.47	21.39
Mean	88.96	103.57	99.07	84.35
Median	87.29	99.94	100.15	86.35
Max	200.55	161.94	150.30	163.57
SD	35.43	34.32	30.76	35.44
SE	3.76	5.64	5.28	4.50
n	89	37	34	62

Table 15: Summary Statistics of First Mover DTW Distances with Window Width 15

Decision	No Trust	Low Trust	Medium Trust	High Trust
Min	16.34	39.49	16.47	20.51
Mean	81.66	94.68	91.14	77.01
Median	78.31	86.17	90.34	76.51
Max	200.55	155.82	145.56	151.05
SD	34.63	35.48	30.57	34.32
SE	3.67	5.83	5.24	4.36
n	89	37	34	62

2. 1. **Hypothesis 3**

Hypothesis 3 tests whether participants who demonstrated trust exhibited PS with their partner, as measured by their DTW distances. Shapiro-Wilk tests were performed on each window width, which indicated that the data of two of the four window widths are not normally distributed: DTW window 0, $W = .96,\ p < .001$; DTW window 5, $W = .99,\ p = .44$; DTW window 10, $W = .99,\ p = .12$; DTW window 15, $W = .98,\ p = .005$. Given that some of the DTW data are not normally distributed, nonparametric Kruskal-Wallis tests were used to compare the four groups.

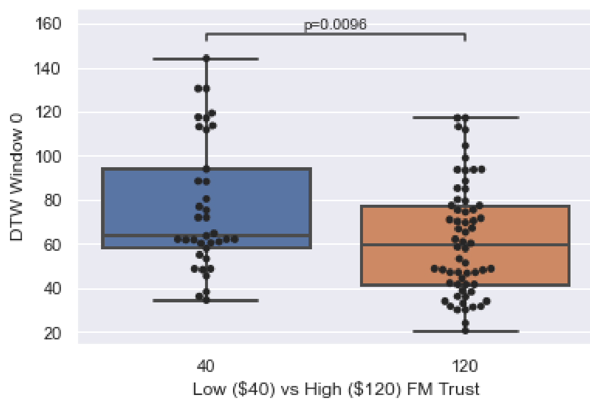
The results of the Kruskal-Wallis test of ranked DTW distances with a window width 0 indicated that the null cannot be rejected, $H(3) = 6.61,\ p = .09$. The results of the test for ranked DTW distances with a window width of 5 indicated that at least one trust level group has significantly different ranked DTW distances, $H(3) = 9.14,\ p = .03$. The results of the test for ranked DTW distances with a window width of 10 ($H(3) = 9.39,\ p = .02$) and 15 ($H(3) = 8.52,\ p = .04$) are also significant. Therefore, comparisons of three out of four DTW window widths found at least one trust level group had significantly different ranked DTW distances. Post hoc pairwise comparisons for ranked DTW distances of window widths 5, 10, and 15 were conducted using Dunn's test and Bonferroni corrected p-values, reported in the following sections.

2. 1. 1. ***DTW: Window 0***

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (alpha level .10 adjusted to .0167 (.1/6) and .05 to .0083 (.05/6)) found that the median DTW distances with window width 0 of the high trust group were smaller than those of the low trust group ($U =$

$1472.0,\ p = .0096$), as hypothesized. This reveals that high trust participants exhibited more PS than low trust participants. Box plots of the distributions and a comparison of the DTW distances with window width 0 of the low and high trust individuals and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 20.

Figure 20: Box Plots of Distributions and Comparison of Low ($40) vs High ($120) Trust Groups – DTW Window Width 0



2. 1. 2. ***DTW: Window 5***

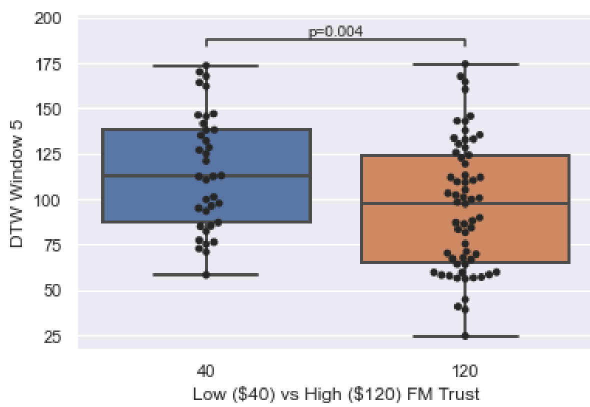
The results of the Kruskal-Wallis test for hypothesis 3 indicated there was a statistically significant difference between the trust groups of the ranked DTW distances with window width 5. A post hoc Dunn's test with Bonferroni corrected p-values found no statistically significant differences between the trust groups, as shown in Table 16.

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (.05 adjusted to .0083) indicated that the median DTW distances with window width 5 of the high trust group were smaller than those of the low trust group, $U = 1514.0,\ p = .004$. This reveals that high trust participants exhibited more PS than low trust participants, as hypothesized. Box plots of the distributions and a comparison of the DTW distances with window width 5 of the low and high trust individuals and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 21.

Table 16: Dunn's Test Results with Bonferroni Corrected P-Values – DTW Window Width 5

Decision	No Trust	Low Trust	Medium Trust	High Trust
No Trust	1.00	0.30	0.77	1.00
Low Trust	0.30	1.00	1.00	0.057
Medium Trust	0.77	1.00	1.00	0.18
High Trust	1.00	0.057	0.18	1.00

Figure 21: Box Plots of Distributions and Comparison of Low ($40) vs High ($120) Trust Groups – DTW Window 5



2. 1. 3. ***DTW: Window 10***

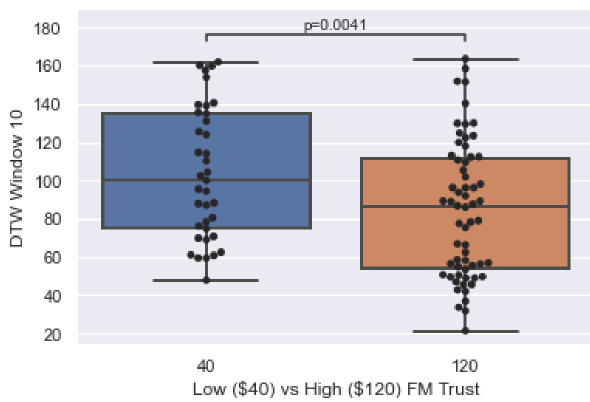
The results of the Kruskal-Wallis test for hypothesis 3 indicated there was a statistically significant difference between the trust groups of the ranked DTW distances with window width 10. Dunn's test with Bonferroni corrected p-values found no statistically significant differences between the trust groups, as shown in Table 17.

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (.05 adjusted to .0083) indicated that the median DTW distances with window width 10 of the high trust group were smaller than those of the low trust group, $U = 1513.0,\ p = .004$. This reveals that participants who displayed high trust exhibited more PS with their partner than those who displayed low trust, as hypothesized. Box plots of the distributions and a comparison of the DTW distances with window width 10 of the low and high trust individuals and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 22.

Table 17: Dunn's Test with Bonferroni Corrected P-Values – DTW Window Width 10

Decision	No Trust	Low Trust	Medium Trust	High Trust
No Trust	1.00	0.26	0.63	1.00
Low Trust	0.26	1.00	1.00	0.058
Medium Trust	0.63	1.00	1.00	0.17
High Trust	1.00	0.058	0.17	1.00

Figure 22: Box Plots of Distributions and Comparison of Low ($40) vs High ($120) Trust Groups – DTW Window 10



2. 1. 4. ***DTW: Window 15***

The results of the Kruskal-Wallis test for hypothesis 3 indicated there was a statistically significant difference between the trust level for the ranked DTW distances with window width

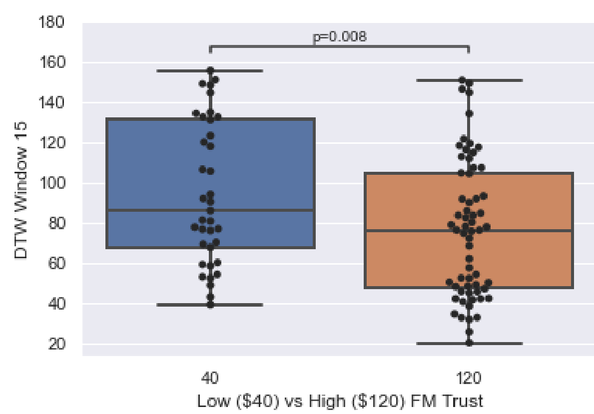
15. Dunn's test with Bonferroni corrected p-values found no statistically significant differences between the trust groups, as shown in Table 18.

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (.05 adjusted to .0083) indicated that the median DTW distances with window width 10 of the high trust group were smaller than those of the low trust group, $U = 1480.5,\ p = .008$. This reveals that high trust participants exhibited more PS with their partner than low trust participants, as hypothesized. Box plots of the distributions and a comparison of the DTW distances with window width 15 of the low and high trust individuals and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 23.

Table 18: Dunn's Test with Bonferroni Corrected P-Values – DTW Window Width 15.

Decision	**No Trust**	**Low Trust**	**Medium Trust**	**High Trust**
No Trust	1.00	0.43	0.69	1.00
Low Trust	0.43	1.00	1.00	0.09
Medium Trust	0.69	1.00	1.00	0.17
High Trust	1.00	0.09	0.17	1.00

Figure 23: Box Plots of Distributions and Comparison of Low ($40) vs High ($120) Trust Groups – DTW Window Width 15



2. 2. **Hypothesis 4**

Hypothesis 4 tested whether FMs who demonstrated more trust by sending *significantly* more money in the trust game than in the dictator game have smaller DTW distances than those who exhibited less trust by sending *slightly* more money. Three one-tailed Mann-Whitney U tests were conducted: median DTW distances of high versus medium, medium versus low, and low versus no trust participants. Alpha levels .05 and .10 were Bonferroni corrected to .0167 and .033, respectively. One-tailed Mann-Whitney U tests results with Bonferroni corrected p-values for window widths 0, 5, 10, and 15 are reported in Tables 19, 20, 21, and 22, respectively. No significant difference was found between the low and medium trust groups. The results indicated significant differences between the no versus low trust and medium versus high trust groups, which are discussed in the following two sections.

Table 19: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – DTW Window Width 0

Decision	**No Trust**	**Low Trust**	**Medium Trust**
Low Trust			
MWU statistic	1462.0		
p-value	0.84		
Medium Trust			
MWU statistic		665.0	
p-value		0.34	
High Trust			
MWU statistic			1295.5*
p-value			0.032*

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 20: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – DTW Window Width 5

Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	1291.0		
p-value	0.97		
Medium Trust			
MWU statistic		659.0	
p-value		0.38	
High Trust			
MWU statistic			1328.0*
p-value			0.018*

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 21: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – DTW Window Width 10

Decision	No Trust	Low Trust	Medium Trust
Low Trust			
MWU statistic	1281.0		
p-value	0.98		
Medium Trust			
MWU statistic		655.0	
p-value		0.38	
High Trust			
MWU statistic			1331.0*
p-value			0.017*

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

Table 22: One-Tailed Mann-Whitney U Tests of Ordinal Trust Groups – DTW Window Width 15

Decision	**No Trust**	**Low Trust**	**Medium Trust**
Low Trust			
MWU statistic	1315.0		
p-value	0.96		
Medium Trust			
MWU statistic		646.0	
p-value		0.42	
High Trust			
MWU statistic			1334.0**
p-value			0.016**

Adjusted p-values: 10% level = .033; 5% level = .0167. * and ** indicate significance at the 10% and 5% level, respectively.

2. 2. 1. ***No vs Low Trust***

In line with the findings in hypothesis 1, a one-tailed Mann-Whitney U test with Bonferroni corrected p-values (alpha level .05 adjusted to .0167 and .10 to .033) found that median DTW distances with widow widths 5 and 10 of the low trust participants were significantly greater than those of the no trust participants: window 5, $U = 1291.0,\ p = .029$; window 10, $U = 1281.0,\ p = .025$. Again, this was an unexpected result given the hypothesis that individuals who trusted their partner would have lower DTW distances, and thus greater synchrony, than those who did not trust. Box plots of the distributions and a comparison of the DTW distances with window widths 5 and 10 of the no and low trust participants and the corresponding one-tailed Mann-Whitney U tests p-values are shown in Figures 24 and 25, respectively.

Figure 24: Box Plots of Distributions and Comparison of No ($0) vs Low ($40) Trust – DTW Window Width 5

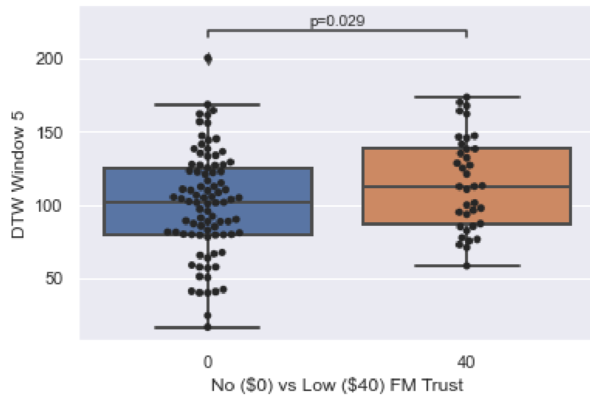
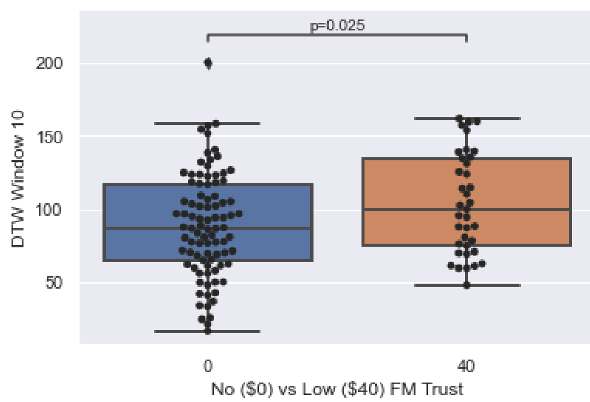


Figure 25: Box Plots of Distributions and Comparison of No ($0) vs Low ($40) Trust – DTW Window Width 10



2. 2. 2. ***Medium vs High Trust***

A one-tailed Mann-Whitney U test with Bonferroni corrected p-values (alpha level .10 adjusted to .033 and .05 to .0167) indicated that median DTW distances with widow widths 0, 5, 10, and 15 of the high trust group were lower than those of the medium trust group: window 0, $U = 1295.0,\ p =$ .032; window 5, $U = 1328.0,\ p =$ .018; window 10, $U = 1331.0,\ p =$ .017; window 15, $U = 13340,\ p =$ .016. Therefore, individuals exhibiting high trust had greater PS than those with medium trust, consistent with my hypothesis. Box plots of the distributions and a comparison of the DTW distances with windows 0, 5, 10, and 15 of the medium and high trust participants and the corresponding one-tailed Mann-Whitney U tests p-values are shown in Figures 27, 28, 29 and 30, respectively.

Figure 26: Box Plots of Distributions and Comparison of Medium ($80) vs High ($120) Trust – DTW Window Width 0

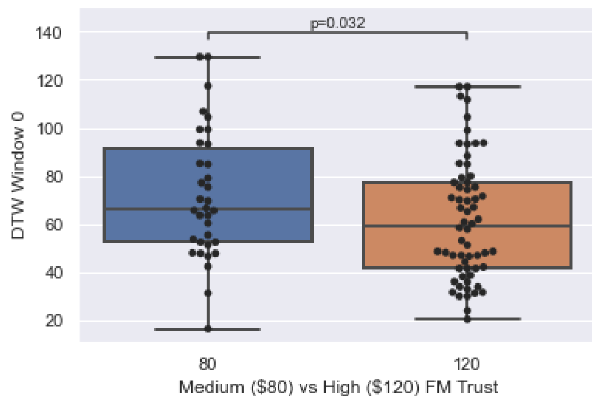


Figure 27: Box Plots of Distributions and Comparison of Medium ($80) vs High ($120) Trust – DTW Window Width 5

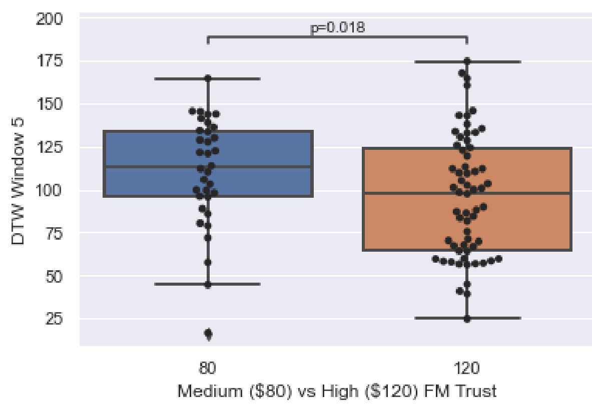


Figure 28: Box Plots of Distributions and Comparison of Medium ($80) vs High ($120) Trust – DTW Window Width 10

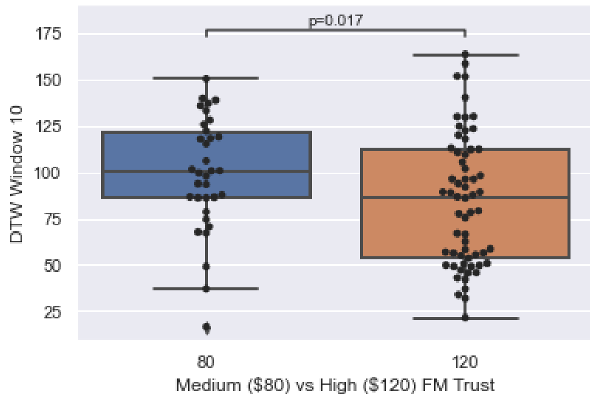
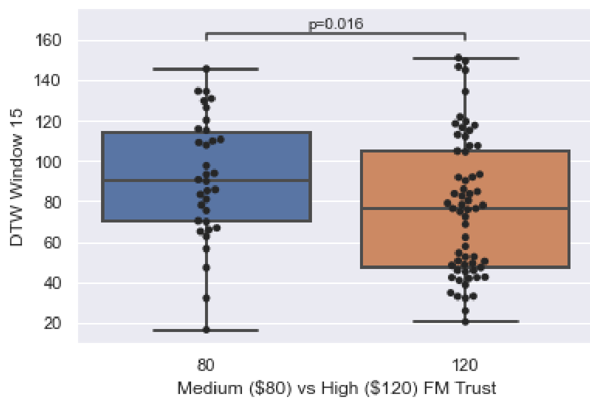


Figure 29: Box Plots of Distributions and Comparison of Medium ($80) vs High ($120) Trust – DTW Window Width 15



2. **Second Mover (SM) Results**

Trustworthiness was measured as the SMs' decision, r_{TG}, in response to the FM showing maximum trust by sending $120. Since the SM had three options, keep all the money (return $0), return the $120 the FM sent, or split the total equally, I define trustworthy behavior as splitting the total equally by returning $240 to the FM (shown in the second row of Table 23). The decisions to keep all the money and return $120 are defined as untrustworthy, i.e., betrayal, as these decisions either make the FM worse off or no better off than had they not extended trust, respectively (Coleman, 1990; Evans & Krueger, 2011; Hong & Bohnet, 2007; Thielmann & Hilbig, 2015). There was a positive correlation between FM trust and SMs' decisions ($r(220) = .39, p < .001$). The results of the SMs' decisions are reported in Table 23 and Figure 30. Figures 31 and 32 show SMs' decisions by session and by interaction, respectively.

Table 23: Second Mover Decisions by Trustworthiness

Definition	s_{TG}	r_{TG}	Count
Trustworthy, n = 158	$120	$240	158
Untrustworthy (Betrayal), n = 64	$120	$120	26
	$120	$0	38

Figure 30: Distribution of Second Mover Decisions

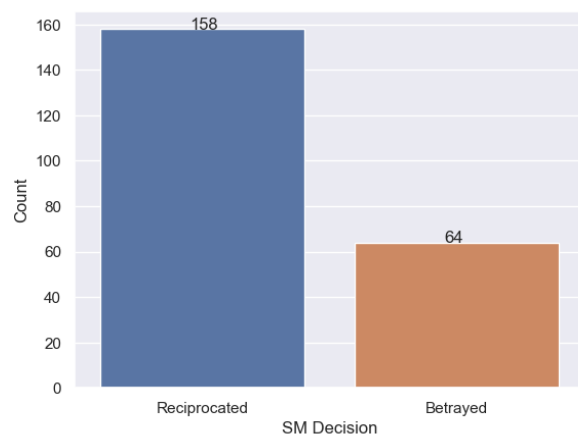


Figure 31: Distribution of Second Mover Decisions by Session

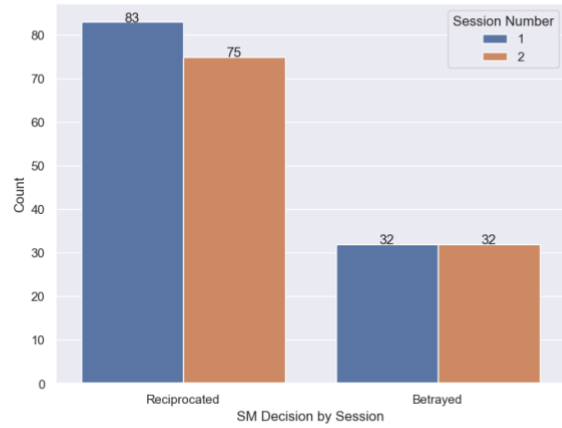
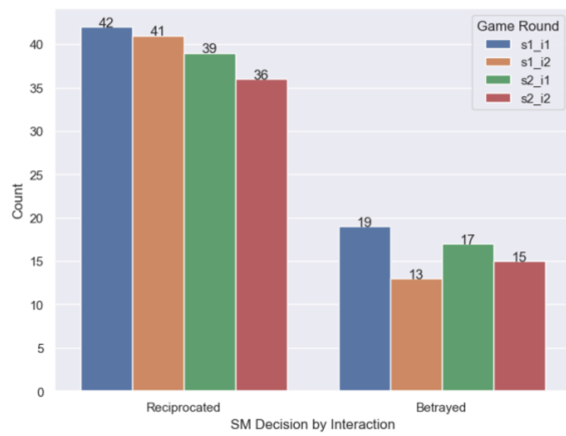


Figure 32: Distribution of Second Mover Decisions by Interaction



1. ***SM ISC Results***

Trustworthy and untrustworthy participants' ISCs were calculated with window sizes 2, 3, 5, and 8, each with a lag of 1. Summary statistics of SM ISCs with window sizes 2, 3, 5, and 8 with lags of 1 are reported in Tables 24, 25, 26, and 27, respectively. Trustworthy and untrustworthy participants' ISCs were compared using Mann-Whitney U tests for hypothesis 5, which are reported in the following section.

Table 24: Summary Statistics of Second Mover ISCs with Window Size 2, Lag 1

Decision	**Trustworthy**	**Untrustworthy**
Min	0.08	.08
Mean	0.45	.49
Median	0.43	.50
Max	0.77	0.73
SD	0.15	0.14
SE	0.01	.02
n	158	64

Table 25: Summary Statistics of Second Mover ISCs with Window Size 3, Lag 1

Decision	**Trustworthy**	**Untrustworthy**
Min	0.32	0.51
Mean	0.80	0.81
Median	0.82	0.82
Max	1.00	0.98
SD	0.12	0.11
SE	0.01	0.01
n	158	64

Table 26: Summary Statistics of Second Mover ISCs with Window Size 5, Lag 1

Decision	**Trustworthy**	**Untrustworthy**
Min	0.41	0.35
Mean	0.74	0.74
Median	0.75	0.77
Max	0.93	0.94
SD	0.10	0.12
SE	0.01	0.01
n	158	64

Table 27: Summary Statistics of Second Mover ISCs with Window Size 8, Lag 1

Decision	Trustworthy	Untrustworthy
Min	0.24	0.17
Mean	0.69	0.70
Median	0.70	0.74
Max	0.94	0.95
SD	0.12	0.14
SE	0.01	0.02
n	158	64

1. 1. **Hypothesis 5**

Hypothesis 5 tested whether SMs who showed trustworthy behavior by splitting the total in response to their partner sending $120 in the first round have higher ISCs than participants who did not exhibit trustworthy behavior by either choosing to return $0 or the amount sent. Results of the one-tailed Mann-Whitney U tests for ISCs of window sizes 2, 3, 5, and 8, each with a lag of 1, are shown in Tables 28, 29, 30 and 31, respectively. A one-tailed Mann-Whitney U test indicated that median ISCs with window size 2, lag 1 of the untrustworthy participants were higher than those of the trustworthy participants, $U = 4095.0,\ p = .013$. Therefore, the participants who betrayed their partner exhibited greater PS than those who were trustworthy, according to their ISCs. This result was unexpected given the hypothesis that those who were trustworthy would have higher ISCs than those who were untrustworthy. It's possible the trustworthy participants were less engaged, having decided they would reciprocate in kind if shown trust, whereas the untrustworthy participants were more engaged in an effort to deceive their partner. PS may have been the result of a strong desire or need for untrustworthy participants to deceive their partner regarding their own trustworthiness. Box plots of the distributions and a comparison of ISCs with window 2, lag 1 of trustworthy and untrustworthy participants and the corresponding one-tailed Mann-Whitney U test p-value are shown in Figure 33. No statistically significant differences were found between participant groups with tests of the other window sizes.

Table 28: One-Tailed Mann-Whitney U Test of Second Mover Decisions – ISC Window 2, Lag 1

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	4095.0
p-value	0.99

* and ** indicate significance at the 10% and 5% level, respectively.

Figure 33: Box Plots of Distributions and Comparison of Trustworthy vs Untrustworthy Second Movers – ISC Window Size 2, Lag 1

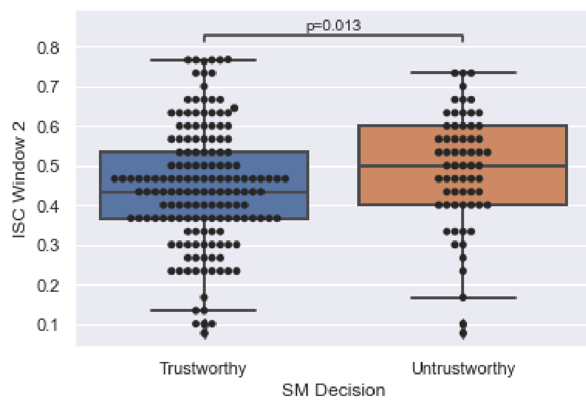


Table 29: One-Tailed Mann-Whitney U Test of Second Mover Decisions – ISC Window Size 3, Lag 1

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	4914.0
p-value	0.63

* and ** indicate significance at the 10% and 5% level, respectively.

Table 30: One-Tailed Mann-Whitney U Test of Second Mover Decisions – ISC Window Size 5, Lag 1

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	4653.0
p-value	0.82

* and ** indicate significance at the 10% and 5% level, respectively.

Table 31: One-Tailed Mann-Whitney U Test of Second Mover Decisions – ISC Window Size 8, Lag 1

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	4564.0
p-value	0.87

* and ** indicate significance at the 10% and 5% level, respectively.

2. ***SM DTW Results***

Trustworthy and untrustworthy participants' DTW distances were calculated using the Euclidean distance and four windows widths: 0, 5, 10, and 15. Summary statistics of SM DTW distances with window widths 0, 5, 10, and 15 are reported in Tables 32, 33, 34, and 35, respectively. Trustworthy and untrustworthy participants' DTW distances were compared using Mann-Whitney U tests for hypothesis 6, which are reported in the following section.

Table 32: Summary Statistics of Second Mover DTW Distances with Window Width 0

Decision	**Trustworthy**	**Untrustworthy**
Min	16.47	16.47
Mean	67.44	72.36
Median	63.62	66.89
Max	144.15	200.55
SD	27.61	30.70
SE	2.20	3.84
n	158	64

Table 33: Summary Statistics of Second Mover DTW Distances with Window Width 5

Decision	**Trustworthy**	**Untrustworthy**
Min	16.34	16.47
Mean	103.14	103.52
Median	104.62	101.20
Max	174.53	200.55
SD	35.23	34.22
SE	2.80	4.28
n	158	64

Table 34: Summary Statistics of Second Mover DTW Distances with Window Width 10

Decision	**Trustworthy**	**Untrustworthy**
Min	16.34	16.47
Mean	91.40	92.30
Median	93.56	86.51
Max	163.57	200.55
SD	35.41	34.47
SE	2.82	4.31
n	158	64

Table 35: Summary Statistics of Second Mover DTW Distances with Window Width 15

Decision	Trustworthy	Untrustworthy
Min	16.34	16.47
Mean	83.29	85.69
Median	81.04	81.07
Max	155.82	200.55
SD	34.55	34.58
SE	2.75	4.32
n	158	64

2. 1. **Hypothesis 6**

Hypothesis 6 tested whether SMs who showed trustworthy behavior by splitting the total in response to their partner sending $120 in the first round have smaller DTW distances than participants who did not exhibit trustworthy behavior by either choosing to return $0 or the amount sent. Results of the one-tailed Mann-Whitney U tests for DTW distances of window widths 0, 5, 10, and 15 are shown in Tables 36, 37, 38 and 39, respectively. No statistically significant differences were found between the trustworthy and untrustworthy groups; therefore, the null cannot be rejected. Thus, it's possible PS plays a larger role in FM behavior than in SM behavior.

Table 36: One-Tailed Mann-Whitney U Test of Second Mover Decisions – DTW Window Width 0

Decision	Trustworthy
Untrustworthy	
MWU statistic	4606.5
p-value	0.15

* and ** indicate significance at the 10% and 5% level, respectively.

Table 37: One-Tailed Mann-Whitney U Test of Second Mover Decisions – DTW Window 5

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	5105.0
p-value	0.54

* and ** indicate significance at the 10% and 5% level, respectively.

Table 38: One-Tailed Mann-Whitney U Test of Second Mover Decisions – DTW Window 10

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	5076.0
p-value	0.52

* and ** indicate significance at the 10% and 5% level, respectively.

Table 39: One-Tailed Mann-Whitney U Test of Second Mover Decisions – DTW Window 15

Decision	**Trustworthy**
Untrustworthy	
MWU statistic	4878.0
p-value	0.34

* and ** indicate significance at the 10% and 5% level, respectively.

VI. Discussion

This research aimed to determine whether participants who cooperated in the trust game experienced PS with their partner. Participant behavior in the trust game resulted in four trust level groups: no, low, medium, and high trust. This section briefly concludes with a discussion of the results, limitations, and suggestions for future research.

1. First Mover (FM) Results

The analyses of participants' FM decisions indicated that some groups experienced significantly more PS than others, but the results were not comprehensive nor conclusive. A comparison of the four trust groups' ISCs with window 2, lag 1 found that the 89 individuals who did not trust their partner exhibited significantly more PS than the 37 participants who trusted their partner with $40. This finding was confirmed with a similar comparison of the four trust groups' DTW distances with window widths 5 and 10. As discussed in 2.2.3. Mechanisms of Physiological Synchrony, Levenson and Gottman (1983) found that distressed couples engaged in a heated argument exhibited significantly more PS than non-distressed couples. It's possible that the no trust participants were more tuned to their partner in response to feeling more hesitant or suspicious, resulting in paying closer attention, i.e., being more vigilant. This may explain why the participants who did not trust exhibited greater PS than the low trust participants. Alternatively, it is possible the participants who did not trust decided not to send any money in the games prior to the interaction, and in turn engaged with their partner without also trying to assess their trustworthiness.

It is also plausible that the no trust participants were making an effort to deceive their partners, and thus more engaged in the interaction. Research has found that deception is more cognitively demanding than making an unintentional error (Lee et al., 2009). The no trust participants may have been paying closer attention to their partner in order to appear trustworthy and deceive them. Finally, this finding may be the result of these participants having a negative interaction with their partner ((Levenson & Gottman, 1983), but analysis of the self-report data is necessary to confirm this. Nevertheless, the discovery was unexpected; it implies that the no trust participants were more attuned to their partner despite the fact they did not entrust them with any of their endowment in the trust game.

Previous research has shown trust is associated with lower arousal Abbott, Middlemiss, Bruce, Smailes, & Dudley, 2018; Kéri & Kiss, 2011), but it is unclear which group experienced less arousal. To examine participants' overall arousal levels, the average SCLs of the two groups were calculated and compared. Shapiro-Wilks tests indicated that the SCLs of the no trust group were not normally distributed, $W = .68,\ p < .001$ (low trust group SCLs, $W = .98,\ p = .80$). Therefore, a one-tailed Mann-Whitney U test was used, with the no trust group hypothesized to have a higher median SCL than the low trust group. The results were not statistically significant, $U = 1539.0,\ p = \ .72$. Thus, the hypothesis that the no trust participants were more aroused cannot be confirmed and remains merely a conjecture.

Additional findings related to participants' FM decisions were in accordance with the hypotheses that trusting participants would exhibit more PS. DTW analyses of window widths 5 and 10 showed that the 62 individuals who trusted their partner with their entire endowment ($120) exhibited greater PS than the 37 participants who only trusted their partner with $40. Shapiro-Wilks tests indicated both groups were normally distributed: low trust group SCLs, $W = .98,\ p = .80$, high trust group SCLs, $W = .96,\ p = .07$. A two-tailed t-test of the overall arousal of the two groups was not statistically significant, $t = 1178.5,\ p = .83$. Thus, no conclusions can be made regarding their comparative arousal levels. Finally, the DTW distances of all widow widths showed the 62 high trust individuals exhibited greater PS than the 34 medium trust individuals. A Shapiro-Wilks test indicated the medium trust group SCLs were not normally distributed, $W = .26,\ p < .001$. A two-tailed Mann-Whitney U test of the arousal of these two groups was not statistically significant, $U = 1050.0,\ p = .98$. Thus, again, no conclusions can be made regarding their comparative arousal levels.

2. **Second Mover (SM) Results**

The ISC analyses of participants' SM decisions found that untrustworthy participants exhibited greater PS than trustworthy participants. Again, this finding may be explained by participants making an effort to deceive their partner, or a negative partner interaction. Although only ISCs with window size 2, lag 1 indicated this finding, this result invalidated the general hypothesis that trustworthy participants would exhibit more PS than untrustworthy participants. Unfortunately, none of the DTW analyses were statistically significant.

Shapiro-Wilks tests indicated that the SCLs of the trustworthy group were not normally distributed, $W = .09,\ p < .001$ (untrustworthy group SCLs, $W = .97,\ p = .08$). A one-tailed Mann-Whitney U test found that the untrustworthy participants experienced significantly lower arousal than the trustworthy participants, $U = 3883.0,\ p < .01$. This finding contradicts previous research showing individuals who trust are less aroused. Although, other research indicates that the relationship between arousal and trust are person-dependent (Potts, McCuddy, Jayan, & Porcelli, 2019; Song, Colasante, & Malti, 2020). Participants who were untrustworthy may have been paying closer attention to their partner during the interaction in an effort to deceive. At the same time, it's possible these untrustworthy participants were less aroused and thus not compelled to reciprocate trust. Regardless, this is an interesting discovery that would require additional research to confirm**.**

Based on the FM and SM results, I suppose that the relationship between arousal and PS follows the Yerkes-Dodson law, a principle that states, “the relationship between arousal and behavioral performance can be linear or curvilinear, depending on the difficulty of the task” (p. 1, Diamond, Campbell, Park, Halonen, & Zoladz, 2007), shown in Figure 34. I suspect arousal and PS are curvilinear, such that a moderate amount of arousal is necessary for PS. If participants experience low arousal, it could be the result of a lack of interest; if participants experience hyperarousal, possibly due to feelings of suspicion, they are more likely to be out of synch with their partner. The ISC findings that the untrustworthy participants exhibited less arousal should be tested in future research.

Figure 34: Diagram of the Yerkes-Dodson Law, from Diamond et al., 2006.

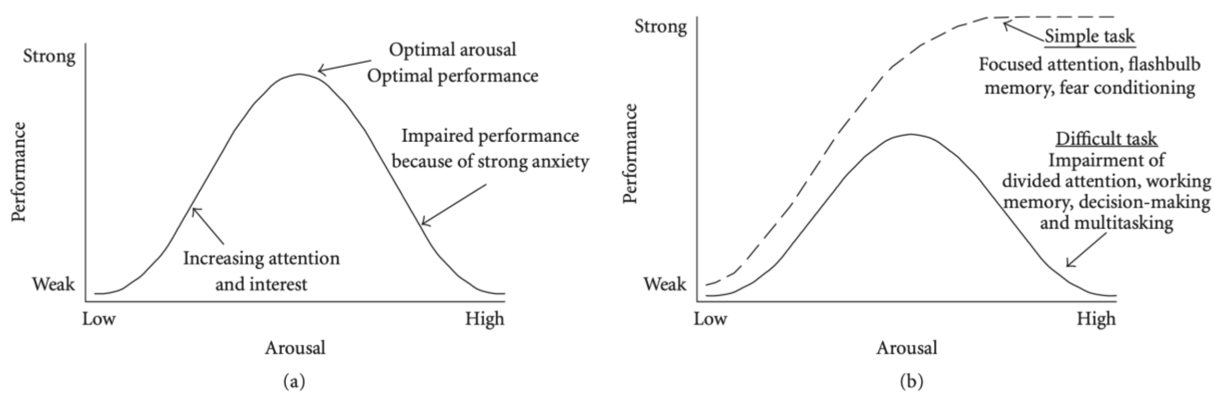


FIGURE 2: A comparison of the Hebbian version of the Yerkes-Dodson law, as it has been commonly represented for the past 50 years (a), and the original version, based on the actual findings and theorizing of Yerkes and Dodson ([38]; (b)). The Hebbian version incorrectly states that high levels of stress, anxiety, or motivation produce a monolithic impairment of performance. The original version based on the actual [38] Yerkes-Dodson findings takes into account the finding that strong emotionality can enhance performance under “simple” learning conditions, such as when learning involves focused attention on a restricted range of cues, and impairs performance under more complex or challenging learning situations, such as in divided attention, multitasking, and working memory tasks. Graph (a) is adapted from 5 decades of publications and books, for example, Hebb [53], Loftus [54], and Radvansky [55].

3. **Comparing ISC and DTW Results**

The purpose of this analysis was to determine whether PS between two individuals while they discuss trusting behaviors will predict their subsequent decisions in a high-stakes trust game. The methods used to measure a statistic of PS, ISC and DTW, were selected because cross-correlation is the most frequently used method and DTW has not been applied to PS research.

The results reveal differences between the two methods. First, ISC requires specifying four parameters: window size, window increment, maximum lag, and lag increment; although the analysis can also be done with only window size and maximum lag and setting both window and lag increment to 1, as it was here. These parameters remain fixed across all participants, whereas DTW calculates minimum distances within a given window. The dynamic nature of this calculation makes it preferable to cross-correlation methods such as ISC.

Next, the methodology of ISC and DTW suggest the two variables are inversely related. Interestingly, an analysis of the correlations between all ISCs and DTW distances finds no or slightly negative correlation between the ISCs and DTW distances, as shown in Table 41. Comparing the no trust group's ISCs and DTW distances shows a slightly negative correlation between the two variables, shown in Table 42. The low trust group's correlation matrix, shown in Table 43, indicates a no or slightly negative correlation between the ISCs and DTW distances. The correlation matrix of the medium trust group's ISCs and DTW distances show the most variation, with correlations ranging from slightly positive to negative. Finally, the correlations of the high trust group's ISCs and DTW distances were slightly negative, as shown in Table 44. A comparison of the second mover groups' ISCs and DTW distances shows the correlations between the two variables are slightly negative for both the trustworthy and untrustworthy groups, as shown in Tables 45 and 46, respectively.

Table 40: ISC and DTW Correlation Matrix, All Data

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.14	-0.11	-0.16	-0.18
DTW Window 5	-0.09	-0.10	-0.11	-0.15
DTW Window 10	-0.09	-0.08	-0.11	-0.14
DTW Window 15	-0.11	-0.10	-0.13	-0.15

Note: All correlations have Mann-Whitney U test p-values of $p < .001$.

Table 41: No Trust Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.35	-0.39	-0.31	-0.29
DTW Window 5	-0.23	-0.27	-0.21	-0.24
DTW Window 10	-0.23	-0.27	-0.23	-0.26
DTW Window 15	-0.25	-0.31	-0.26	-0.27

Note: All correlations have Mann-Whitney U test p-values of $p < .001$.

Table 42: Low Trust Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.13	-0.06	-0.14	-0.28
DTW Window 5	-0.09	0.02	0.06	-0.20
DTW Window 10	-0.09	0.01	0.03	-0.21
DTW Window 15	-0.13	-0.31	-0.02	-0.25

Note: All correlations have Mann-Whitney U test p-values of p < .001.

Table 43: Medium Trust Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	0.02	-0.04	-0.22	-0.02
DTW Window 5	0.09	0.03	-0.23	0.08
DTW Window 10	0.11	0.03	-0.22	0.08
DTW Window 15	0.07	-0.001	-0.24	0.03

Note: All correlations have Mann-Whitney U test p-values of p < .001.

Table 44: High Trust Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.09	-0.10	-0.22	-0.02
DTW Window 5	-0.08	-0.16	-0.16	-0.15
DTW Window 10	-0.08	-0.14	-0.16	-0.12
DTW Window 15	0.07	-0.001	-0.24	0.03

Note: All correlations have Mann-Whitney U test p-values of p < .001.

Table 45: Trustworthy Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.16	-0.09	-0.15	-0.18
DTW Window 5	-0.11	-0.11	-0.13	-0.16
DTW Window 10	-0.10	-0.10	-0.12	-0.14
DTW Window 15	-0.14	-0.12	-0.13	-0.15

Note: All correlations have Mann-Whitney U test p-values of $p < .001$.

Table 46: Untrustworthy Group's ISC and DTW Correlation Matrix

	ISC Window 2	ISC Window 3	ISC Window 5	ISC Window 8
DTW Window 0	-0.14	-0.17	-0.18	-0.18
DTW Window 5	-0.05	-0.06	-0.08	-0.12
DTW Window 10	-0.05	-0.05	-0.09	-0.14
DTW Window 15	-0.07	-0.07	-0.12	-0.16

Note: All correlations have Mann-Whitney U test p-values of $p < .001$.

4. Limitations

The first limitation pertains to the number of data points associated with participants' decisions; the same ISCs and DTW distances were used for both the FM and SM analyses. Participants interacted with their partner once but made two decisions – as both FM and SM – based on this interaction. Each participant's physiological response may have been influenced by the contemplation of just one or both decisions. Thus, it is unclear whether the interaction captures participants' feelings of trust, trustworthiness, or both.

Another limitation is that the data analyzed were from an experiment designed to measure several hormones and physiological responses, including ECG and EEG, not aimed at examining PS. Participants had their blood drawn more than once and were connected to several electrodes, which may have impeded their ability to experience PS. Thus, it is possible a simpler experiment, focused solely on measuring PS, would produce more conclusive results. Furthermore, due to the lengthy protocol, the experiment took place on two different days, one week apart, which may have impacted participants' dispositions and thus their physiology.

The challenge of measuring participants' PS in the trust game may be confounded by their risk aversion and risk preferences (Chetty, Hofmeyr, Kincaid, & Monroe, 2021). Because reactions to risky decisions are captured by autonomic responses (Bechara et al., 1997; Figner et al., 2009), it is possible participants' SCLs were affected by the high stakes of this trust game. In other words, detection may have been more difficult because of the large sums of money at stake. More specifically, it is possible the rate of change of individuals' SCLs was impacted by their responses to risk taking. Furthermore, if individuals' reactions to risk taking are strongly reflected in their autonomic responses, it may be that only participants with similar risk preferences show synchrony when large sums are at stake. Thus, it is possible it would be easier to detect PS when the games are played with small amounts of money, i.e., when less risk is involved. Future research could test this theory, and measure and control for participants' risk preferences.

Regarding the data analysis, a key limitation is the absence of validation by null hypothesis testing the ISCs and DTW distances to confirm the presence of synchrony (McAssey et al., 2013; Moulder, Boker, Ramseyer, & Tschacher, 2018; Palumbo et al., 2017). According to Moulder et al. (2018), "many methods for assessing synchrony rely on standard null-hypothesis testing. However, in behavioral time series, standard null-hypothesis testing methods tend to show significant synchrony between two individuals even when there is in reality no synchrony to be found (Ramseyer & Tschacher, 2010)" (p. 4). Due to the nature of the data and interaction, participants' SCLs will always show some degree of synchrony (McAssey et al., 2013). Many studies on PS either bootstrap participant data, as in Golland et al. (2014), or generate surrogate data by randomly pairing participants, creating new dyads, and testing whether synchrony in the real dyads is significantly different from the simulated data (McAssey et al., 2013; Moulder et al., 2018; Palumbo et al., 2017). Comparisons with randomized dyads is necessary to rule out the chance of spurious correlations (Palumbo et al., 2017). Another approach is to compare "periods when participants interacted to periods when they did not" (Palumbo et al., 2017, p. 110). The results could be validated by using one or more of these methods to null hypothesis test the synchrony measures.

In addition, regression analyses could be conducted using the survey measures, including personality measures and demographics. These data were not analyzed and may provide valuable insights. For example, Behrens, Snijdewint, et al. (2020) conducted a multilevel linear regression with synchrony as the dependent variable and dyads included as a random intercept. Finally, this analysis could have focused solely on the first or last participant interactions, rather than all 222 interactions. It is possible synchrony varied in different rounds, which could be tested using a regression including the game round as a control variable. For example, synchrony may have been higher in the first or last round. Further analyses would be necessary to determine whether differences in synchrony exist between rounds.

5. **Future Research**

Considering this is the first study measuring PS in the trust game, significant research should be conducted to confirm the presence of synchrony and establish experiment design standards. The first design suggestion is to have a longer interaction period, which may provide better results. For example, Danyluck and Page-Gould (2018) measured participants' sympathetic reactivity over the course of a five-minute interaction. In addition, the experiment design should focus on measuring PS and determining whether individuals who do not trust exhibit synchrony due to negative interactions. Another approach would be to isolate the portion of the interaction when participants made promises to each other and examine the SCRs during these periods. As discussed in the previous section, future research should also test whether risk aversion and risk preferences influence participants' PS.

In an experiment using the prisoner's dilemma game, Verplaetse, Vanneste, and Braeckman (2007) found that participants were able to accurately predict noncooperative individuals from their photograph only if the picture was taken during the decision-making moment (versus photos taken prior to the game and during a practice round). Based on their finding, future research could analyze participants' PS during decision making. SCRs could also be examined during these decision-making periods.

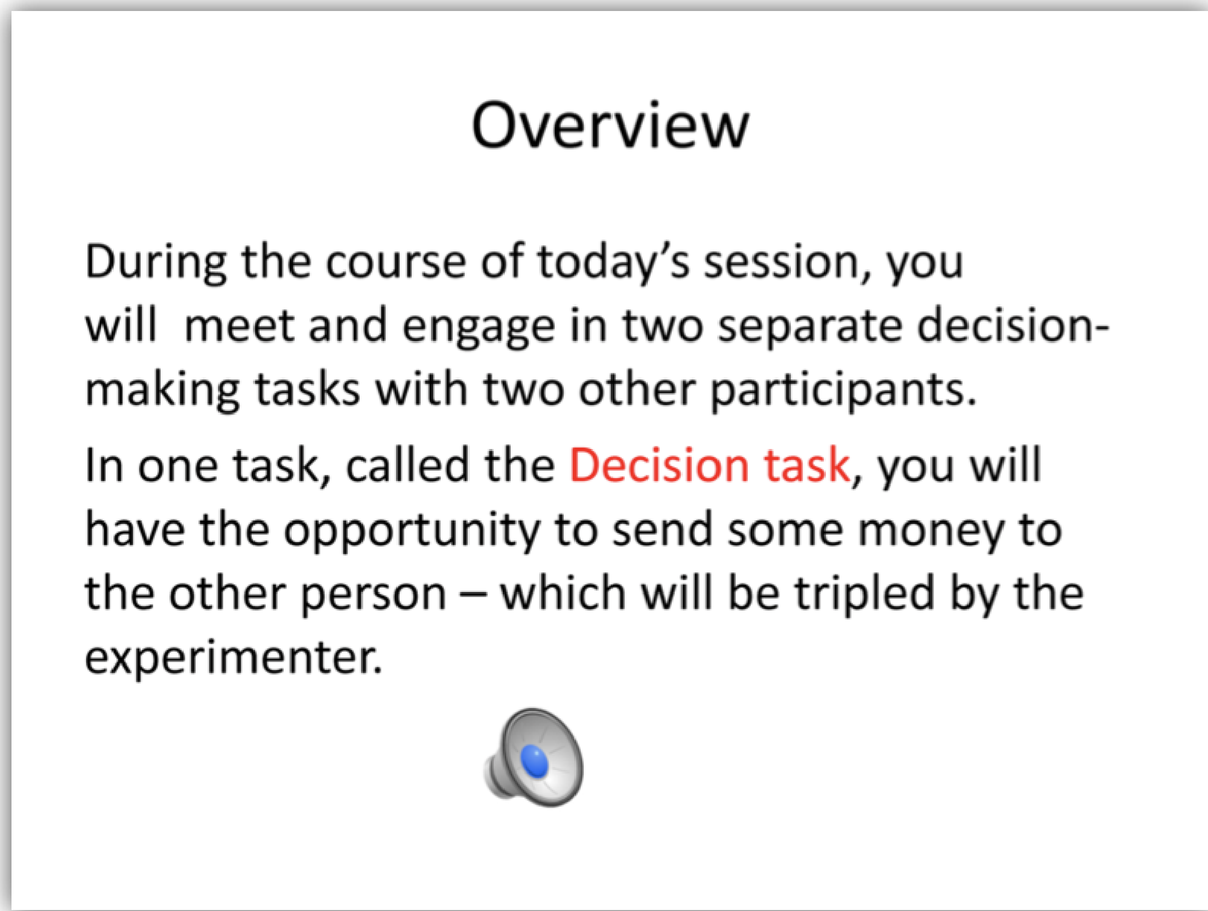
Most importantly, future researchers should collect one period of physiology data per decision to ensure a one-to-one relationship between the physiological data and participants' game decisions.

Appendix

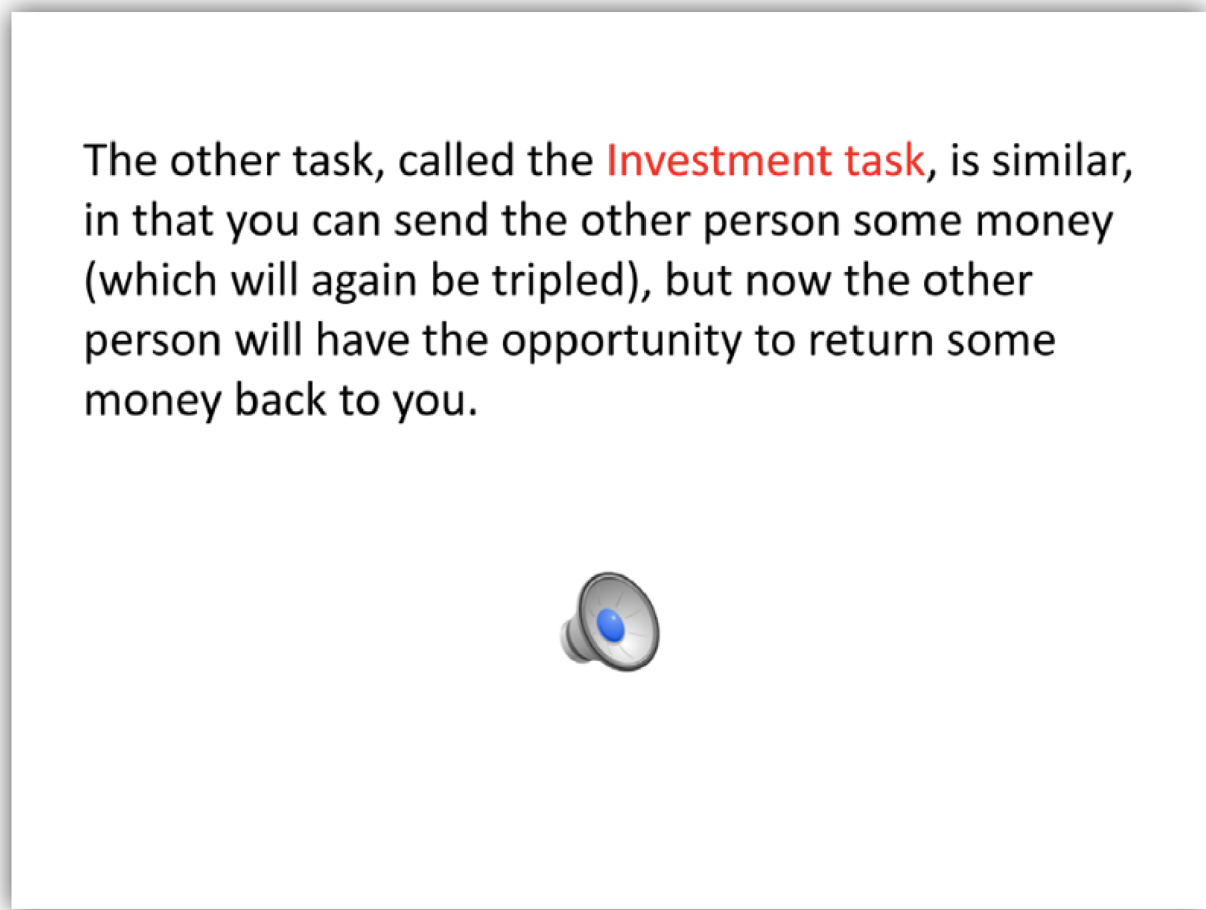
Subjects were given instructions about the economic games via a set of viewgraphs with audio instructions on laptop computers. This appendix contains copies of the viewgraphs and transcripts of the audio instructions for four instructional segments: Overview of the games, Dictator Game instructions, Investment Game instructions, and the randomized selection of the decision for payoff (the "Reveal").

OVERVIEW INSTRUCTIONS

> ***Narration:*** *During the course of today's session, you will meet and engage in two separate decision-making tasks with two other participants.*
>
> *In one task, called the Decision task, you will have the opportunity to send some money to the other person – which will be tripled by the experimenter*

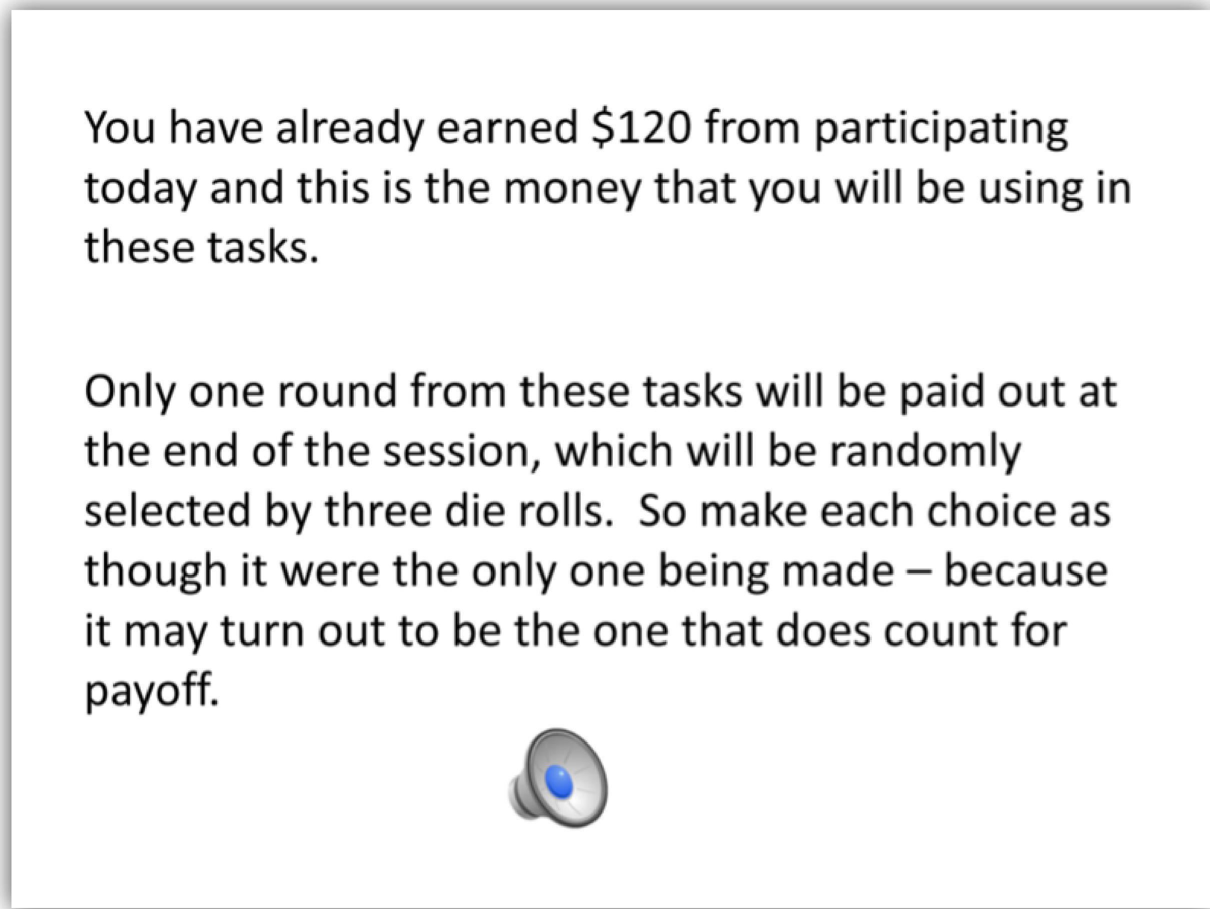


The other task, called the Investment task, is similar, in that you can send the other person some money (which will again be tripled), but now the other person will have the opportunity to return some money back to you.



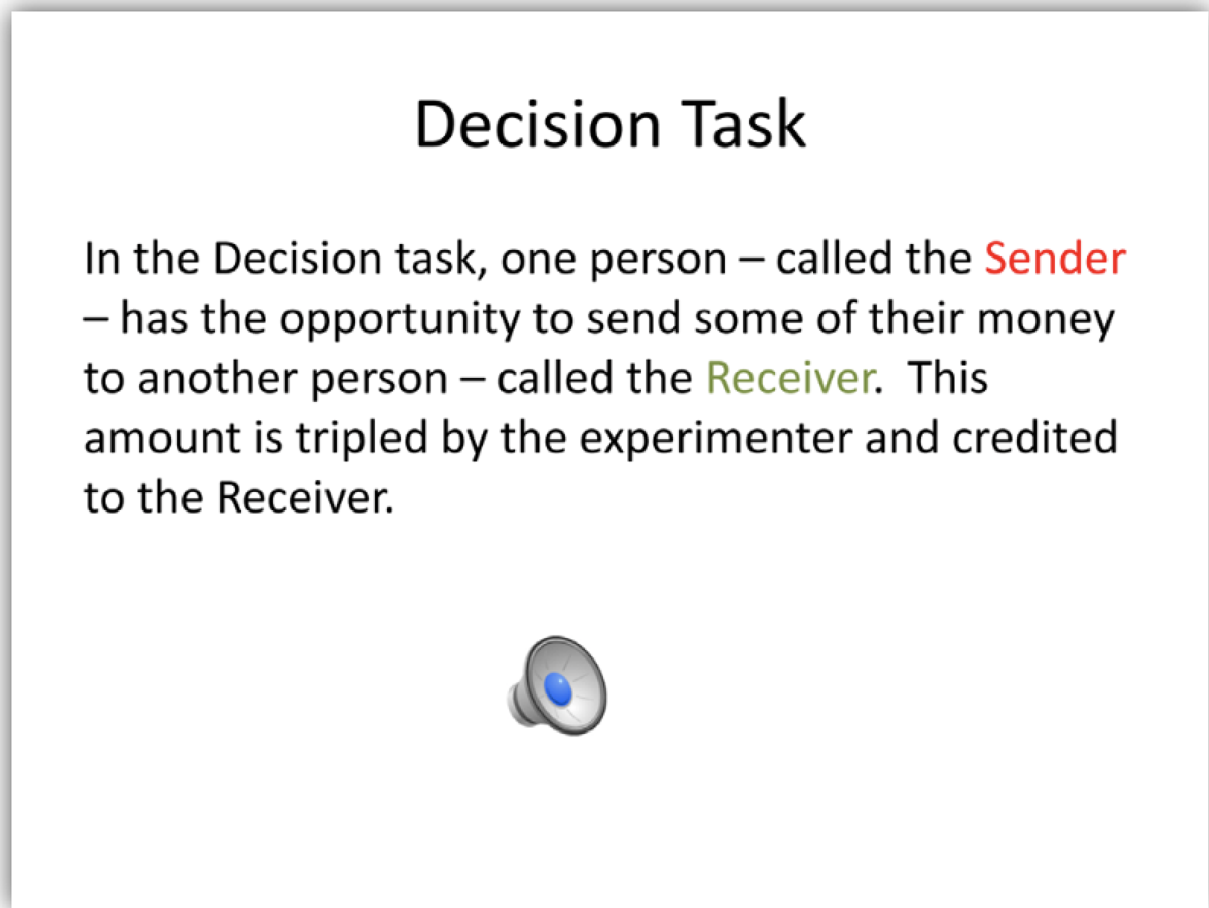
You have already earned $120 from participating today and this is the money that you will be using in these tasks.

Only one round from these tasks will be paid out at the end of the session, which will be randomly selected by three die rolls. So make each choice as though it were the only one being made – because it may turn out to be the one that does count for payoff.



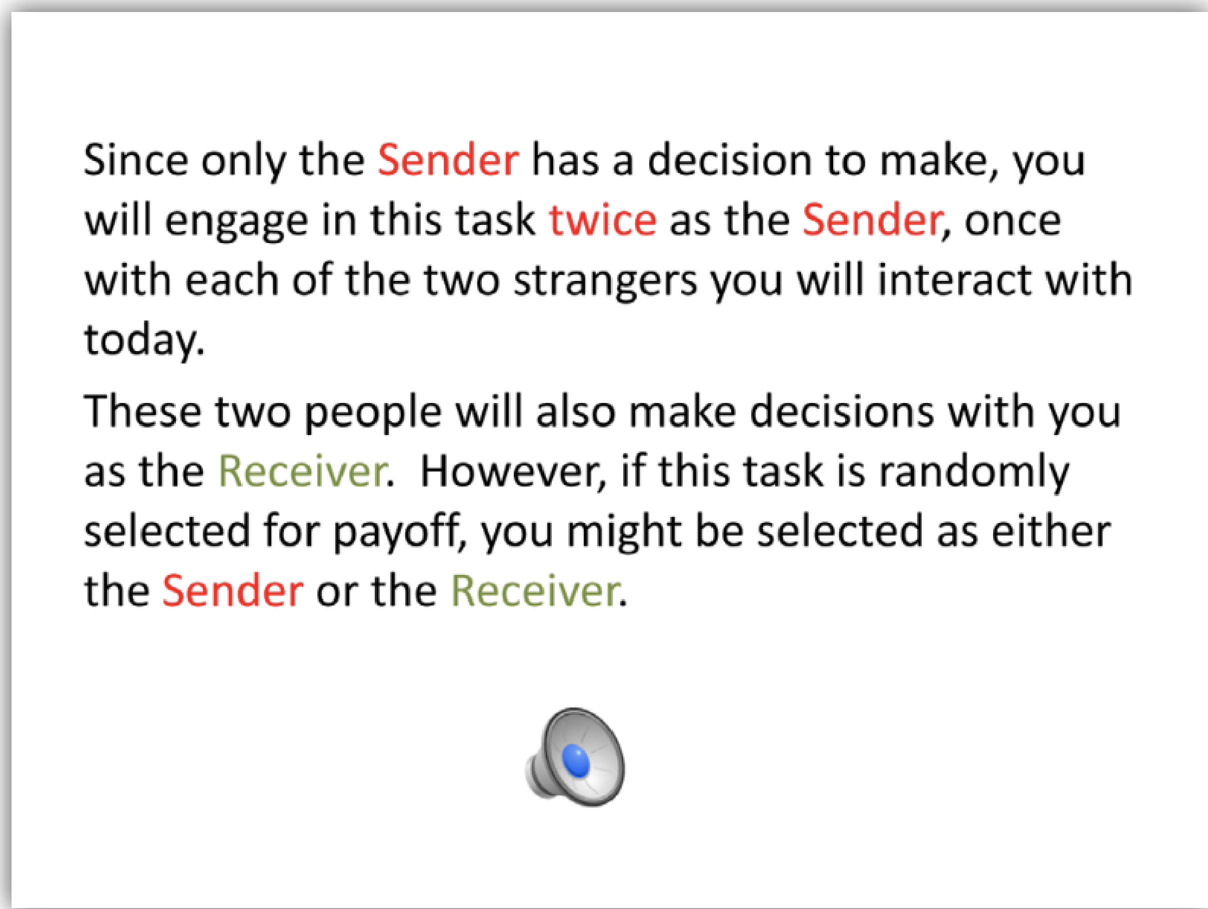
DICTATOR GAME INSTRUCTIONS

> *In the Decision task, one person – called the Sender – has the opportunity to send some of their money to another person – called the Receiver. This amount is tripled by the experimenter and credited to the Receiver.*



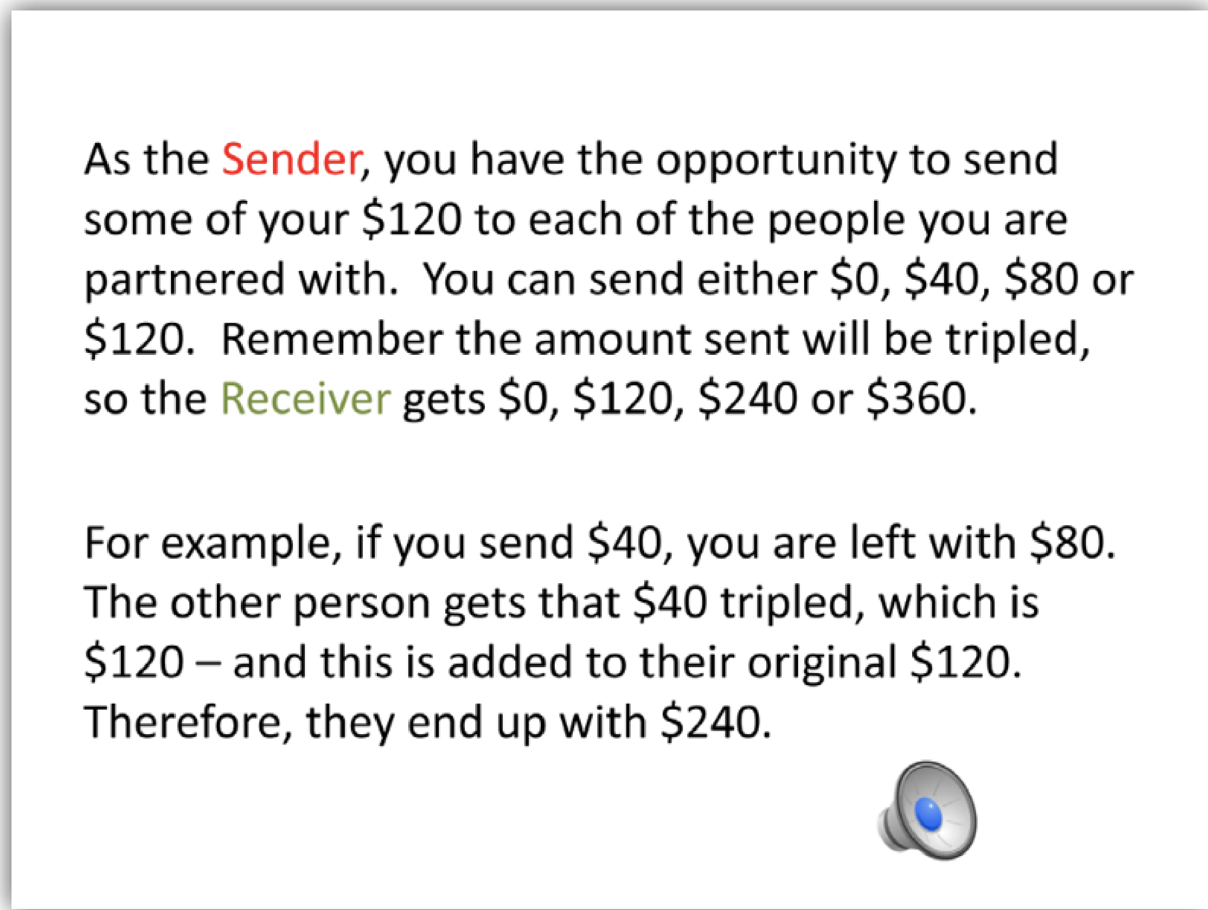
Since only the Sender has a decision to make, you will engage in this task twice as the Sender, once with each of the two strangers you will interact with today.

These two people will also make decisions with you as the Receiver. However, if this task is randomly selected for payoff, you might be selected as either the Sender or the Receiver.

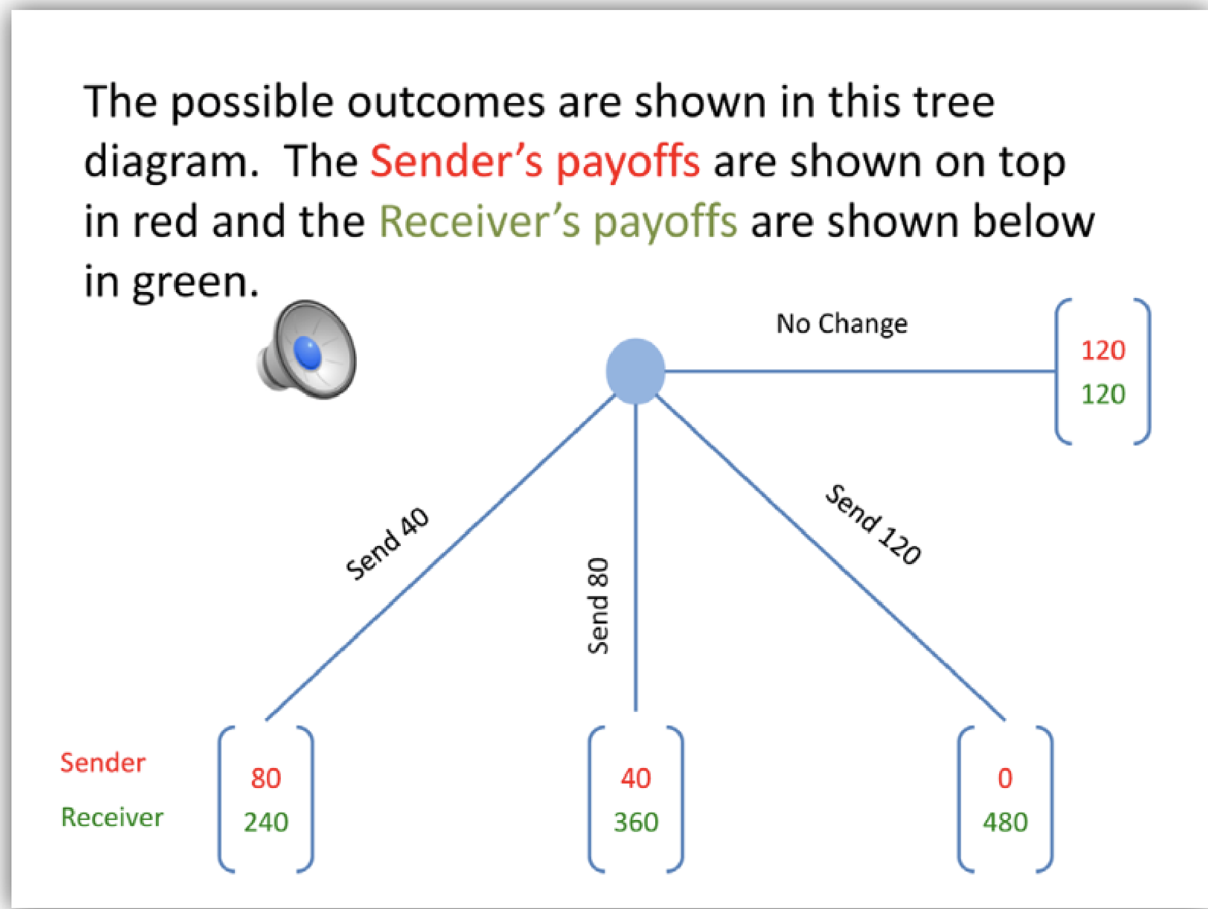


As the Sender, you have the opportunity to send some of your $120 to each of the people you are partnered with. You can send either $0, $40, $80 or $120. Remember the amount sent will be tripled, so the Receiver gets $0, $120, $240 or $360.

For example, if you send $40, you are left with $80. The other person gets that $40 tripled, which is $120 – and this is added to their original $120. Therefore, they end up with $240.

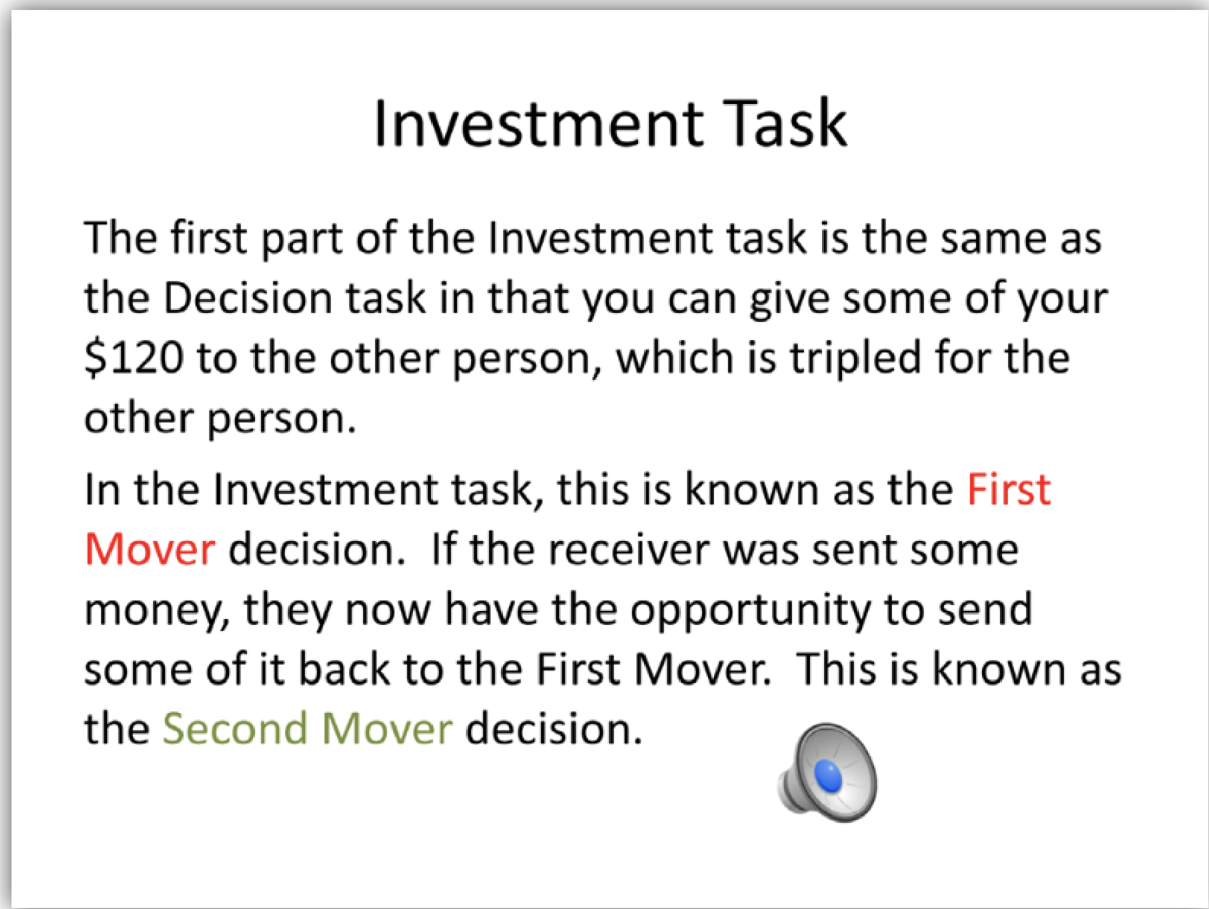


The possible outcomes are shown in this tree diagram. The Sender's payoffs are shown on top in red and the Receiver's payoffs are shown below in green.



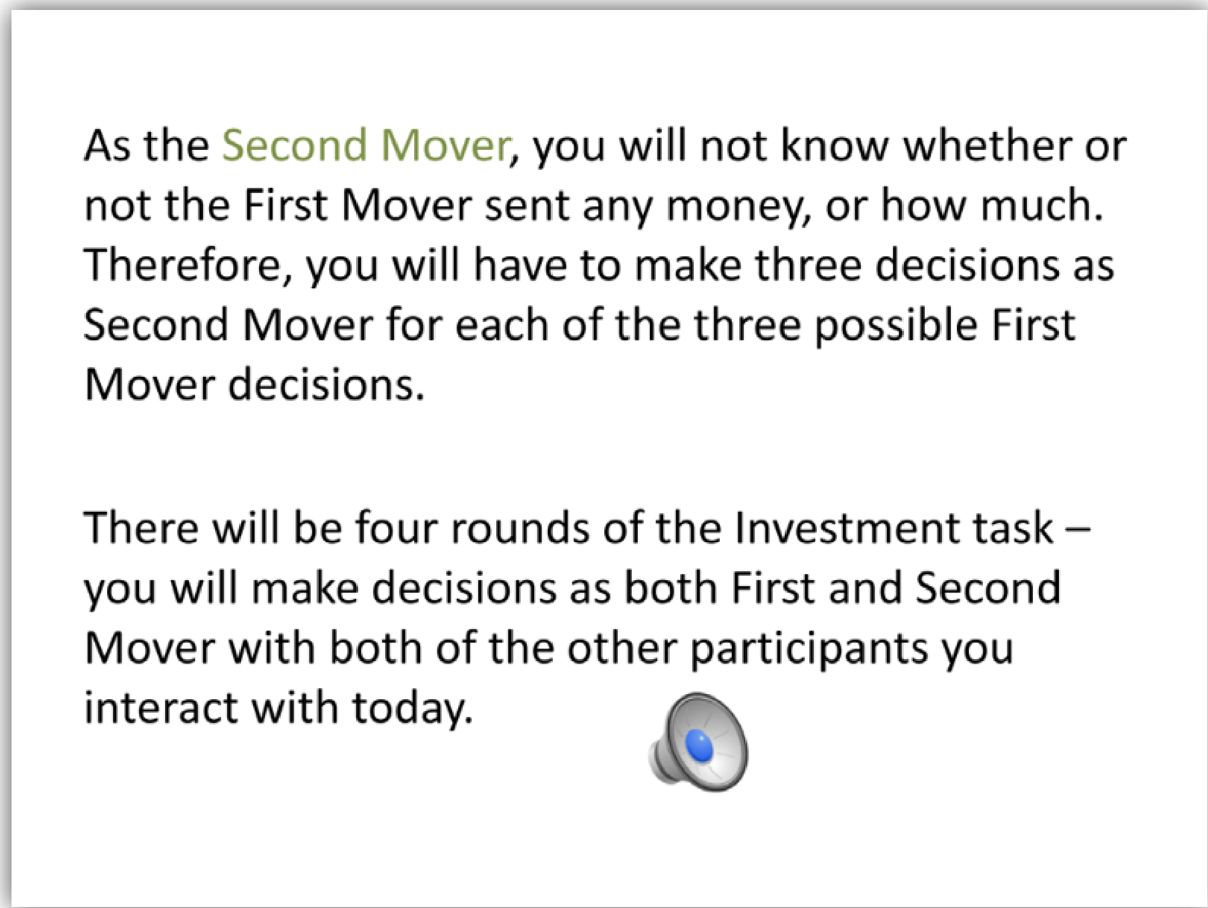
INVESTMENT GAME INSTRUCTIONS

> *The first part of the Investment task is the same as the Decision task in that you can give some of your $120 to the other person, which is tripled for the other person.*
>
> *In the Investment task, this is known as the First Mover decision. If the receiver was sent some money, they now have the opportunity to send some of it back to the First Mover. This is known as the Second Mover decision.*

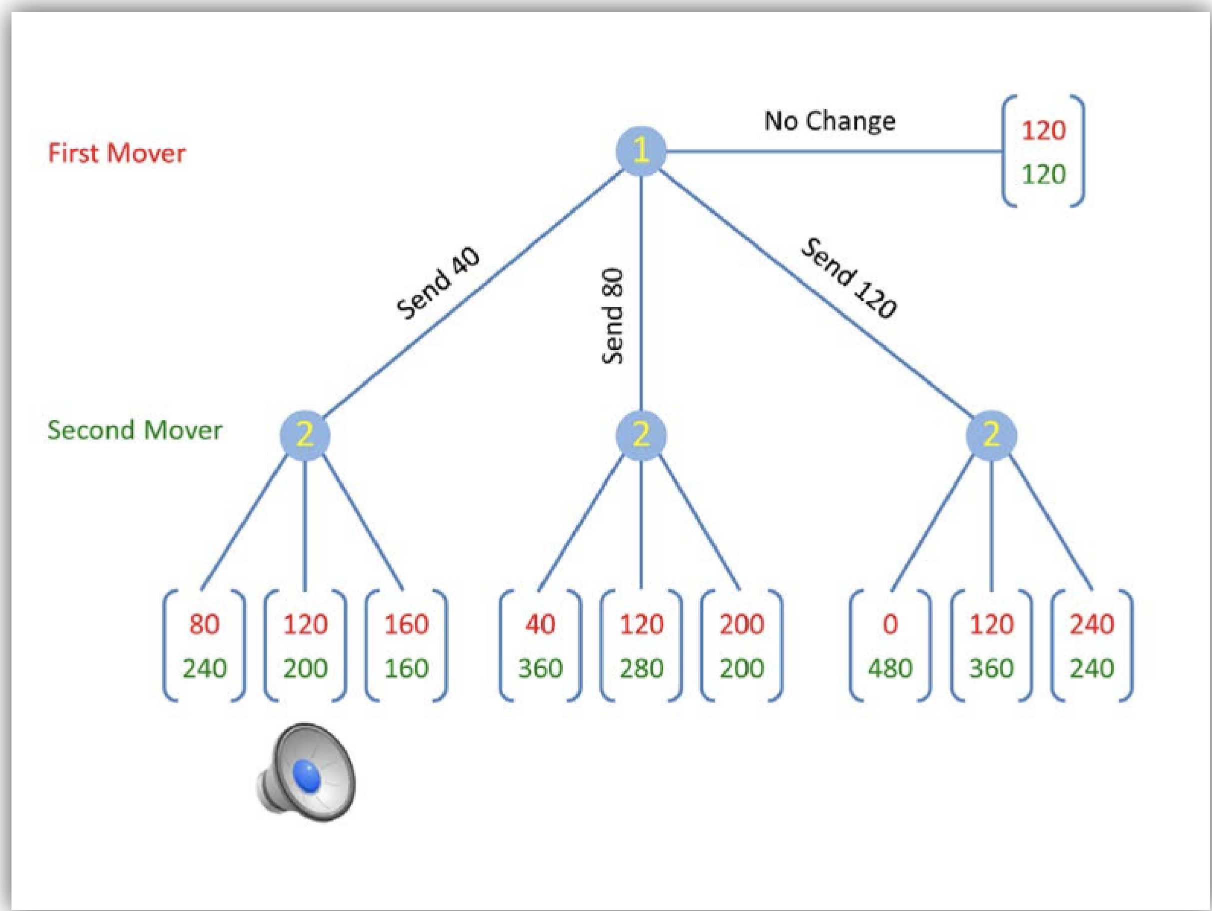


As the Second Mover, you will not know whether or not the First Mover sent any money, or how much. Therefore, you will have to make three decisions as Second Mover for each of the three possible First Mover decisions.

There will be four rounds of the Investment task – you will make decisions as both First and Second Mover with both of the other participants you interact with today.

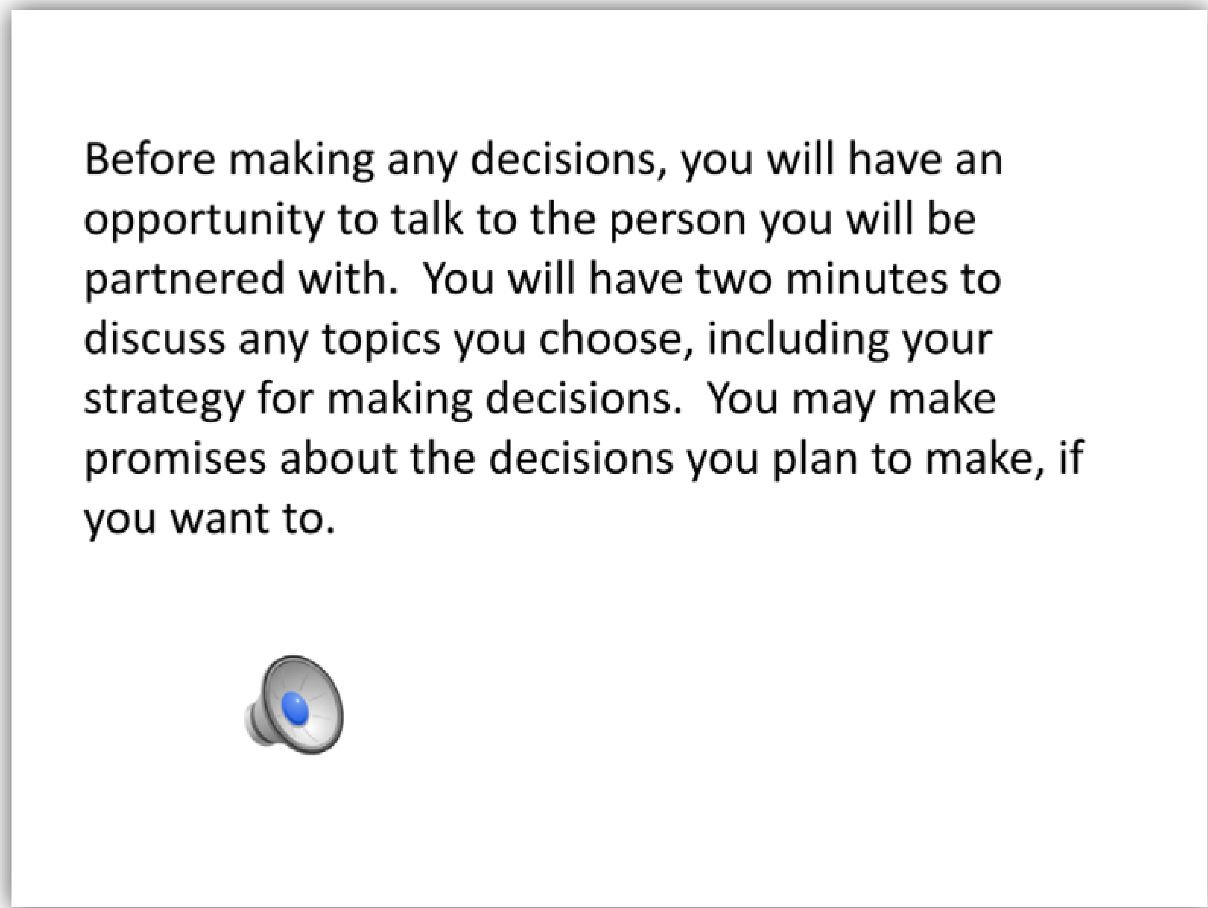


This tree diagram shows all the possibilities. At the end of each branch the final payoff is shown with the first mover payoff on top in red, and the second mover payoff below in green. The First Mover decision is indicated at the top by the blue circle with the number one inside. The First mover can send 0, 40, 80 or 120 dollars. This money is tripled and credited to the Second Mover.



If the Second Mover was not sent any money, they can do nothing. However, if they were sent some money, they have three options. They can either keep it all, return the amount that was sent to them or split the money equally. For example, looking at the central branch, if the First Mover sends 80 dollars, that gets tripled into 240 dollars and sent to the Second mover. At this point the First mover has 40 dollars and the second mover has 360 dollars. The second mover can return no money, so the payoff would remain 40 dollars to the first mover and 360 dollars to the second. Or the second mover can return the 80 dollars the first mover sent them, so the payoff would be 120 dollars for the first mover and 280 dollars for the second mover. Or the second mover can split the money equally so both the first and second mover get 200 dollars. In each of these possible outcomes, the total amount of money is the same, it is just divided up differently.

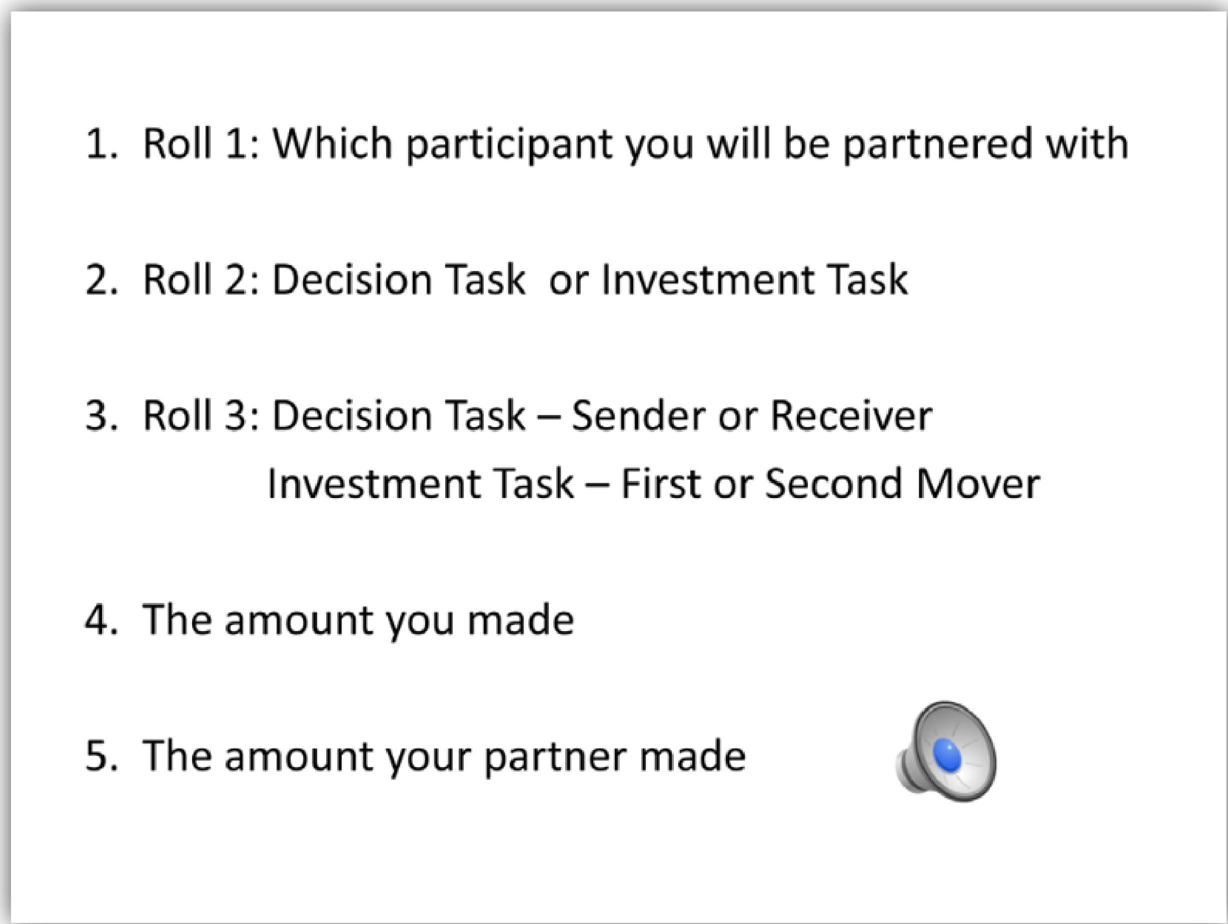
Before making any decisions, you will have an opportunity to talk to the person you will be partnered with. You will have two minutes to discuss any topics you choose, including your strategy for making decisions. You may make promises about the decisions you plan to make, if you want to.



REVEAL INSTRUCTIONS

Three die rolls are going to be used to determine which of the decisions will be paid out. The first die roll will determine which participant you will be paired with. The next will determine which task will be paid out, either the Decision Task or the Investment task. The final roll will determine who will assume each role in the task that was selected. If the decision task is selected, one person will be the one who sends their money – called the sender. The other one will be the person who receives the money – called the receiver. If the Investment task was selected, one person will be the first mover, the other the second mover.

Following these die rolls, you will learn first how much you made in this session and second, how much the person you were partnered with made.



References

Abbott, J., Middlemiss, M., Bruce, V., Smailes, D., & Dudley, R. (2018). The effect of arousal and eye gaze direction on trust evaluations of stranger's faces: A potential pathway to paranoid thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, *60*, 29–36.

Aimone, J., Ball, S., & King-Casas, B. (2015). The Betrayal Aversion Elicitation Task: An Individual Level Betrayal Aversion Measure. *PloS ONE*, *10*(9), 2–12.

Aimone, J. A., & Houser, D. (2012). What you don't know won't hurt you: A laboratory analysis of betrayal aversion. *Experimental Economics*, *15*, 571–588.

Alos-Ferrer, C., & Farolfi, F. (2019). Trust Games and Beyond. *Frontiers in Neuroscience*, *13*, 1–14.

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the reciprocal trust less? *Economics Letters*, *99*, 454–457.

Appelhans, B. M., & Luecken, L. J. (2006). Heart Rate Variability as an Index of Regulated Emotional Responding. *Review of General Psychology*, *10*(3), 226–240.

Arrow, K. (1974). *The limits of organization*, electronic edition. New York, NY: Norton & Company Inc.

Baer, M. D., & Colquitt, J. A. (2018). Why Do People Trust? Moving towards a more comprehensive consideration of the antecedents of trust. In *The Routledge Companion to Trust* (pp. 163–182), electronic edition. London: Routledge.

Balliet, D., & Van Lange, P. A. (2013). Trust, conflict, and cooperation: a meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, *275*(5304), 1293–1295.

Behrens, F., Moulder, R. G., Boker, S. M., & Kret, M. E. (2020). Quantifying Physiological Synchrony through Windowed Cross-Correlation Analysis: Statistical and Theoretical Considerations. *bioRxiv*, 1–53.

Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M. et al. (2020). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*, *10*(1), 1–9.

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79.

Ben-Ner, A., & Putterman, L. (2009). Trust, communication and contracts: An experiment. *Journal of Economic Behavior & Organization*, *70*(1– 2), 6–121.

Ben-Ner, A., Putterman, L., & Ren, T. (2011). Lavish returns on cheap talk: Two-way communication in trust games. *The Journal of Socio-Economics*, *40*, 1–13.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142.

Berndet, D. J., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. *Workshops on Knowledge Discovery in Databases*, 359–370.

Bicchieri, C. (2002). Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society*, *14*(2), 192–228.

Bicchieri, C., Lev-On, A. & Chavez, A. (2010). The medium or the message? Communication relevance and richness in trust games. *Synthese*, *176*, 125–147, Bigoni, M., Bortolotti, S., Casari, M., & Gambetta, D. (2013). It takes two to cheat: An experiment on derived trust. *European Economic Review, 64*, 129–146.

BIOPAC Knowledge Base. (2020). FIR Filter – Acqknowledge 3.9.2 (Mac). Retrieved Nov. 8, 2019, from https://www.biopac.com/knowledge-base/fir-filter-acqknowledge-3-9-2-mac/.

Blau, P.M. (1986). *Exchange and Power in Social Life* (2nd ed.), electronic edition. New York: Routledge.

Blume, L. E., & Easley, D. (2018). Rationality. In M. P. Ltd (Ed.), *The New Palgrave Dictionary of Economics* (Third ed.). London: Palgrave Macmillan.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review, 98*, 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the reference points for trustworthiness in Gulf and Western countries. *The Quarterly Journal of Economics, 125,* 811–828.

Bohnet, I., & Meier, S. (2005). *Deciding to distrust. Public policy discussion papers, Federal Reserve Bank of Boston, No. 05:4*.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, *55*, 467–484.

Boker, S. M., Xu, M., Rotondo, J. L., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychol Methods*, *7*(3), 338–355.

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27,* 269–299.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*(3), 163–182.

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.), electronic edition. New York, NY: Springer.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. *Technical Report, 2nd Version* https://www.biopac.com/wp-content/uploads/EDA-SCR-Analysis.pdf.

Brandts, J., & Charness, G. (2000). Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games. *Experimental Economics*, *2*(3), 227–238.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, *14*, 375–398.

Brosig-Koch, J., Ockenfels, A., & Weimann, J. (2003). The Effect of Communication Media on Cooperation. *German Economic Review*, *4*(2), 1–19.

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization*, *51*, 195–216.

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

Camerer, C. F. & Hare T. A. (2014). The Neural Basis of Strategic Choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain* (2nd ed.) (pp. 479–492). Elsevier Inc.

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Chaudhuri, A., & Gangadharan, L. (2007). An Experimental Analysis of Trust and Trustworthiness. *Southern Economic Journal*, *74*(4), 959–985.

Cheong, J. H. (2019). Four Ways to Quantify Synchrony Between Time Series Data. 2020, from https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9.

Chetty, R., Hofmeyr, A., Kincaid, H., & Monroe, B. (2021). The Trust Game Does Not (Only) Measure Trust: The Risk-Trust Confound Revisited. *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)*, Elsevier, vol. 90(C).

Coleman, J. S. (1990). *Foundations of Social Theory*, electronic edition. Cambridge: Harvard University Press.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281.

Cox, J. C., Kerschbamer, R., & Neururer, D. (2016). What is trustworthiness and what drives it? *Games and Economic Behavior*, *98*, 197–218.

Cox, J. C., Sadiraj, V., & Schmidt, U. (2011). Paradoxes and mechanisms for choice under risk. *Kiel Working Paper, No. 1712*.

Critchley, H.D. & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, *77*(4), 624–638.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, electronic edition. New York, NY: Avon Books.

Dana, J. D., Cain, D. M., & Dawes, R. M. (2005). What You Don't Know Won't Hurt Me: Costly (But Quiet) Exit in a Dictator Game. *Organizational Behavior and Human Decision Processes*, *100*(2), 560–572.

Danyluck, C., & Page-Gould, E. (2018). Intergroup dissimilarity predicts physiological synchrony and affiliation in intergroup interaction. *Journal of Experimental Social Psychology*, *74*, 111–120.

Dawson, M. E., Filion, D. L., & Schell, A. M. (1989). Is elicitation of the autonomic orienting response associated with allocation of processing resources? *Psychophysiology*, *26*(5), 560–572.

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity*, *2007*, 60803.

Dilmi, M. D., Barthès, L., Mallet, C., & Chazottes, A. (2020). Iterative multiscale dynamic time warping (IMs-DTW): a tool for rainfall time series comparison. *International Journal of Data Science and Analytics*, *10*(1), 65–79.

Eckel, C. & Grossman, P. (1996). Altruism in Anonymous Dictator Games, *Games and Economic Behavior*, *16* (2), 181–191.

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization*, *55,* 447–465.

Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology, 36,* 82–95.

Evans, A. M., & Krueger, J. I. (2009). The Psychology (and Economics) of Trust. *Social and Personality Psychology Compass*, *3*(6), 1003–1017.

Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology*, *47*, 171–177.

Fareri, D. S. (2019). Neurobehavioral Mechanisms Supporting Trust and Reciprocity. *Frontiers Human Neuroscience*, *13*, 1–7.

Fehr, E. (2009). On the Economics and Biology of Trust. *Journal of the European Economic Association*, *7*(2-3), 235–266.

Fehr, E., & Krajbich, I. (2014). Social Preferences and the Brain. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain* (2nd ed.) (pp. 193–218). Elsevier Inc.

Fetchenhauer, D., & Dunning, D. A. (2012). Betrayal aversion versus principled trustfulness—How to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior & Organization, 81*, 534–541.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 709–730.

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A Handbook of Process Tracing Methods for Decision Research*, electronic edition. New York, NY: Psychology Press.

Gambetta, D. (1988). Can We Trust Trust? In Gambetta, Diego (ed.) *Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213–237.

Gates, K. M., & Liu, S. (2016). Methods for Quantifying Patterns of Dynamic Interactions in Dyads. *Assessment*, *23*(4), 459–471.

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision Making and the Brain* (2nd ed.). Academic Press.

Golland, Y., Keissar, K., & Levit-Binnun, N. (2014). Studying the dynamics of autonomic activity during emotional experience. *Psychophysiology*, *51*(11), 1101–1111.

Guala, F., & Mittone, L. (2010). Paradigmatic experiments: The Dictator Game. *Journal of Socio-Economics*, *39*(5), 578–584.

Healey, J., Seger, J., & Picard, R. (1999). Quantifying driver stress: developing a system for collecting and processing bio-metric signals in natural situations. *Biomedical sciences instrumentation*, *35*, 193–198.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29.

Hill, C. A., & O'Hara, E. A. (2006). A Cognitive Theory of Trust. *Washington University Law Review*, *84*, 1717–1796.

Holm, H. J., & Nystedt, P. (2008). Trust in surveys and games – A methodological contribution on the influence of money and location. *Journal of Economic Psychology*, *29*, 522–542.

Hong, K., & Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, *28*, 197–213.

Houser, D., & McCabe, K. (2014). Experimental Economics and Experimental Game Theory. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain* (2nd ed.) (pp. 19–34). Elsevier Inc.

Isoni, A., & Sugden, R. (2019). Reciprocity and the Paradox of Trust in psychological game theory. *Journal of Economic Behavior & Organization*, *167*, 219–227.

Johnson, N., D., & Mislin, A., A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kamas, L., & Preston, A. (2012). Distributive and reciprocal fairness: What can we learn from the heterogeneity of social preferences? *Journal of Economic Psychology*, *33*(3), 538–553.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review, 95,* 1688–1699.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An Atlas of Interpersonal Situations,* electronic edition. New York, NY: Cambridge University Press.

Keogh, E., & Pazzani, M. (2002). Derivative Dynamic Time Warping. *Proceedings from First SIAM International Conference on Data Mining*.

Keogh, E., & Ratanamahatana, C. A. (2005). Exact Indexing of Dynamic Time Warping. *Knowledge and Information Systems*, *7,* 358–386.

Kéri, S., & Kiss, I. (2011). Oxytocin response in a trust game and habituation of arousal. *Physiology & Behavior*, *102*(2), 221–224.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*, 673–676.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*, 165–181.

Kret, M. E., Tomonaga, M., & Matsuzawa, T. (2014). Chimpanzees and humans mimic pupil-size of conspecifics. *PLoS ONE*, *9*(8), e104886.

Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in Psychology*, *6*, 1–10.

Kret, M. E., Fischer, A. H., & De Dreu, C. K. (2015). Pupil Mimicry Correlates with Trust in In-Group Partners with Dilating Pupils. *Psychological Science*, *26*(9), 1401–1410.

Kret, M. E., & De Dreu, C. K. W. (2019). The power of pupil size in establishing trust and reciprocity. *Journal of Experimental Psychology: General*, *148*(8), 1299–1311.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730–735.

Lahnakoski, J., & Chang, L. (2021). Intersubject Correlation. 2020, from https://naturalistic-data.org/content/Intersubject_Correlation.html.

Lee, T. M., Au, R. K., Liu, H. L., Ting, K. H., Huang, C. M., & Chan, C. C. (2009). Are errors differentiable from deceptive responses when feigning memory impairment? An fMRI study. *Brain and cognition*, *69*(2), 406–412.

Lev-On, A., Chavez, A., & Bicchieri, C. (2010). Group and dyadic communication in trust games. *Rationality and Society*, *22*(1), 37–54.

Levenson, R. W. (2014). The Autonomic Nervous System and Emotion. *Emotion Review*, *6*(2), 100–112.

Levenson, R. W., & Gottman, J. M. (1983). Marital Interaction: Physiological Linkage and Affective Exchange. *Journal of Personality and Social Psychology*, *45*(3), 587–597.

Lick, D. J., & Johnson, K. L. (2014). “You can’t tell just by looking!” Beliefs in the diagnosticity of visual cues explain response biases in social categorization. *Personality and Social Psychology Bulletin*, *40*(11), 1494–1506.

Liu, S., Zhou, Y., Palumbo, R., & Wang, J. L. (2016). Dynamical correlation: A new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychol Methods*, *21*(3), 291–308.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2010). *Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison*. CGS Working Paper, Vol. 2, No. 3, Cologne Graduate School in Management, Economics and Social Sciences.

Marieb, E. N. (2002). *Anatomy and Physiology*. San Francisco, CA: Pearson Education, Inc.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The Tie That Binds? Coherence Among Emotion, Experience, and Physiology. *Emotion*, *5*, 175–190.

Mayer, R. C., Davis, J. H., & Schoorman, D. F. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734.

McAssey, M. P., Helm, J., Hsieh, F., Sbarra, D. A., & Ferrer, E. (2013). Methodological Advances for Detecting Physiological Synchrony During Dyadic Interactions. *Methodology*, *9*(2), 41–53.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267–275.

McCabe, K. A., & Smith, V. L. (2002). Goodwill Accounting and the Process of Exchange. In G. Gigerenzer & S. Reinhard (Eds.), *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior*, *74*(1), 285–298.

Meeren, H. K., van Heijnsbergen, C. C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16518–16523.

Mitkidis, P., McGras, J. J., Roepstorff, A., & Wallot, S. (2015). Building Trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiology & Behavior*, *149*, 101–106.

Montepare, J.M., & Dobish, H. (2003). The Contribution of Emotion Perceptions and their Overgeneralizations to Trait Impressions. *Journal of Nonverbal Behavior*, *27*, 237–254.

Moulder, R. G., Boker, S. M., Ramseyer, F., & Tschacher, W. (2018). Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychological Methods*, *23*(4), 757–773.

Müller, Meinard. (2007). Dynamic Time Warping. In *Information Retrieval for Music and Motion* (pp. 69–84). Berlin, Heidelberg: Springer.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality Judgments Based on Physical Appearance. *Personality and Social Psychology Bulletin*, *35*(12), 1661–1671.

Norris, C. J., Larsen, J. T., & Cacioppo, J. T. (2007). Neuroticism is associated with larger and more prolonged electrodermal responses to emotionally evocative pictures. Psychophysiology, *44*(5), 823–826.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Palumbo, R. (2015). Interpersonal Physiology: Assessing Interpersonal Relationships Though Physiology. University of Rhode Island.

Palumbo, R., Marraccini, M., Weyandt, L., Wilder-Smith, O., A McGee, H., Liu, S. et al. (2017). Interpersonal Autonomic Physiology: A Systematic Review of the Literature. *Personality and Social Psychology Review*, *21*(2), 99–141.

Pflanzer, R., & McMullen, W. (2012). Electrodermal Activity & Polygraph Introduction. In *BIOPAC Student Lab Manual* (pp. I-1). Goleta, CA: BIOPAC.

Potts, S. R., McCuddy, W. T., Jayan, D., & Porcelli, A. J. (2019). To trust, or not to trust? Individual differences in physiological reactivity predict trust under acute stress. *Psychoneuroendocrinology*, *100*, 75–84.

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *PNAS*, *115*(31), E7265–E7274.

Reed, R. G., Randall, A. K., Post, J. H., & Butler, E. A. (2013). Partner influence and in-phase versus anti-phase physiological linkage in romantic couples. *International Journal of Psychophysiology*, *88*, 309–316.

Rotter, J. B. (1971). Generalized Expectancies for Interpersonal Trust. *American Psychologist*, *26*(5), 443–452.

Rotter, J. B. (1980). Interpersonal Trust, Trustworthiness, and Gullibility. *American Psychologist*, *35*(1), 1–7.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not So Different After All: A Cross-Discipline View of Trust. *The Academy of Management Review*, *23*(3), 393–404.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: behavioral and neural correlates. *Journal of Personality and Social Psychology*, *104*(3), 409–426.

Sally, D. (1995). Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92.

Sapienza, P., Toldra-Simats, A., & Zingales, L. (2013). Understanding trust. *The Economic Journal*, *123*, 1313–1332.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*(5), 617–640.

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization*, *62*, 272–292.

Schoenherr, D., Paulick, J., Worrack, S., Strauss, B. M., Rubel, J. A., Schwartz, B. et al. (2019). Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. *Behavior Research Methods*, *51*(1), 361–383.

Schoorman, D. F., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. *Academy of Management Review*, *32*(2), 344–354.

Schotter, A., Sopher, B. (2006). Trust and trustworthiness in games: An experimental study of intergenerational advice. *Experimental Economics*, *9*, 123–145.

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87–94.

Schulte-Mecklenbeck, M., Kühberger, A., Ranyard, R., eds. (2011). *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*, electronic edition. New York, NY: Routledge.

Searle, R. H., Nienaber, A.-M. I., & Sitkin, S. B. (2018). *The Routledge Companion to Trust*. London: Routledge.

Senin, P. (2008). Dynamic Time Warping Algorithm Review. *Technical Report CSDL-08-04*, 1–23.

Shor, M. (2005). *Non-Cooperative Game*, Dictionary of Game Theory Terms, Game Theory .net, <*http://www.gametheory.net/dictionary/Non-CooperativeGame.html*> Web accessed: September 2020.

Smith, V. L. (2003). Constructivist and Ecological Rationality in Economics. *American Economic Review*, *93*, 465–508.

Smith, V. L. (2010). What would Adam Smith think? *Journal of Economic Behavior & Organization*, *73*(1), 83–86.

Song, J. H., Colasante, T., & Malti, T. (2020). Taming anger and trusting others: Roles of skin conductance, anger regulation, and trust in children's aggression. *The British journal of developmental psychology*, *38*(1), 42–58.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21,* 349–354.

The Definition of Risk. (2019). Www.dictionary.com. https://www.dictionary.com/browse/risk

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality, 50,* 61–65.

Thielmann, I., & Hilbig, B. E. (2015). Trust: An Integrative Review from a Person–Situation Perspective. *Review of General Psychology*, *19*(3), 249–277.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience*, *3*(2), 119–127.

Todorov, A. (2008). Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*, 208–224.

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, *23*(3), 373–380.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, *27*(6), 813–833.

Todorov, A., & Oosterhof, N. N. (2011). Modeling Social Perception of Faces. *IEEE Signal Processing Magazine*, *28*(2), 117–122.

van der Werff, L., Legood, A., Buckley, F., Weibel, A., & de Cremer, D. (2019). Trust motivation: The self-regulatory processes underlying trust decisions. *Organizational Psychology Review*, *9*(2–3), 99–123.

van Witteloostuijn, A. (2003). A Game-Theoretic Framework of Trust. *International Studies of Management & Organization*, *33*(3), 53–71.

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803.

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel.: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, *28*(4), 260–271.

Vollan, B. (2011). The difference between kinship and friendship: (Field-) experimental evidence on trust and punishment. *The Journal of Socio-Economics*, *40*(1), 14–25.

Vyrastekova, J., & Garikipati, S. (2005). Beliefs and Trust: An Experiment. Discussion Paper 88, Center for Economic Research, Tilburg University.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal Acts of Irrational Trust: Motivated Attributions and the Trust Development Process. *Research in Organizational Behavior*, 26, 75–101.

Werner, J. (2008). Risk Aversion. In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London.

Williamson, O. E. (1993). Calculativeness, Trust, and Economic Organization. *Journal of Law & Economics*, *36*(1), 453–486.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Willoughby, A. R., Barraza, J. A., Javitz, H., Roach, B. J., Harrison, M. T., de Zambotti, M. et al. (2012). *Electrophysiological and Neuroendocrine Correlates of Trust in the Investment Game*, 1–22.

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y. et al. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes*, *120*(2), 260–271.

Zak, P. J. (2012). The Moral Molecule: The Source of Love and Prosperity. New York, NY: Dutton.

Zak, P. J. (2011). The physiology of moral sentiments. *Journal of Economic Behavior & Organization*, *77*, 53-65.

Zak, P. J., Barraza, J. A., Hu, X., Zahedzadeh, G., & Murray, J. (2022). Predicting Dishonesty When the Stakes Are High: Physiologic Responses During Face-to-Face Interactions Identifies Who Reneges on Promises to Cooperate. *Frontiers in Behavioral Neuroscience*, *15*, 1–10.

Zak, P. J., Borja, K., Matzner, W. T., & Kurzban, R. (2005). The neuroeconomics of distrust: sex differences in behavior and physiology. *American Economic Review*, 95(2), 360–363.

Zak, P. J., Kurzban, R., & Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Hormones and Behavior*, 48(5), 522–527.

Zak, P. J., Kurzban, R., & Matzner, W. T. (2004). The neurobiology of trust. *Annals of the New York Academy of Sciences*, 1032, 224–227.

Zaki, J., & Mitchell, J. P. (2013). Intuitive Prosociality. *Current Directions in Psychological Science*, *22*(6), 466–470.

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling. *Personality and Social Psychology Revie*, *7*(3), 194–215.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*(2), 175–189.

Zebrowitz, L. A., & Montepare, J.M. (2008). Social Psychological Face Perception: Why Appearance Matters, *Social and Personality Psychology Compass*, *2*(3), 1497–1517.