

# Why Do We Trust Strangers? Social Trust, Moral Reasoning and Identity

**Maria Banu**

**ANNALS of the University of Bucharest  
Philosophy Series**

Vol. LXVII, no. 2, 2018  
pp. 39 – 66.

**ANALELE  
UNIVERSITATII  
BUCURESTII**

# WHY DO WE TRUST STRANGERS? SOCIAL TRUST, MORAL REASONING, AND IDENTITY

MARIA BANU<sup>1</sup>

## *Abstract*

Behavioral economists have extensively argued against consequentialist theories of social trust. The most recent studies show that trust decisions are mainly expressive. Trust-taking behavior is non-consequential and linked to betrayal aversion, norms, and self-identity. Trustfulness is thus granted an affective and normative dimension. Yet these studies lack an integrative theoretical framework. In light of these results, this paper argues that reaching a more comprehensive understanding of the notion of social trust may draw on conceptual resources and empirical insights from moral psychology. Specifically, future studies may test and explore further the non-consequential aspects of social trust in connection with moral reasoning and moral identity.

*Keywords:* social trust, trustfulness, moral psychology, identity, moral reasoning.

## **Introduction**

Social trust, defined as the expectancy about the trustworthiness or benevolence of strangers (Rotter 1971), is one of the most puzzling topics in the literature on trust since there is no consensus on its definition and research methodology, nor on how to adequately measure it across individuals and societies. Yet, social trust is the “social glue” which reduces the transaction costs that its absence entails, enhances cooperation, and it is generally thought to be critical in solving social dilemmas. Trust brings about great individual and social benefits, as

---

<sup>1</sup> University of Bucharest, Faculty of Philosophy. Email: <mariabanu221@gmail.com>.

high trusters tend to cooperate more with strangers and behave more prosocially (Rönnerstrand 2013; Sønderskov 2011), whereas high trust societies tend to perform better in terms of government efficiency and economic growth (Knack & Keefer 1997; Zak & Knack 2001).

Yet, what motivates our decision to trust others, especially strangers? The dominant view is that people are rational actors who trust merely as a means of maximizing their utility. Trust is, thus, instrumental and consequential. Different accounts of social trust in this sense converge on two main conceptions. Trustworthiness-based social trust is a direct consequence of the perceived trustworthiness of others (*e.g.*, Hardin 1993; 2002; 2006). In contrast, trustfulness-based social trust distinguishes our own general propensity to trust others, especially strangers, from evaluations of their trustworthiness (*e.g.*, Yamagishi & Yamagishi 1994; Yamagishi 1998). Both concepts bear an important cognitive aspect since trust decisions eventually rely, in different ways and degrees, on information about how trustworthy we think people are.

Behavioral assessments, however, show that people seem to trust strangers excessively. Trust is risky and unwise (Berg *et al.* 1995). If trust indeed relied on information about the others' trustworthiness, then the wealth of empirical evidence of trust in strangers does not make much sense. Yet the consequentialist approach, based on risk and expectation, has been only partly confirmed (*e.g.*, Ashraf *et al.* 2006; Evans & Krueger 2011; Glaeser *et al.* 2010). Behavioral trust seems to be non-consequential and different from cognitive trust. It is expressive rather than instrumental; it is motivated by the act of trust itself, regardless of what people stand to gain. Consequently, behavioral trust has been linked to social preferences such as betrayal aversion (Bohnet & Zeckhauser 2004; Bohnet *et al.* 2008), norms (Dunning *et al.* 2014; Fetchenhauer *et al.* 2017), and preferences about self-identity (Yamagishi *et al.* 2015). In this sense, an important affective and normative dimension seems to characterize our propensity to trust others (Engelmann & Fehr 2017; Dunning *et al.* 2014; Fetchenhauer *et al.* 2017).

Given these competing views and wide-ranging empirical data on trust, what social trust is and why people generally trust others, especially strangers, is not at all settled. In fact, it would appear that people trust "both too little and too much" (Fetchenhauer *et al.* 2017). When asked

about how trustworthy they think others to be most people seem to be very distrustful, but behavioral assessments would indicate otherwise. Yet, these studies in behavioral economics lack an integrative theoretical framework that could further their explanatory power. As Yamagishi *et al.* (2015) argue, the next stage of study is to find a way to integrate the non-consequential aspects of social trust related to betrayal aversion, norms, and self-identity.

In this paper, I suggest that further studies focusing on the affective and normative dimension of social trust may benefit from drawing conceptual and empirical resources from moral psychology. This could constitute a first step in addressing the need to reach a conceptual integration of the non-consequential aspects of social trust. First, I will review the main research trends in both the consequentialist and non-consequentialist approaches to social trust. I will then point to their primary limitations in reaching a more in-depth understanding of what social trust is and why we generally trust others. Second, I will argue that recent developments in moral psychology concerning moral reasoning and moral identity may address the need to test the normative dimension of trust decisions and explore further the non-consequential aspects of social trust.

The paper is structured as follows. In **Section 1**, I focus on the *consequentialist* approach where social trust is an efficient strategy in securing cooperation and beneficial social interactions. Its main limitation, however, is that it often fails to predict real-life trust-taking behavior and decision-making. **Section 2** deals with *non-consequentialist* accounts of social trust which show that trustfulness is biologically rooted and that trust-taking is closely linked to betrayal aversion, norms, and preferences about self-identity. Nonetheless, I argue, these studies can only partly answer the question of why we trust strangers. **Section 3** tackles a so-called *normative* approach to social trust. I argue that moral psychology may offer viable insights for a conceptual integration of the non-consequential aspects of social trust. Turning our attention to the agents' moral reasoning and moral identity may provide further research directions in assessing the normative features of trust decisions. My suggestions in this sense focus on the relationship between trustfulness and trustworthiness in an individual and the situation-

dependency of trust decisions in connection with the amount of information available to the individuals.

### 1. Consequentialist approaches to social trust

The Trust Game (TG) parsimoniously illustrates the decision to trust others. This is a two-player, two-move game with a payoff matrix designed such that mutual cooperation (C, C) and mutual defection (D, D) are both Nash equilibria. According to the payoff matrix in *Fig. 1*, in the TG (C, C) is the unanimously preferred equilibrium.

		<b>B</b>	
		C	D
<b>A</b>	C	4, 4	0, 3
	D	3, 0	3, 3

*Fig. 1.*

The TG becomes a dilemma when rationality dictates that both players choose the action leading to the Pareto-inferior equilibrium (D, D). If I am certain that my partner will cooperate then I should join her and if I am certain that she will defect then I should defect as well. This is, in fact, why this is called the trust or assurance game. In conditions of uncertainty, however, according to standard decision theory, I must estimate the probability of my partner playing C or D. If I lack information to form any such estimates, then rationality suggests that I ought to treat all options as equally likely. I ought to defect if and only if, regardless whether my partner cooperates or defects, the total payoff I get by playing D in both cases (3 + 3) is higher than the joint payoff I get from playing C (4 + 0).

My decision to cooperate or defect depends very much on information about what the other player will do. For (C, C) to obtain, I need to know that my partner will cooperate. If I trust that she will play C then I will play C and we will both be better off. This is the basic idea that lies at the core of trustworthiness-based social trust. Trustworthiness-based

social trust primarily refers to Hardin's (1993; 2002; 2006) trust as encapsulated interest<sup>2</sup> and related definitions (e.g., Levi & Stocker 2000; Herreros 2004; Delhey & Newton 2005), including that in social capital theories (e.g., Coleman 1998; 1990; Putnam *et al.* 1993; Fukuyama 1995)<sup>3</sup>. They all ground social trust in the trustworthiness of others and place

---

- <sup>2</sup> According to Hardin, trust is a three-party relationship in which A trusts B with respect to  $x$  (Hardin 1993, 506). Specifically, stating that A trusts B is actually saying that A thinks that B is trustworthy in relationship with A in a given context (Hardin 2002). A trusts B because it is in B's interest to take into consideration A's interests in a certain matter, *i.e.*, B encapsulates A's interests in her own. B's interests may typically come from her valuation of the relationship with A and her interest in its continuation.
- <sup>3</sup> For the purposes of the paper, I have focused here on the distinction between trustworthiness-based and trustfulness-based social trust and referenced to the most relevant works where the roots of this distinction become clear. However, as already mentioned, there is yet no consensus on the definition of trust or research methodologies and many differences and similarities can be found across the literature. For illustration, in a recent study from 2014, Walterbusch *et al.* have collected 121 definitions of trust, spanning over 50 years, and used quantitative and qualitative analysis to study their similarities and differences. Their analysis focuses on (i) the most common used terms, (ii) the temporal distribution, and (iii) the clusters in all trust definitions. I have selected here some of their findings focusing on the key words used in the definitions discussed in this paper. First, they found that the word *expectation* was used in 47.9% of the cases, *vulnerability* in 23.1% of the cases, *belief* in 22.3% of the cases, *behavior* and *risk* each in 17.4% of the cases. Second, little research on trust was done before 1980, since they found only 10 definitions with approximately 25,000 cites. In the 1980s and especially in the 1990s, there was an increasing research trend on the topic with 21 and 57 definitions, respectively, and approximately 120,000 cites. Starting with 2000, they found 31 new definitions with almost 12,000 cites. However, as the study conveys, this is not an indication of a research decline in the topic. Finally, the authors found a high co-occurrence of the clusters *action/will* and *expect/subject*, as many definitions state that the trustor [*subject*] *expects* the trustee to perform an *action* in the future [*will*]. Moreover, the co-occurrence of *positive* and *negative* is explained due to the fact that trustors not only expect the trustee to act in a way that produces positive effects, but also to prevent negative effects. Other highly occurring clusters are *risk*, *situation*, and *involve*, as many definitions describe trust in a situation involving risk. At the same time, the clusters *willingness* and *risk* and/or *vulnerability* often occur together since many definitions imply that trust includes the willingness to take risks and make oneself vulnerable to the potentially malevolent or opportunistic actions of another.

very much emphasis on our experiences<sup>4</sup> with our social environment. For Hardin (2002), the others' trustworthiness gives rise to trust, and it is trustworthiness (not trust) that is the base of social cooperation.

Trust in this account is rather what Yamagishi (1998; Yamagishi & Yamagishi 1994) calls by assurance, which he distinguishes from trust. Assurance is based on encapsulated interest. It is secured by assessing the others' trustworthiness, *i.e.*, knowing their interests and motivations to behave trustworthily in relation to us. Yet trust, Yamagishi argues, is salient in circumstances of risk and social uncertainty, where I do not know what my partner will do. In such situations, the challenge is to explain why people often trust strangers since rational decision-making in conditions of uncertainty compels us to choose the risk-dominant equilibrium (**D, D**). The common answer, in the trustworthiness-based account, is that social trust extends to unknown parties. Social capital theories, for instance, argue that particularized trust – the trust we have in those close to us (*e.g.*, family members, friends, etc.) – is the basis for generalized trust, *i.e.*, trust as a rather abstract attitude, which includes social relations with people outside our immediate social environment, even utter strangers. Hence, how trustful one generally is essentially depends on how trustworthy those in their close social environment are.

However, it is not clear how we can validly infer the emergence of generalized trust from particularized trust<sup>5</sup> beyond the confines of a

---

<sup>4</sup> There are two broad theoretical conceptions of social trust: the experiential and the dispositional explanation (Dinesen & Bekkers 2017). The dividing line is what the causal prior of social trust is in each of these accounts (Nannestad 2008). The experiential explanation grounds social trust in the (perceived) trustworthiness of others. The fluctuating experience, broadly understood, with a particular social environment, shapes how trustful one is. In the dispositional account, on the other hand, social trust is an individual disposition that is only to a very limited extent subject to experiential influences after the formative years. The latter view endorses trustfulness, independently of the others' trustworthiness, as an autonomous, individual-level cause of social trust.

<sup>5</sup> Simply put, the process of building social trust in this view goes like this. Close social relations create trustworthiness within social structures. Closure allows the formation of reputation and collective sanctions that ensure trustworthiness (Coleman 1988). When trustworthiness is pervasive in a community, the aggregate subjective expectations which it creates, that its members will act

small group or community. Particularized trust *could* extend toward strangers, but it does not necessarily entail the individuals' general openness to engage in new, hypothetically more beneficial social interactions. And this is the main assumption at work in discussing social trust. Quite the contrary, based on particularized trust all we can say is that individuals are mainly prone to trust those in their close social environment, whom they know or acquire information about via other trusted (and trustworthy) individuals. In Yamagishi's terms, particularized trust, based on closure and trustworthiness, can only secure commitment relations. This mechanism of building social trust rather emphasizes the relation-consolidation role of trust. Yet, Yamagishi argues, the relation-extension role of trust is equally important, because of its potential to produce more individual and social benefits.

In his emancipation theory of trust Yamagishi (1998; Yamagishi & Yamagishi 1994) provides a different concept of social trust, based on the agents' trustfulness as a mainly stable, individual disposition (see note 4) to trust others, *especially* strangers. This account fares better in explaining the high trusters' tendency and openness to trust and cooperate with strangers. It cannot be subject to the same objections above because trustfulness does not causally rely on how trustworthy the others are. Being trustful is an adaptive<sup>6</sup> and efficient strategy, which

---

trustworthily, can be accounted for as social trust. High social trust, in turn, yields the economic and political performance of social organizations. Thus, particularized trust is *causally* linked to generalized trust and, then, to social benefits in terms of economic and political efficiency.

- <sup>6</sup> To Yamagishi (1998), social trust is an evolutionarily selected strategy. It is influenced by the social environment in the sense that it is an advantageous only in a high trust social environment. Conversely, it can prove disadvantageous in a low trust society. Yamagishi & Yamagishi (1994) explain the individual function of trustfulness as such. Commitment relations, where cooperation is assured and trust is consolidated, are the standard solution to the problem of social uncertainty, but only where opportunity costs are small. Commitment relations save the agent of incurring the transaction costs of forming new cooperation and trust relations. However, the actor who stays with committed relations despite better outside opportunities forgoes the potential extra benefit (*i.e.*, opportunity costs). Hence, commitment becomes a liability rather than an asset as opportunity



brings about the most benefits we enjoy from social interactions. That is because a high (vs. low) truster tends to overestimate the benignity of the partner's intentions beyond the level warranted by the cautious assessment of the available information. This yields the high trusters' openness to new social interactions that can potentially bring about more benefits. Empirical data, in fact, shows that high trusters tend to cooperate more with strangers and behave more prosocially (Balliet & Van Lange 2013; Bekkers 2012; Rönnerstrand 2013; Sønderskov 2011; Van Lange 2015).

Now, is this type of trust rational? If we look at the strategic choices the players have to make in the TG, it seems that playing C in circumstances of social uncertainty, where I do not know what my partner will do, is not rational. On a strictly rational perspective, trust is risky (Berg *et al.* 1995). Not only it is per se risky, but also people are generally risk-averse (Kahneman & Tversky 1979). Yamagishi (1998; 2001) anticipates this by arguing that high trusters possess some kind of social intelligence that enables them to "read" signs of (un)trustworthiness in others. Despite having incomplete information, they can assess whether or not their trust is (at least partly) warranted in specific situations. Most empirical studies have confirmed this link between trustfulness and social intelligence (Carl & Billari 2014; Hooghe *et al.* 2012; Oskarsson *et al.* 2012; Sturgis *et al.* 2010b). Hence, the decision to trust strangers remains within the boundaries of rationality. Even in circumstances of uncertainty, where trust is indeed relevant, high trusters are equipped to acquire information about their partners in the TG. The trustfulness-based account thus conforms to the consequentialist approach of social trust.

However, serious concerns exist regarding the adequacy of attitudinal measures designed to investigate individual and societal differences in social trust<sup>7</sup>. Attitudinal measures account for the individuals' belief-

---

costs increase. In such cases, high trusters leaving the committed relations are better off than low trusters who stay in committed relations.

<sup>7</sup> Early research assessed social trust using survey questions about the trustworthiness or fairness of others. *E.g.*, the General Social Survey (GSS), the World Value Survey (WVS), and the American National Election Studies (ANES) have relied on such questions to evaluate trust (Wilson & Eckel 2011). Glaeser *et al.* (2010), however,

based assessment about the trustworthiness of others (Knack & Keefer 1997; Kosugi & Yamagishi 1998; La Porta *et al.* 1997). In this sense, they bear an important cognitive component. People infer an unknown interaction partner's trustworthiness based on their expectation concerning the other's trustworthiness in order to judge the consequences of choosing to trust. Yet this relationship has been only partly confirmed in empirical studies. Recent behavioral studies argue against the methodological adequacy of attitudinal measures and the accuracy of their results. They show that the trusting behavior cannot be captured by beliefs of other people's trustworthiness (Fehr *et al.* 2005; Fehr 2009; Yamagishi *et al.* 2015; Fetchenhauer *et al.* 2017). Behavioral trust is different from attitudinal trust since, at a cognitive level, when asked about how trustworthy they generally perceive others to be, people do not seem to trust others at all, while behaviorally, they trust excessively (Fetchenhauer *et al.* 2017). Thus, attitudinal measures based on the consequentialist approach fail to predict actual trusting behavior (behavioral trust) in economic games<sup>8</sup>.

---

found little statistical correlation between an individual's answers to the GSS questions and her behavior in the TG. Furthermore, they show that attitudinal measures of trust are more strongly related to behavioral trustworthiness rather than behavioral trust. These findings have been confirmed by similar results in other studies (Ashraf *et al.* 2006; Ermisch *et al.* 2009; Lazzarini *et al.* 2004). Others have found that attitudinal trust predicts neither behavioral trust nor behavioral trustworthiness (Ben-Ner & Halldorson 2010).

- <sup>8</sup> Even though behavioral trust cannot be inferred based on attitudinal trust, the empirical correlations between the two are, in fact, more complex. In a study from 2009, Ahmed and Salas assessed the relationship between attitudinal trust, measured by survey items, and behavioral trust, measured by experimental games, in five culturally and economically varied countries (Chile, Columbia, India, Mexico, and Sweden) and found that behavioral trust and attitudinal trust differ significantly. First, at the level of behavioral trust, the level of trust and cooperation differs across countries: behavioral trust is highest in Sweden, followed by Latin America, and lowest in India. Second, attitudinal trust is significantly different among countries: Chile and Sweden have the highest level of self-reported trust, while Columbia has the lowest. However, the self-reported trustworthiness was not significantly different across countries. Finally, the authors note, the most intriguing finding was that no single survey item predicts actual trust measured in experimental games across all countries. Attitudinal

## 2. Non-consequentialist approaches to social trust

Behavioral assessments of social trust mainly use laboratory versions of the TG to show that trust extended to an anonymous counterpart is the main driver of the predominant choices that happen in the game (Berg *et al.* 1995; Brühlhart & Usunier 2012). The sequential TG (or the investment game) accurately captures the notion of trust as a behavior that makes an individual A (the trustor or investor) vulnerable to the actions of another person B (the trustee)<sup>9</sup>. Experimental results show that an important part of the subjects in B's role reciprocate A's trust and that an even greater part of the A players trusted B (Fehr *et al.* 2005). Moreover, when A was presented the option to reward or punish B, even at a (high) personal cost, then many A players rewarded B players that reciprocated their trust and punished those who didn't (De Quervain *et al.* 2004). But, even though punishing cheaters is experienced as satisfactory, most individuals actually prefer mutual cooperation (Fehr *et al.* 2005). The trustor's motive for taking this social risk is that mutual cooperation can increase not only her financial well-being (and that of the trustee), but also that it is itself rewarding (Engelmann & Fehr 2017).

These results are supported by a series pharmacological and neuroimaging studies investigating the role of oxytocin (OXT) in trust decisions (Mikolajczak *et al.* 2010; De Dreu *et al.* 2010; Van Ijzendoorn & Bakermanc-Kranenburg 2012; De Dreu & Giffin 2017). OXT is a neuropeptide

---

trust is significantly related to behavioral trust in some but not all countries. In Sweden, both attitudinal and behavioral trust were relatively high. However, while behavioral trust measured in Columbia was significantly higher, the self-reported trust measured in Chile was significantly higher compared to Columbia.

<sup>9</sup> The experiments are designed as follows. A (the trustor) and B (the trustee) each own, say, 100 m.u. First, A decides whether to keep this amount of money or to transfer (part of) it to B. If A decides to transfer some of the money to B, then B faces the same choice. The experimenter doubles each transfer, so that both players are better off if they transfer their money rather than if they decide to keep it. From a strictly rational point of view, B has a strong incentive to keep her money regardless of A's decision. If A anticipates this behavior, though, she has even fewer reasons to transfer it in the first place. Thus, while it is rational for both A and B to keep their money, a mutually beneficial exchange can take place only if A trusts B and if B behaves non-selfishly.

whose action on the brain can influence social behavior, as well as social motivation, sexual behavior, and social cognition (Engelmann & Fehr 2017). It inhibits the natural tendency to avoidance and defensive behavior and it facilitates attachment and parental care (Churchland 2011). In a most cited study, Kosfeld *et al.* (2005) tested the hypothesis that OXT enhances trust in the TG by administering it in two groups of participants – one that received synthetic OXT and the control group which received an inactive placebo. Their results show increased trust-taking in the OXT group compared to the control group. Most importantly, OXT affected only trust-taking behavior, but not the transfers in non-social risk games, nor those of the trustees to the trustors. Specifically, the A subjects who received OXT exhibited a significantly more predisposed behavior to trust, but OXT did not affect how trustworthy B players were. The percentage of B players in whom A trusted grew from 21% to 45% while the transfers B did to A remained constant in both groups.

Further studies show that the mechanism whereby OXT mediates trust-taking is via reducing betrayal or exploitation aversion, *i.e.*, A's fear of not being fooled by B (Bohnet & Zeckhauser 2004; Bohnet *et al.* 2008; Fehr *et al.* 2005). Betrayal aversion is a form of social anxiety that inhibits the propensity to trust compared to other non-social risk games where financial loss is not associated with intentional betrayal by another person (Engelman & Fehr 2017). That is, OXT does not affect the subjects' beliefs on how trustworthy the others are in the TG, but only their willingness to incur the risk of being betrayed. This means that, beyond the cognitive component of trust decisions discussed in the previous section, trust-taking behavior also bears an important affective dimension. Taken together, these results show that emotions are important mediators of trust decisions and endorse the dispositional, trustfulness-based concept of social trust. Trustfulness is a biologically rooted disposition<sup>10</sup> that translates into a trusting behavior via

---

<sup>10</sup> It is difficult to pinpoint what kind of biological disposition trustfulness really is since research on the matter is still in its initial stage. Trust could be innate (*i.e.*, shaped by genetic differences), acquired (formed by environmental conditioning), or shaped by an interplay between genes and environment. Studies using

mechanisms run by emotions, which are closely linked to social preferences, *e.g.*, betrayal aversion (Fehr 2009).

More recent studies confirm these results and argue against the assumption that trust as anticipated trustworthiness implies a strong connection between behavior and cognition. People often display behavioral trust despite having insufficient cognitive trust to support it. Thus, additional factors, other than betrayal aversion, can importantly influence when and if people trust, *e.g.*, norms (Dunning *et al.* 2014; Fetchenhauer *et al.* 2017) and preferences about self-identity (Yamagishi *et al.* 2015). In contrast to the literature on trust discussed in the first section, trust in this account is non-consequential and different from attitudinal trust. People do not trust in order to gain something but rather they have other reasons which may be difficult to integrate into a rational concept of trust. Trust seems to be motivated simply by the act of trust itself and what it represents, regardless of the outcomes (Dunning *et al.* 2012). The agents may hold pessimistic beliefs about the others' trustworthiness, but still appear to follow an injunctive norm that impels them to trust other people (Fetchenhauer *et al.* 2017). The emotions associated with fulfilling a social duty or responsibility (*e.g.*, guilt, anxiety) play an important role in this decision (Dunning *et al.* 2014). Behavioral trust is, thus, at least in part an expressive act. Yamagishi *et al.*'s (2015) hypothesis is that behavioral trust may be a preference about self-identity, *i.e.*, what kind of person one wants to be. People seem to derive personal satisfaction from being trustful and acting in a trustful manner. This is distinct from the general pro-social preference defined

---

samples of twins to partition sources of trust into genetic and environmental factors provide convincing evidence on either side (Cesarini *et al.* 2008; Oskarsson *et al.* 2012; Sturgis *et al.* 2010a; Van Lange *et al.* 2014). It is, thus, difficult to draw any conclusion yet on whether it is a genetic predisposition (Krueger *et al.* 2012; Riedl & Javor 2012; Sturgis *et al.* 2009) or an acquired one (Skyrms 2008; Yamagishi 1998). Either way, trustfulness seems to be biologically rooted. This is enough reason to back the dispositional account of trustfulness as an essentially individual disposition and autonomous feature of social trust. Even if we accept that trustfulness is an acquired trait, it remains unique to the individual (Dinesen & Bekkers 2017) and, therefore, cannot be grounded in the perceived trustworthiness of others.

as a preference for the distribution of the outcomes of an action<sup>11</sup>. It is merely a preference for trusting *per se* and it is non-consequential. It implies that one wants to be a trustful person regardless of the possible negative consequences of trust (Yamagishi *et al.* 2015).

These studies succeed in isolating important variables, other than the preference for maximizing outcomes, that influence trust-taking behavior and reveal significant features of social trust. However, my concern is that the absence of a theoretical framework to integrate them conceptually hinders their explanatory potential. On the one hand, since other competing explanations for social trust remain, nonetheless, valid (see note 11), we need to assess further the circumstances in which preferences such as betrayal aversion, norms, and self-identity are relevant. On the other hand, even though it may not be directly within the scope of behavioral research to address questions like why people exhibit such preferences, this endeavor is particularly valuable in order to assess their conceptual implications for our understanding of social trust. Behavioral assessments only give us a glimpse into trust decision-making processes and the reasons people have for taking such a high social risk to trust someone they do not know. Yet, why individuals are actually motivated to follow an injunctive norm or what may motivate their preference for building a self-identity as trustful persons need further qualification.

---

<sup>11</sup> This is not to say, however, that individuals do not have other reasons as well for acting in a trustful and trustworthy manner in the TG. In a study from 2016, Espín *et al.* study the impact of different motives on the individuals' choices in a dual-role TG, such as altruism, spitefulness, egalitarianism, and efficiency, next to self-interest and strategic motives. They argue that there is considerable heterogeneity in motives in the TG. Even though social motives can differ significantly, an important percentage of individuals seem to be motivated by (strategic) self-interest whereas others are driven by efficiency considerations (understood as a preference for maximizing the total outcome). According to the authors, the subjects' trustworthiness, in this sense, can be an important indicator of the motives individuals have for acting trustfully. That is because, while those individuals acting out of (strategic) self-interest are not trustworthy as trustees, those acting out of efficiency concerns are.

Labeling departures from standard economic rationality as expressive acts only partly answers the question of why people trust strangers. Betrayal aversion links trust to emotions and behavior and it shows that trustfulness has a critical, inherent affective nature. But what it really tells us is why people are reluctant to trust rather than actually trust. This idea is at odds with the fact that people prefer mutual cooperation. In fact, it may only add up to the explanation that, in conditions of uncertainty, it is rational to distrust rather than trust. Thus, why people trust each other in circumstances of uncertainty and prefer mutual cooperation, despite their crippling betrayal aversion, is yet to be explained. Norm-driven concepts of trust, on the other hand, link the decision to trust to norms. Bicchieri *et al.* (2011) tested the hypothesis that people view trust as normative and found that they do not, in fact, behave as though trust is a norm, but rather as though trustworthiness is. Assuming that a social norm imposes punishment for non-compliance, we don't really punish one for being distrustful, but for being untrustworthy in relation to us. Yet Dunning *et al.* (2014) emphasize that the norm people follow is, in fact, more internal and private rather than social. It is a moral norm in nature; trusting follows a moral standard that is internal to the individual and requires no social pressure to evoke. People trust strangers in compliance with an injunctive norm even if it conflicts with their preferences.

This is, however, an incomplete answer to the main question of this paper. Further aspects need to be addressed. Why do people feel compelled to trust others, if there is no peer pressure or social sanction for distrust? Why do they feel like they should trust when they do not necessarily want to? What kind of moral norm are they following and what does it mean that its source is internal? These are all *prima facie* questions. Yet they are particularly important since other studies argue against the predominant role of injunctive norms for trust behavior and found evidence that the decision to trust in the TG is more strongly influenced by *want* rather than *should* even when controlling for the fairness aspect (Thielmann & Hilbig 2016). Now, why would people *want* to exhibit trust to strangers and make themselves vulnerable to the actions of another? Moreover, why do they derive satisfaction from being trustful and acting trustfully? Why is it important for them to

build an identity as trustful individuals? Answering these questions is crucial in order to reach a comprehensive understanding of the notion of social trust.

### 3. A normative approach to social trust

Research on social trust needs to explore and build further on the notion that trust decisions have a significant emotional and moral component. Identifying the possible variables that may influence trust-taking behavior is just the first step in reaching a thorough explanation of why people trust strangers. Recent developments in moral psychology may provide a more unitary conceptual framework to integrate these results and the possibility to actually test the normative aspects of social trust. Potential studies on the agents' trust-taking decisions and behavior in connection to their moral reasoning and moral identity may help us sort out the potential moral considerations that lie at the core of our notion of trust.

Before outlining some suggestions on how this may be the case, there is first one important aspect concerning the potential normative character of trust decisions that needs to be addressed. Accepting that trust decisions are indeed normative, how are people motivated to trust others simply by the thought that they should do it? Korsgaard (2010), for instance, in her critique to the current attempts at tracing the evolution of morality, argues that what we actually need to explain is exactly this capacity to be motivated to do something simply by the thought that I should do it. She contends that morality is not just about altruism and cooperation. It is "our unique human capacity to take responsibility for ourselves, to give shape and form to our own identities and characters, and to make laws for our own conduct" (Korsgaard 2010, 7). To get a grip on what morality is and how it came to be we need to explain what Korsgaard calls *normative self-government*. This is a specific human capacity to be motivated to do something by the thought that I *ought* to do it. It enables us to assess the potential grounds of our beliefs and actions, to ask whether they constitute good reasons for actions, and to regulate them accordingly. Acting from a sense of *obligation* is an expression of normative self-government.



Studies on the evolution of morality usually overlook explaining how this capacity evolved. Nevertheless, according to Korsgaard, morality lies in the normative self-government rather than in altruism or cooperation and I think that she makes an essential point here. As many theorists on the evolution of morality notice, a disputably large number of instances of biological altruism and cooperation can be found in the animal world (see, *e.g.*, Trivers 1971; Axelrod & Hamilton 1981). Yet, most attempts at explaining altruism and cooperation, both in animals and human beings, rest on a maximizing conception that in certain circumstances may fail to uphold altruistic behavior<sup>12</sup>. Some of our arguably altruistic decisions – that standard economic rationality deems as irrational – pose an important challenge to this account since they sometimes seem to be motivated by something other than the outcomes of an action<sup>13</sup>.

In particular, there are plenty of instances where trust decisions are oblivious to the possible gains from cooperation and social interactions. People often feel like they should trust others even when they hold low expectations concerning future returns. Besides the prospects of mutual cooperation rendering everybody better off and the satisfaction derived from it, the decision to trust seems to be also about doing the right thing. What is challenging about social trust in this account is not the plain fact that people follow a norm despite incurring the risk of becoming worse-off. Rather, it is why doing the right thing motivates trust-taking behavior

---

<sup>12</sup> One general objection to this model of weak altruism that Trivers (1971) and Axelrod (1984), for instance, propose is that this is not, in fact, biological altruism. Altruistic behaviors that evolve via mutual reciprocation directly benefit the altruistic individuals, thus, they do not reduce their fitness in the long-term (Sober 1988). Second, studies show that cooperation in iterated interactions tends to deteriorate in the final rounds (Gintis 2000).

<sup>13</sup> An alternative model to weak altruism is strong reciprocity (Gintis 2000; Gintis *et al.* 2003; Bowles & Gintis 2004) which refers to an actual human behavior, which probably has a genetic component. In this account, altruism lies not merely in cooperation, but mostly in the predisposition to punish free-riders at a high personal cost, even when the probability for future interactions and recovering these costs is small. This type of altruistic behavior is motivated by the negative emotions triggered by free-riding (Fehr & Gächter 2000; 2002).

and it often prevails over standard economic rationality and people's betrayal aversion. Norm-driven trust decisions seem to be influenced by emotions associated with fulfilling a social duty or responsibility (*e.g.*, guilt, anxiety). Yet, whatever their intensity, emotions *per se* cannot confer normativity to trust-taking actions. What may confer normativity to our trust decisions may be precisely this capacity for normative self-government which determines the standards of reason, which we then take to govern our own actions. It might explain how trust-taking behavior carries a normative character and what it means that the norm people follow is internal and private. It might also explain how people may overcome their instinctive betrayal aversion by exerting some kind of control over it, and why trust-taking behavior is connected to our own identity.

Yet, to gain a broader understanding of social trust and the assessed normative feature of trust decisions, the first task is to evaluate how our moral intuitions may guide such behaviors, if at all. In this sense, behavioral assessments may need to address the normative character of trust decisions by integrating measurements from moral psychology concerning the agents' moral reasoning and moral identity. The moral rationales that advise trust decisions and their alignment with one's own identity may offer insights into what motivates trust decisions. Moral reasoning is the process whereby we attempt to reach an answer to the question of what is the right thing to do and to formulate justifications for our moral beliefs. Also, morality is intimately linked to our own identities in the sense that it is not just about how we stand in relation to other people but, most importantly, it is an expression of how we stand in relation to ourselves (Korsgaard 2010).

### *3.1. Trustfulness and moral reasoning*

Moral reasoning directed toward deciding what to do involves forming judgments about what one ought morally to do. Korsgaard's account only attempts at finding out what it is that renders our actions normative. Yet the content of our moral judgments depends on the moral theories they amount to, hence projecting different models of moral reasoning. Dual-process theories of moral judgment suggest that

there are two main moral principles that guide our responses to moral dilemmas (Conway & Gawronski 2013; Greene 2009; 2014; Greene *et al.* 2001). The morality of deontological actions depends on their intrinsic nature (*e.g.*, harming others is wrong regardless of its consequences), while the moral status of a utilitarian action is determined by its consequences (*e.g.*, harming others is acceptable if it increases the well-being of a greater number of people). Behavioral and neurobiological evidence shows that intuitive and automatic judgments tend to be typically deontological, whereas consequentialist judgments are often the result of slow, deliberative cognitive processes (Greene *et al.* 2007).

Two recent studies converge on the idea that people who make characteristically deontological judgments are preferred as social partners, perceived as more moral and trustworthy, and are trusted more in economic games. Everett *et al.* (2016) argue that intuitive moral judgments, which typically follow deontological prescriptions, confer an adaptive function by increasing a person's likelihood of being chosen as a cooperation partner. Similarly, Sacco *et al.* (2017) found that deontological decision-making fosters perceptions of trustworthiness and likeability. Decisions consistent with deontological (vs. utilitarian) moral reasoning may signal trustworthiness. Consequently, their bearers are trusted more in economic games.

These results may be explored further in order to determine what the relationship between trustfulness and trustworthiness in an individual is and whether the same type of moral reasoning that motivates trustworthy behavior may also motivate trustful behavior. Trustful (vs. distrustful) individuals tend to be more trustworthy and are preferred as cooperation partners by both high and low and trusters (Yamagishi 1998; Yamagishi & Yamagishi 1994). Colquitt *et al.* (2007), for instance, suggest that a high truster may socially project her own trustworthiness onto others. In this sense, trustfulness may also signal trustworthiness. After all, regardless of our strategic decisions, we do want our partners to trust us. If they do from the very start, then we may likely perceive them as willing to cooperate further and, hence, as more trustworthy. The relationship between trustfulness and trustworthiness in an individual has been frequently observed, but insufficiently explained (Thielmann & Hilbig 2015). In this respect, it is

worth exploring whether deontological moral reasoning, which signals trustworthiness, may also signal trustfulness and whether individuals who tend to reason deontologically are not only more trustworthy, but also more trustful. The assessed normativity of trust decisions seems to have a special affinity to deontological moral reasoning, but this is not yet a settled relationship.

Secondly, further studies on social trust need to investigate whether different types of moral decision-making may guide trust-taking behavior depending on the nature of the situation, or if deontological judgments are more fundamentally linked to social trust. This evaluation may determine the moral consistency of trust-taking behavior. Recent studies show that the experimental design of moral dilemmas may determine the “default” type of moral reasoning engaged in responding, as opposed to what dual-process theories contend (Koop 2013). We tend to reason deontologically when moral dilemmas are more personal or when we would like other people in general to respond in a certain way. For instance, if I do not want to be lied to, then I may think that no one should tell lies regardless of the possible negative consequences. Similarly, if it is important for A to be trusted by others, then she may hold the belief that people should be trustful even if they incur the risk of their trust being betrayed and she may, thus, act accordingly. In contrast, utilitarian moral reasoning is the default type when moral dilemmas are more impersonal or the subject is directly involved. I may think that lying is permitted if it bears positive consequences for a greater number of people, or that it is acceptable for me to lie in order to avoid harm to others. In the same vein, in some situations, A may think that she is justified not to trust some stranger to prevent negative consequences for herself and others.

Additionally, the amount of information available to individuals in various situations may determine important shifts in their default mode of processing trust decisions. This may be an important indicator in assessing the consequential or non-consequential features of social trust. Deontological decision-making seems to prevail when the actor has limited information (Baron *et al.* 2015). That is, deontologically motivated decisions are the default response when individuals deal with new actors and situations where they have little information available. As noted,

social trust primarily refers to trust in strangers, *i.e.*, people we have no information about, and behavioral assessments usually measure such behavior in conditions of anonymity. Thus, it is likely that deontological moral reasoning is the default type in such situations and only these. We need to evaluate whether, once one acquires more information about their partners in the trust game, they tend to reason in a more consequentialist fashion. If such a conclusion were reached, then it would pose an important challenge to the assessed normative character of trust decisions. It would suggest, in fact, that trust decisions are not inherently normative or morally motivated. Rather, this seemingly normative or moral character may simply conceal the heuristics of trust decisions. In situations where we deal with new actors and limited amounts of information, default trust requires less cognitive effort and, as discussed in the first section, it is more efficient in terms of producing individual benefits.

### ***3.2. Trustfulness and moral identity***

Future research needs to integrate assessments of the correlation between trustfulness, identity, and the normativity of trust decisions as well. Trust decisions seem to be motivated by a preference for being trustful and acting in a trustful manner, regardless of possible negative consequences (Yamagishi *et al.* 2015). Given the postulated normative features of trust decisions, we need to assess if, how, and what kind of moral considerations may guide this preference for self-identity. What is it about trustfulness that makes it a desirable characteristic in building one's own identity? The concept of identity holds multiple meanings and it has been tackled from various fields and angles. In the context of morally motivated trust decisions, my focus here, however, is with moral identity.

The connection between self-identity and morality is deeply rooted in both normative ethics and moral psychology. Previous research primarily emphasized moral reasoning and moral emotions in guiding moral actions. However, it has been extensively argued that these alone cannot motivate moral behavior (Aquino & Reed 2002). Since morality is

central to shaping self-identity, moral identity may play an important part in moral motivation (Hardy & Carlo 2005; 2011; Hardy 2006; Lapsley 2015). Strohminger & Nichols (2014) show that individuals consider morality indispensable to selfhood; the moral self is essential to our identity more than personality traits, memory, or desires. Moral identity motivates moral emotions, *e.g.*, guilt (Stets & Carter 2006), concern for out-group members (Aquino & Reed 2002; Hardy *et al.* 2010), and it may even motivate and predict moral actions (Hardy 2006; Perugini & Leone 2009). Agents with a stronger sense of moral identity may be more likely to exhibit congruence with moral judgments (Hardy & Carlo 2011). Thus, we need to assess the likelihood that one's own sense of moral identity may also underlie their preference for being trustful and acting trustfully. In this sense, perhaps those with a stronger sense of moral identity are more likely to exhibit trust in others, especially strangers, and disregard the possible negative consequences of misplaced trust.

First, the potential correlation between moral identity and trustful behavior may offer a better understanding of the relationship between trustfulness and trustworthiness in an individual. Newman *et al.* (2014), for instance, found that individuals tend to believe that their "true self" is fundamentally good, that is, that deep inside they believe that there is something motivating them to behave virtuously. Consistently, observers are more likely to see a person's true self reflected in what they deem as morally good behaviors than those they assess as bad, and their own moral values influence what they judge to be another person's true self. If my decision to trust others is morally motivated in this sense, then I may deem others as more trustworthy and, accordingly, behave trustfully in relation to them. This hypothesis is consistent with the notion that high trusters tend to overstate the benevolence of their partners in the trust game beyond the prudent assessment of their actual trustworthiness (*e.g.*, Yamagishi & Yamagishi 1994).

Second, Newman *et al.* (2014) argue that this normative view of the true self is independent of the particular type of mental state (beliefs *vs.* feelings) that is seen as responsible for an agent's behavior. This may explain why, even though trustfulness is important to one's own identity, actual trust decisions may depend, in fact, on the specifics of

the situations the individuals find themselves in and the amount of information that is available to them. Thielmann & Hilbig (2015), for instance, argue that trust among strangers can be explained through an interaction between features of the trust situation and a trustor's personality characteristics. They argue that people might decide to trust for various reasons. On the one hand, various features of the trust situation (*e.g.*, trustee characteristics, temptation to betray, availability to potential sanctions) might signal the likelihood of encountering a trustworthy (*vs.* untrustworthy) interaction partner. On the other hand, different personality characteristics (*e.g.*, risk and loss aversion, trustworthiness, and betrayal sensitivity) might influence an individual's willingness to trust irrespective of (or in interaction with) the situations at hand.

Finally, Aquino *et al.* (2009) examined the joint influence of situational factors and the centrality of moral identity to moral intentions and behaviors. Their hypothesis is that if a situational factor emphasizes one's moral identity, then it strengthens the motivation to act morally. In contrast, if a situational factor significantly fails to relate to one's moral identity, then it weakens the motivation to act morally. As argued above, the specifics of certain situations may influence the "default" type of moral reasoning engaged in trust decision-making. Deontologically motivated decisions are more prevalent when individuals interact with new actors of whom they have very little information. In such circumstances, the individuals' moral identity may play a greater role in trust decisions together with deontological judgments. Yet, it is likely that this emphasis on moral identity in trust decision-making to decrease once the agents acquired more information about their partners. They may shift to a more consequentialist approach where moral identity plays a smaller role in appraising trust decisions. Surely, this may not always be the case. In fact, Aquino *et al.* (2009) expected the influence of situational factors to vary depending on the extent to which moral identity was central to a person's overall self-conception. If one has a stronger sense of her moral identity, then the threshold for this shift in moral motivation to trust or distrust may be higher despite the increasing amount of information about their partners' trustworthiness.

### **Concluding Remarks**

Moral psychology may provide viable insights for a conceptual integration of the non-consequential aspects of social trust related to betrayal sensitivity, norms, and self-identity. Specifically, further research may use such conceptual and empirical resources to test and assess the presumed normative character of trust decisions by integrating evaluations of the individuals' moral reasoning and identity. In this sense, my suggestions in this paper focused on two aspects: (i) the insufficiently explained relationship between trustfulness and trustworthiness in an individual, (ii) the situation-dependency of trust decisions in relation to the amount of information available to the individuals. Addressing their importance in influencing trust decisions may provide us with a clearer notion of the normative features of social trust.

On the one hand, further studies need to address, first, if deontological moral reasoning associated with trustworthiness is also correlated with trustfulness. Second, if deontological decision-making is more fundamentally linked to social trust or the default type of moral reasoning that underlies such decisions varies depending on the particular situations the individuals find themselves in and the amount of information available to them. On the other hand, we need to assess the likelihood that a stronger sense of moral identity is directly correlated to being trustful and acting trustfully. First, this potential correlation may offer a better understanding of the relationship between trustfulness and trustworthiness in an individual. Second, it is important in order to assess whether situational factors and the amount of information available to individuals may override or enhance the primacy of self-identity in trust decision-making. These aspects, I contend, are important indicators in assessing the non-consequential features of social trust.

### **Acknowledgements**

I thank Professors Emanuel Socaciu and Emilian Mihailov, and my colleague Toni Gibeau, for their expert advice and support in writing this paper.



## REFERENCES

- Ashraf, N. *et al.* (2006). "Decomposing trust and trustworthiness." *Experimental Economics* 9: 193-208.
- Ahmed, A.M., Salas, O. (2009). "The Relationship between Behavioral and Attitudinal Trust: A Cross-cultural Study." *Review of Social Economy* 67(4):457-82.
- Aquino, K., Reed, A. (2002). "The Self-Importance of Moral Identity." *Journal of Personality and Moral Psychology* 83(6): 1423-40.
- Aquino, K. *et al.* (2009). "Testing a Social-Cognitive Model of Moral Behavior: The Interactive Influence of Situations and Moral Identity Centrality." *Journal of Personality and Social Psychology* 97(1): 123-41.
- Axelrod, R., Hamilton, W.D. (1981). "The Evolution of Cooperation." *Science* 211(4489): 1390-6.
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Balliet, D., Van Lange, P.A.M. (2013). "Trust, Punishment, and Cooperation Across 18 Societies: A Meta-Analysis." *Perspectives on Psychological Science* 8(4): 363-79.
- Baron, J. *et al.* (2015). "Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)?" *Journal of Applied Research and Cognition* 4: 265-84.
- Bekkers, R. (2012). "Trust and Volunteering: Selection or Causation? Evidence From a 4 Year Panel Study." *Political Behavior* 34: 225-47.
- Ben-Ner, A., Halldorson, F. (2010). "Trusting and trustworthiness: What are they, how to measure them and what affects them." *Journal of Economic Psychology* 31(1): 64-79.
- Berg, J., *et al.* (1995). "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10: 122-42.
- Bicchieri, C. *et al.* (2011). "Trustworthiness is a social norm, but trusting is not." *Politics, Philosophy & Economics* 10(2): 170-87.
- Bohnet, I., Zeckhauser, R. (2004). "Trust, risk, and betrayal." *Journal of Economic Behavior and Organization* 55: 467-84.
- Bohnet, I. *et al.* (2008). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States." *American Economic Review* 98(1): 294-310.
- Bowles, S., Gintis, H. (2004). "The Evolution of Strong Reciprocity." *Theoretical Population Biology* 65: 17-28.
- Brühlhart, M., Usunier, J.-C. (2012). "Does the trust game measure trust?" *Economic Letters* 115: 20-3.
- Carl, N., Billari, F.C. (2014). "Generalized Trust and Intelligence in the United States." *PLOS One* 9(3): 1-10.
- Cesarini, D. *et al.* (2008). "Heritability of cooperative behavior in the trust game." *Proceedings of the National Academy of Sciences* 105: 3721-6.
- Churchland, P.S. (2011). *Braintrust: What neuroscience tells us about morality*. Princeton: Princeton University Press.
- Coleman, J.S. (1988). "Social Capital in the Creation of Human Capital." *The American Journal of Sociology* 94(Supplement): S95-S120.
- Coleman, J.S. (1990). *Foundations of Social Theory*. Cambridge: Harvard University Press.

- Colquitt, J.A. *et al.* (2007). "Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships With Risk Taking and Job Performance." *Journal of Applied Psychology* 92(4): 909-27.
- Conway, P., Gawronski, B. (2013). "Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociations Approach." *Journal of Personality and Social Psychology* 104(2): 261-35.
- De Dreu, C.K. *et al.* (2010). "The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans." *Science* 328(5984): 1408-11.
- De Dreu, C.K., Giffin, M. (2017). "Neuroendocrine Pathways to In-Group Bounded Trust and Cooperation." In *Trust in Social Dilemmas*, edited by P. A. M. Van Lange, B. Rockenbach, T. Yamagishi, 57-76. New York: Oxford University Press.
- De Quervain, D.J.-F. *et al.* (2004). "The Neural Basis of Altruistic Punishment." *Science* 305: 1254-58.
- Delhey, J., Newton, K. (2005). "Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?." *European Sociological Review* 21(4): 311-27.
- Dinesen, P.T., Bekkers, R. (2017). "The Foundations of Individuals' Generalized Social Trust: A Review." In *Trust in Social Dilemmas*, edited by P.A.M. Van Lange, B. Rockenbach, T. Yamagishi, 77-100. New York: Oxford University Press.
- Dunning, D. *et al.* (2012). "Trust as a social and emotional act: Noneconomic considerations in trust behavior." *Journal of Economic Psychology* 33(3): 686-94.
- Dunning, D. *et al.* (2014). "Trust at Zero Acquaintance: More a Matter of Respect Than Expectation of Reward." *Journal of Personality and Social Psychology* 107(1): 122-41.
- Engelmann, J.B., Fehr, E. (2017). "The Neurobiology of Trust and Social Decision-Making: The Important Role of Emotions." In *Trust in Social Dilemmas*, edited by P.A.M. Van Lange, B. Rockenbach, T. Yamagishi, 33-56. New York: Oxford University Press.
- Ermisch, J. *et al.* (2009). "Measuring people's trust." *Journal of Royal Statistical Society: Series A (Statistics in Society)* 172(4): 749-69.
- Espín, A.M. *et al.* (2016). "Heterogeneous Motives in the Trust Game: A Tale of Two Roles." *Frontiers in Psychology* 7: 1-11.
- Evans, A.M., Krueger, J.I. (2011). "Elements of trust: Risk and perspective-taking." *Journal of Experimental Social Psychology* 47: 171-7.
- Everett, J. *et al.* (2016). "Inference of Trustworthiness From Intuitive Moral Judgments." *Journal of Experimental Psychology* 145(6): 772-87.
- Fehr, E. (2009). "On the Economics and Biology of Trust." *Journal of the European Economic Association* 7(2-3): 235-66.
- Fehr, E. *et al.* (2005). "Neuroeconomic Foundations of Trust and Social Preferences." *IZA Discussion Paper No. 1641*.
- Fehr, E., Gächter, S. (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Association* 90(4): 980-94.
- Fehr, E., Gächter, S. (2002). "Altruistic punishment in humans." *Nature* 415: 137-40.
- Fetchenhauer, D. *et al.* (2017). "The Mysteries of Trust: Trusting Too Little and Too Much at the Same Time." In *Trust in Social Dilemmas*, edited by P.A.M. Van Lange, B. Rockenbach, T. Yamagishi, 139-54. New York: Oxford University Press.

- Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York, London, Toronto, Sydney, Tokyo, Singapore: Simon & Schuster.
- Gintis, H. (2000). "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206: 169-79.
- Gintis, H. et al. (2003). "Explaining altruistic behavior in humans." *Evolution and Human Behavior* 24: 153-72.
- Glaeser, E.L. et al. (2000). "Measuring Trust." *The Quarterly Journal of Economics* 115(3): 811-46.
- Greene, J.D. et al. (2001). "An fMRI investigation of Emotional Engagement in Moral Judgment." *Science* 293: 2105-8.
- Greene, J.D. et al. (2007). "Cognitive load selectively interferes with utilitarian moral judgment." *Cognition* 107(3): 1144-54.
- Greene, J.D. (2009). "Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie." *Journal of Experimental Social Psychology* 45(3): 581-4.
- Greene, J.D. (2014). "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics." *Ethics* 124(4): 695-726.
- Hardin, R. (1993). "The Street-Level Epistemology of Trust." *Politics & Society* 21 (4): 505-29.
- Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.
- Hardin, R. (2006). *Trust*. Cambridge, Malden: Polity.
- Hardy, S.A. (2006). "Identity, Reasoning, and Emotion: An Empirical Comparison of Three Sources of Moral Motivation." *Motivation and Emotion* 30: 207-15.
- Hardy, S.A., Carlo, G. (2005). "Identity as a Source of Moral Motivation." *Human Development* 48: 232-56.
- Hardy, S.A., Carlo, G. (2011). "Moral Identity: What Is It, How Does It Develop, And Is It Linked to Moral Action?" *Child Development Perspectives* 5(3): 212-8.
- Hardy, S.A. et al. (2010). "Moral identity and psychological distance: The case of adolescent parental socialization." *Journal of Adolescence* 33: 111-23.
- Herreros, F. (2004). *The Problem of Forming Social Capital: Why Trust?*. New York: Palgrave Macmillan.
- Hooghe, M. et al. (2012). "The cognitive basis of trust. The relations between education, cognitive ability, and generalized and political trust." *Intelligence* 20: 604-13.
- Kahneman, D., Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263-91.
- Knack, S., Keefer, Ph. (1997). "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112(4): 1251-88.
- Koop, G.J. (2013). "An assessment of the temporal dynamics of moral decisions." *Judgment and Decision Making* 8(5): 527-39.
- Korsgaard, C.M. (2010). "Reflections on the evolution of morality." *Amherst Lecture in Philosophy*. The Department of Philosophy at Amherst College.
- Kosfeld, M. et al. (2005). "Oxytocin Increases Trust in Humans." *Nature* 435: 673-76.
- Kosugi, M., Yamagishi, T. (1998). "General trust and judgments of trustworthiness." *Japanese Journal of Psychology* 69(5): 349-57.

- Krueger, F. *et al.* (2012). "Oxytocin Receptor genetic variation promotes human trust behavior." *Frontiers in Human Neuroscience* 6: 1-9.
- La Porta, R. *et al.* (1997). "Trust in Large Organizations." *The American Economic Review* 87(2): 333-8.
- Lapsley, D. (2015). "Moral Identity and Developmental Theory." *Human Development* 58: 164-71.
- Lazzarini, S.G. *et al.* (2004). "Measuring trust: An experiment in Brazil." *Ibmec Working Paper No. 49*.
- Levi, M., Stocker, L. (2000). "Political Trust and Trustworthiness." *Annual Review of Political Science* 3: 475-507.
- Mikolajczak, M. *et al.* (2010). "Oxytocin makes people trusting, not gullible." *Psychological Science* 21(8): 1072-4.
- Nannestad, P. (2008). "What Have We Learned About Generalized Trust, If Anything?" *Annual Review of Political Science* 11: 413-36.
- Newman, G.E. *et al.* (2014). "Value Judgments and the True Self." *Personality and Social Psychology Bulletin* 40(2): 203-16.
- Oskarsson, S. *et al.* (2012). "The genetic origins of the relationship between psychological traits and social trust." *Twins Research and Human Genetics* 15(1): 21-3.
- Perugini, M., Leone, I. (2009). "Implicit self-concept and moral action." *Journal of Research in Personality* 43: 747-54.
- Putnam, R.D., Leonardi, R., Nonetti, R. Y. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, New Jersey: Princeton University Press.
- Riedl, R., Javor, A. (2012). "The Biology of Trust: Integrating Evidence From Genetics, Endocrinology, and Functional Brain Imaging." *Journal of Neuroscience, Psychology, and Economics* 5(2): 63-91.
- Rotter, J.B. (1971). "Generalized Expectancies for Interpersonal Trust." *American Psychologist* 26(5): 443-52.
- Rönnerstrand, B. (2013). "Social capital and immunisation against the 2009 A(H1N1) pandemic in Sweden." *Scandinavian Journal of Public Health* 41: 853-9.
- Sacco, D.F. *et al.* (2017). "The Adaptive Utility of Deontology: Deontological Moral Decision-Making Fosters Perceptions of Trust and Likeability." *Evolutionary Psychological Science* 3:125-32.
- Skyrms, B. (2008). "Trust, risk, and the social contract." *Synthese* 160: 21-5.
- Sober, E. (1988). "What Is Evolutionary Altruism?" *Canadian Journal of Philosophy* 14: 75-99.
- Sønderskov, K.M. (2011). "Explaining large-N cooperation: generalized social trust and the social exchange heuristic." *Rationality and Society* 23(1): 51-74.
- Stets, J.E., Carter, M.J. (2006). "The moral identity: A principle level identity." In *Purpose, meaning, and action: Control system theories in sociology*, edited by K.A. McClelland & T.J. Fararo, 293-316. New York: Palgrave MacMillan.
- Strohinger, N., Nichols, S. (2014). "The essential moral self." *Cognition* 131: 159-71.
- Sturgis, P. *et al.* (2010a). "A Genetic Basis for Social Trust?" *Political Behavior* 32: 205-30.
- Sturgis, P. *et al.* (2010b). "Does intelligence foster generalized trust? An empirical test using the UK birth cohort studies." *Intelligence* 38: 45-54.

- Thielmann, I., Hilbig, B.E. (2015). "Trust: An Integrative Review From a Person-Situation Perspective." *Review of General Psychology* 19(3): 249-77.
- Thielmann, I., Hilbig, B.E. (2016). "Should versus want: On the relative contribution of injunctive norms and preferences on trust decisions." *Journal of Behavioral Decision Making* 30: 446-52.
- Trivers, R.L. (1971). "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology* 46(1): 35-57.
- Van Lange, P.A.M. (2015). "Generalized Trust: Four Lessons From Genetics and Culture." *Current Directions in Psychological Science* 24(1): 71-6.
- Van Lange, P.A.M. *et al.* (2014). "Genetic influences are virtually absent for trust." *PloS One* 9(4): e93880.
- Van IJzendoorn, M.H. & Bakermanc-Kranenburg, M.J. (2012). "A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group." *Psychoneuroendocrinology* 37(3): 438-43.
- Yamagishi, T. (1998). *The Structure of Trust: An Evolutionary Game of Mind and Society*. Tokio: Tokio University Press.
- Yamagishi, T. (2001). "Trust as a Form of Social Intelligence." In *Trust in Society*, edited by K.S. Cook, 121-147. New York: Russell Sage Foundation.
- Yamagishi, T., Yamagishi, M. (1994). "Trust and Commitment in the United States and Japan," *Motivation and Emotion* 18(2): 129-66.
- Yamagishi, T. *et al.* (2015). "Two-component Model of General Trust: Predicting Behavioral Trust from Attitudinal Trust." *Social Cognition* 33(5): 436-58.
- Zak, P.J., Knack, S. (2001). "Trust and Growth." *The Economic Journal* 111(April): 295-321.
- Walterbusch, M. *et al.* (2014). "How Trust is Defined: A Qualitative and Quantitative Analysis of Scientific Literature." *Twentieth Americas Conference on Information Systems*: 1-10.
- Wilson, R.K., Eckel, C.C. (2011). "Trust and Social Exchange." In *Cambridge Handbook of Experimental Political Science*, edited by J.N. Druckman, D.P. Green, J.H. Kulinski, A. Lupia, 243-57. Cambridge, New York: Cambridge University Press.