Drivers of trust and trustworthiness

Jon Reiersen
School of Business, University College of Southeast Norway, Borre, Norway

Received 18 January 2018
Revised 23 April 2018
Accepted 26 April 2018

Abstract

Purpose – The purpose of this paper is to contribute to the understanding of why people act trustworthily in anonymous non-repeated meetings where trustworthiness benefits the trustor and runs against the trustee's material self-interest.

Design/methodology/approach – The paper uses a survey originally developed by Bicchieri *et al.* (2011). The survey makes it possible to explore whether trustworthiness has a normative element. Is there a norm of trustworthiness that inflicts punishment for disobedience?

Findings – The participants in the experiment strongly believe that most people will punish untrustworthy behavior, lending support to the idea that trustworthiness is norm driven. The data provide little evidence for a parallel norm of trust.

Originality/value – The theory of repeated games explains how trust can emerge among players in ongoing interactions. But why do people choose to trust others who they do not know in non-ongoing interactions? The results offer an explanation. When trustors are aware that trustworthiness is rooted in norms, they have reason to expect trustees to act trustworthily. Then, it makes sense to trust since trustors will benefit from their trusting.

Keywords Trust, Beliefs, Punishment, Trustworthiness, Norms

Paper type Research paper

1. Introduction

In recent years, trust has been accepted as a central concept for understanding economic behavior and organization. Many activities, such as trade and cooperation, become much easier if people trust each other. Mistrust leads to control and precautions that are both costly and time consuming.

The increased awareness of the varied benefits of trust have led to an explosion of research on the topic, and we now have a reasonably good understanding of what trust is and why it is important (e.g. Fukuyama, 1995; Rousseau *et al.*, 1998; Fehr, 2009; Algan and Cahuc, 2013; Sapienza *et al.*, 2013; Granovetter, 2017). Much less research has been directed at understanding the related, yet separate concept of trustworthiness. This is somewhat surprising given that trustworthiness is fundamental for trust. As Hardin (2002) notes: "[…] much of the literature on trust hardly mentions trustworthiness, even though implicitly much of it is primarily about trustworthiness, not about trust. […] Commonly, the best device for creating trust is to establish and support trustworthiness" (p. 30). Trust is not viable in the presence of widespread untrustworthiness. It makes no sense to trust others if others are untrustworthy. On the contrary, trusting can bring harm if others are untrustworthy. We trust others if we have reason to believe that they are trustworthy, i.e., if we think that they do not abuse the trust we show. If we want to understand trust in society we therefore need to better understand its foundation, namely, trustworthiness[1].

The aim of the present paper is to identify drivers of trust and trustworthiness in non-ongoing relationships. What motivates us to trust and be trustworthy in interactions with people we do not know and will not meet again? The paper is inspired by the framework developed by Bicchieri *et al.* (2011). By studying people's expectations regarding the

International Journal of Social Economics
Vol. 46 No. 1, 2019
pp. 2-17

0306-8293
DOI 10.1108/IJSE-01-2018-0025

JEL Classification — C72, C91, D63, Z10

The author of this paper has not made their research data set openly available. Any enquiries regarding the data set can be directed to the corresponding author.

punishment of untrustworthy behavior in a trust game, they find that there exists a norm of trustworthiness. Participants strongly believe that most people will punish untrustworthy trustees. In contrast, they find no evidence for a corresponding norm of trust. Participants do not believe that those who do not trust others will be punished for this. They conclude that norms are important for trustworthiness and therefore also for trust. When trustors are aware that trustworthiness is rooted in norms, trustors have reason to believe that trustees will act trustworthy. Therefore, it makes sense to trust since trust will be reciprocated.

The experiment in Bicchieri *et al.* (2011) was carried out in the USA with participants recruited from an American university. We are interested in investigating whether we can draw the same conclusions as Bicchieri *et al.* (2011) when we conduct the same experiment, but with participants from a different societal background. We know that norms are often highly contextual, in that they only cover a specific type of situation (Krupka and Weber, 2013). We also know that norms are influenced and shaped by the society in which people live (Ostrom, 2000). By redoing an experiment, it is possible to keep the context constant but vary the social and cultural environment of the participants[2].

This paper builds on the framework developed Bicchieri *et al.* (2011), using Norwegian participants. Are there reasons to believe that social and cultural diversity may produce different expectations regarding trusting and trustworthy behavior among American and Norwegian participants? Along many dimensions Norway and the USA are not very different. Both are wealthy and well established western democracies. Still, there are some important differences between the two societies. Norway has a more generous welfare state, higher social mobility, smaller income differences and a lower lever level of ethnic diversity (Alesina and Glaeser, 2004; Pontusson, 2005). Many argue that these elements of a societies' socioeconomic system are strongly connected with trust (e.g. Rothstein and Uslaner, 2005). If we look at the data on social trust from the World Value Survey, we see that the list of trusting citizens is topped by Norwegians, followed by Dutch and Swedes[3]. In total, 74 percent of Norwegians think that most people can be trusted, while the corresponding figure in the USA is 38 percent. This places the Americans as number 18 in the list of trusting citizens.

It is interesting to observe that despite substantial differences between Norway and the USA regarding social trust at the macro level, Norwegian and American respondents have just about the same expectations about which behaviors, in the context of the trust game, that they view as socially accepted and not. Our data show that norms are important in understanding how trustees behave. Norwegian participants strongly believe that others will punish those who act untrustworthily. Most respondents also state that they themselves would punish untrustworthy behavior. This lends support to the idea that trustworthiness is norm driven. Our results provide little evidence about the existence of a norm of trust. The obtained results are highly consistent with the results reported by Bicchieri *et al.* (2011).

The rest of the paper is organized as follows. Section 2 gives a brief overview of the importance and problem of trust and trustworthiness, focusing on the claim that trust is not sustainable in the absence of trustworthiness. Results from previous experimental studies of trust and trustworthiness are reported in Section 3 together with a discussion of the importance of norms as a driver for trust and trustworthiness. Section 4 introduces the experimental design followed by a presentation and discussion of the experimental results. Section 5 concludes the paper.

2. Trust is difficult without trustworthiness

This paper follows Gunnthorsdottir *et al.* (2002) who define trust as the "[…] voluntary transfer of a good or favor to someone else, with future reciprocation expected but not guaranteed" (p. 50). This is a representative definition and can be further specified in the following way: if A (the trustor) trusts B (the trustee), it is often the case that A transfers

something to B that will increase in value if B performs trustworthily. At the same time, A makes himself vulnerable by trusting. What A transfer to B is valuable to A. A therefore suffers a loss if B does not honor A's trust. The basic problem for a trustor is that he does not know if the trustee is trustworthy. If the trustor has this knowledge, there is no trust problem (Rousseau *et al.*, 1998; Ashraf *et al.*, 2006; Ben-Ner and Halldorsson, 2010; Reiersen, 2017).

The term trustworthiness is often used, but seldom with an explicit definition. One exception is Ben-Ner and Halldorsson (2010) who define trustworthiness as "[…] the willingness of a person B [the trustee] to act favorably towards a person A [the trustor], when A has placed an implicit or explicit demand or expectation for action on B" (p. 65). A problem that often adds to the trust-problem, and that is not covered in the above definition, is that B in many situations has an economic incentive not to honor A's trust. By exploiting A's trust, B can secure an extra economic gain. If A is aware of this, no trust relationship is formed, and both parties will lose out. Hence, the solution to the trust-problem rests with B. If B can get A to believe that B is trustworthy – that B will honor A's trust (despite B having an economic incentive to do the opposite), A will trust B. This underscores the point stressed in the Introduction. It is trustworthiness, that is, important and not trust by itself.

This last point is also the main insight arising from the studies conducted in Chaudhuri *et al.* (2003), Chaudhuri and Gangadharan (2007) and Dasgupta and Menon (2011). They show, with the help of experiments and scores on trust surveys, that people who are trustworthy are also generally more trusting. But the opposite is not necessarily true. There are many players that chose to trust in the role as trustors, but who chose not to reciprocate the trust of others in the role as trustees. For these players, both the act of trusting and not-reciprocating trust are motivated by self-interest. They trust in the hope that the trustee will reciprocate, which gives them a higher payoff – and they exploit the trust of others in increasing their own payoff. This demonstrates that if one observes a lot of trusting behavior, this does not automatically mean that there is a lot of trustworthiness out there. Or as Chaudhuri *et al.* (2003) note: "[T]rust and trustworthiness are fundamentally different constructs. […] What prior studies […] have interpreted as trust has two distinct components. One is 'pure trust' while the other is an element of calculated risk taking or a predilection for accepting a gamble" (p. 337). Pure trustors are both trusting and trustworthy, while calculating trustors are trusting but do not reciprocate others' trust.

3. The foundation of trust and trustworthiness

When a trustor is unaware of the trustworthiness of the trustee, he must form a belief about this, and act on the basis of this belief. What forces contribute in forming these beliefs and what can explain that trustees are viewed as trustworthy by a trustor? It is possible to distinguish between two general classes of explanations of why trustees may perform trustworthily, and hence why trustors should trust them (Torsvik, 2000). The first highlights traits of the trustees (as in the study by Chaudhuri *et al.*, 2003 commented on above), while the second explanation emphasizes the characteristics describing the interactive situation. It is the first explanation that is further explored in this paper, but the second explanation also deserves some comments.

Formation of trust through repeated interaction has received a lot of attention in economics. A key result in game theory is that a rational and fully self-interested trustor and trustee can solve the trust-problem described in Section 2. If the interaction is repeated enough times and the players are not too impatient, "trust" and "honor trust" can establish itself as a self-enforcing equilibrium. The main intuition of this result was noted by Thomas Schelling (1960) long ago: "[…] trust is often achieved simply by the continuity of

the relation between parties and the recognition by each that what he might gain by cheating in a given instance is outweighed by the value of the tradition of trust that makes possible a long sequence of future agreement" (pp. 134-135)[4].

When economists talk about trust, it is often trust through repeated interaction they are referring to (Torsvik, 2000; James, 2002)[5]. This is natural, since this type of trust is based on behavioral assumptions that are central to economic theory. Trust is rooted in maximization of long-term self-interest. The players are rational, they look to the future and they stick to the strategy that maximizes their payoff in the long-run[6].

Many trust relationships are undoubtedly grounded in repeated interaction. At the same time, it can be argued that a theory of trust that is unable to explain trust in non-ongoing meetings has its limitation. We know that this type of trusting exists. Both everyday observations and results from a series of experimental studies show that many are willing to trust strangers they know they will never meet again. Many also honor the trust they are offered by strangers. It is this type of trust and trustworthiness we are interested in studying in this paper.

Johnson and Mislin (2011) collected data from a total of 162 experiments with the trust game, involving more than 23,000 participants from different countries and cultures. In these experiments, participants act anonymously and they interact in one-shot meetings. There is thus no possibility to build trust through repeated interaction. All the experiments reported use the following variant of the trust game – also sometimes referred to as the investment game, introduced by Berg *et al.* (1995): participants are divided into pairs and are assigned the role of either the trustor (A) or the trustee (B). Both A and B are given an initial amount of money – of m, say (in some experiments, only A is given money). In the first stage of the game, A decides on the amount $x \in [0, m]$ he wants to transfer to B. A knows that if he transfers x, B will receive tx, with $t > 1$ (typically, t is 3). In the second stage, B observes what A has done and then decides on the amount $y \in [0, tx]$ that he wants to return to A. After this decision the game ends with material payoffs of $u = m - x + y$ for A and $v = m + tx - y$ for B. The idea is that the amount A transfers to B captures the degree of trust, while the amount B returns to A captures the degree of trustworthiness.

Johnson and Mislin (2011) find that on average, trustors transfer half of their money to the trustee while the trustees return a significant portion of their money to the trustor. On average, trustees return 37 percent of the amount they have available. It is very rare that some appear completely selfish (return nothing). They also find that higher amounts sent from A result in a significant higher percentage amount returned from B. Since the experiments are based on one-shot meetings, the observed trust and trustworthiness must be grounded in factors other than self-interest and the prospect of gaining something through repeated interaction[7].

Transfer of money back to the trustor in the trust game obviously breaks with the trustee's material self-interest. Why do many trustors still choose to send significant amounts of money to a trustee when they know that the trustee has a clear economic incentive to keep everything himself? One potential answer to this is that trust is norm driven (Bicchieri *et al.*, 2011). People believe that they are expected to trust, and thus they do so although a rational calculation tells them to do otherwise. Another potential answer is that trustworthiness is a norm (Uslaner, 2002; Fehr and Fischbacher, 2004; Bowles and Gintis, 2006; Bicchieri *et al.*, 2011; Gintis, 2017). Trustors transfer money because they know that the trustees are bound by norms of trustworthiness and that they therefore will back-transfer a significant amount of money.

The aim of the study presented below is to investigate whether trust is norm driven or if trust is better seen as an action motivated by the prospect of earning money through the trustee's trustworthy behavior.

4. Trust, trustworthiness and norms: an experiment

The existence and importance of social norms cannot be doubted, or as Fehr and Fischbacher (2004) note: "No human society exists without social norms […]. In fact, the ability to develop and enforce social norms is probably one of the distinguishing characteristics of the human species" (p. 64). Although "norms" is a fundamental concept in many of the social sciences, economists have traditionally given social norms little attention (although this is rapidly changing). One important reason for this is that norms are difficult to measure and quantify (Krupka and Weber, 2013). Another reason why economists find the concept of norms problematic is that there is no general agreement among scholars about how to define them. The literature about norms is highly complex and voluminous. However, a representative definition of norms is the following: norms are standards of behavior that are based on widely shared beliefs of how individuals ought to behave in a given situation and that are backed by social sanctions (Hechter and Opp, 2001; Elster, 2009; Horne 2009; Brennan *et al.*, 2015; Bicchieri, 2017).

Given this definition of norms, there are at least two challenges associated with identifying norms of trustworthiness using experiments where participants act anonymously. As discussed in more detail in Bicchieri *et al.* (2011), the first challenge is that pro-social norms (such as a norm of trustworthiness) usually dictate behavior that conflicts with our immediate material self-interest. To deviate from our self-interest, we must be confronted with strong normative expectations and the assumption that those around us are ready to punish us if we do not follow the norms. If these conditions are not present, the norm will lose much of its strength. This implies that a person who normally would follow a norm of trustworthiness can be tempted to deviate if his behavior is not observable, and thus not sanctionable. Observing behavior in anonymous experimental situations is therefore usually insufficient if one wishes to investigate whether a norm exists or not[8].

A solution to this problem is to build into the experiment an opportunity to punish others (see e.g. Fehr and Gächter, 2002; Fehr and Fischbacher, 2004; Horne, 2009). The problem here is that it can be difficult to distinguish between personal values and norms. Some may feel a deep personal commitment to a particular type of behavior, e.g., to behave trustworthy. These individuals may wish to punish untrustworthy behavior irrespective of norms. Others may have personal values that lead them to look at untrustworthy behavior more mildly, even though they are aware that a norm of trustworthiness exists. They will probably not be willing to punish others who behave untrustworthy.

One way out of these difficulties is to ask the participants directly how they think others will react and respond to different choices made by the trustor and trustee. The method of identifying social norms by asking respondents to consider the social appropriateness of different types of behavior is discussed in depth in a recent paper by Krupka and Weber (2013). They also develop a clever survey method for eliciting norms that uses coordination games. Bicchieri *et al.* (2011) use the same method, although their context is different. Their survey builds on the trust game as set out in Section 3. Their aim is to find out whether there is a social norm of trusting that imposes punishment for noncompliance? First, they ask respondents how they themselves would react toward a trustor (A) who does not send anything to the trustee (B). Second, they ask the respondents how they think others would respond to the same situation. Participants can respond by fining A (called a "payoff cut" in the instructions). Note that the money deducted from A does not fall into the hands of B. It is therefore not in B's self-interest to punish A.

If trust is a norm, we should expect that, on average, participants believe that the majority (more than 50 percent) will punish a completely untrusting trustor. However, Bicchieri *et al.* (2011) find that, on average, participants believe that 60 percent will in fact not punish lack of trust. In the case of a lack of trustworthiness, the participants respond

very differently. Participants are asked how they think others will react toward a trustee who returns nothing of the tripled amount received from A. Participants on average believe that only 24 percent will not punish this type of behavior. Stated differently, people think that a clear majority will punish an untrustworthy trustee and that few will punish an untrusting trustor.

To investigate whether the normative status of trust and trustworthiness depends on the relationship between the trustor and the trustee, Bicchieri *et al.* (2011) also conducted a "friend" treatment of the experiment. The only difference is that in the friend treatment the participants were told that the trustor and the trustee are friends. However, they do not find any significant differences between the stranger and the friend treatment. People do not expect that others would punish an untrusting trustor, even when the trustee is the trustor's friend. Based on these results, Bicchieri *et al.* (2011) conclude: "[P]eople do not believe that to trust is a norm. Our participants expected that most people would not punish untrusting investors, regardless of whether the potential trustee was a friend or stranger. On the other hand, our participants behaved as though behaving in a trustworthy manner is a social norm" (p. 181).

Do Norwegians have the same expectations about others' reactions toward trust and trustworthiness? To find out, we recruited Norwegians to participate in the same type of survey as developed by Bicchieri *et al.* (2011). While Bicchieri *et al.* (2011) have observations from 62 participants where all are students at the University of Pennsylvania, we have observations from 174 participants (66.6 percent female, average age $= 25.4$) where all are students at the University College of Southeast Norway. We visited six courses within different study programs at the university, and the experiment was conducted in the lecture hall in agreement with the lecturers who ran the courses we visited. The participants were given detailed information about the experiment together with an illustration of the trust game. They were then asked to anonymously fill out a questionnaire. The participants were invited to answer several questions related to the trust game. The survey took around 25 mins to complete, including the initial instruction period. The participants did not earn money by participating.

As in Bicchieri *et al.* (2011), we focus mainly on the two cases where either the trustor is completely untrusting, or the trustee is completely untrustworthy. In both cases we asked the participants three questions: whether they would impose a fine on either the trustor or the trustee, what their expectations were about how many other participants would chose not to fine the decision-maker and what their expectations were about the amount of punishment most participants would choose. Since Bicchieri *et al.* (2011) did not find any statistically significant differences between the friend and the stranger treatment, we only conducted the stranger treatment. This allowed us to leave out many questions and thus shorten the questionnaire considerably, compared to the original questionnaire developed by Bicchieri *et al.* (2011) (see the questionnaire in the Appendix).

If trust is a norm, we should expect that the participants on average expect that the majority will punish non-trusting behavior. We follow Bicchieri *et al.* (2011) and define the majority as more than 50 percent. Participants were asked to consider a situation where A (the trustor) and B (the trustee) receive NOK100[9] and A does not transfer any of this to B. In total, 67.2 percent of participants state that they do not want to punish A for this. Moreover, on average, 63.2 percent expect other participants to choose the same. This is significantly more than 50 percent (one-tail t-test, $p < 0.01$). Participants therefore believe that a clear majority will not punish untrusting behavior – which indicates that trusting is not perceived as a norm.

This result is very different from what the participants think will happen if B appears untrustworthy. Participants were asked to consider the following situation: "A and B receive NOK100. A transfers NOK50 to B, and B receives NOK150. B does not return anything to A.

A's payoff is NOK50 while B receives a payoff of NOK250." On average, participants believe that 34.9 percent of the other participants will not impose any punishment on a trustee who sends back nothing. This is significantly less than 50 percent (one-tail *t*-test, $p < 0.01$). In other words, participants believe that a clear majority will punish an untrustworthy trustee. The punishment option that the participants think would be chosen most times is to cut 50 percent from B's payoff. In total, 33.3 percent of participants think this would be the most popular choice.

These results provide support to the conclusion drawn by Bicchieri *et al.* (2011). People do not believe that trusting is a norm. They do not expect others to punish those who do not trust. However, people seem to believe that trustworthiness is a norm. Most think that the majority will punish those who do not reciprocate the trust they are shown.

Figure 1 illustrates the distribution of participants' expectations regarding the choice of others to punish or not. The dark-shaded bars provide information on answers given to the question: How many people/participants do you think decided to impose no punishment on A when the amount transferred to B is zero. The light-shaded bars provide information on answers given to the question: How many do you think decided to impose no punishment on B when the amount returned to A is zero?

By comparing the last two dark-shaded bars with the last two light-shaded bars on the right in Figure 1, we observe that significantly more participants expect that at least 50 percent of participants will choose no punishment in the untrusting trustor case than in the untrustworthy trustee case (65.0 percent vs 25.3, two-tail paired *t*-test, $p < 0{,}01$). Again, this gives support to the view that trustworthiness is a norm, but trusting is not[10].

Figure 2 plots the distribution of expectations that participants hold regarding the punishment level that most participants will choose. The dark columns show how much punishment participants think most participants will inflict on an untrusting trustor, while the bright columns show how much punishment participants think most participants will inflict on an untrustworthy trustee.

Figure 2 reveals that participants expect lack of trustworthiness to be punished much harder than lack of trust. In total, 12.6 percent expect that other participants will deduct all the earnings from an untrustworthy trustee, while 65.5 percent believe that others will deduct half or more of the trustee's earnings. If the trustee returns nothing, only 23.6 percent of participants expect zero punishment to be the most popular choice. In contrast, 74.7 percent of participants expect zero punishment to be the most popular choice when a trustor transfers nothing[11]. This evidence also supports the view that trustworthiness is considered as a norm, but trusting is not.

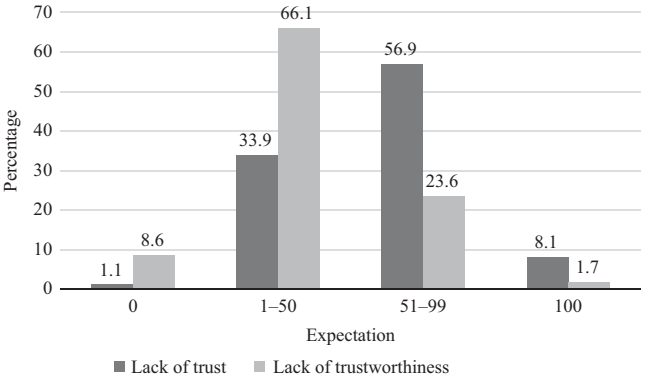

Figure 1.
Distribution of expectations regarding the percentage of people who imposed no punishment

The findings presented above seem to be consistent with recent experimental work showing that people are a lot more sensitive to acts of commission than acts of omission (Cox *et al.*, 2017). Cox *et al.* (2017) define a behavior that actively impose harm or kindness as an act of commission, while a failure to prevent harm or to act kindly is an act of omission. They show, with the help of an experiment, that acts of commission that overturn status quo generate significant stronger reciprocal responses than acts of omission. In the context of the trust game, the act of not transferring money as a trustor when there is an opportunity to do so (upholding status quo) can be viewed as an act of omission. As we have seen, people do not expect that this will be punished very hard. In contrast, the act of not back-transferring anything to the trustor (if the trustor transfer money) can be viewed as an act of commission. The trustee enrich himself at the expense of the trustor and thereby overturns status quo. Our respondents strongly believe that most people will punish this type of behavior.

So far, we have focused on expectations regarding punishment decisions when the trustee returns nothing (acts completely untrustworthy). What are the participants expectations regarding punishment when B is trustworthy to some degree? Participants were asked to consider the following three scenarios: A transfers NOK50 to B, which is tripled in value, and: B returns zero, B returns NOK50, B returns NOK100.

Figure 3 illustrates that 94.0 percent of participants expect no punishment to be the most popular choice when B returns NOK100 (remember that if B returns NOK100 both A and B get a payoff of NOK150). When B returns zero, only 23.6 percent of participants expect that most participants will choose no punishment. The average expected punishment imposed when B returns zero is NOK128.9, while it is NOK50.5 when B returns NOK50 (the difference between the two figures is statistically significant, two-tail paired t-test $p < 0.01$). The average expected punishment imposed when B returns NOK100 is close to zero.

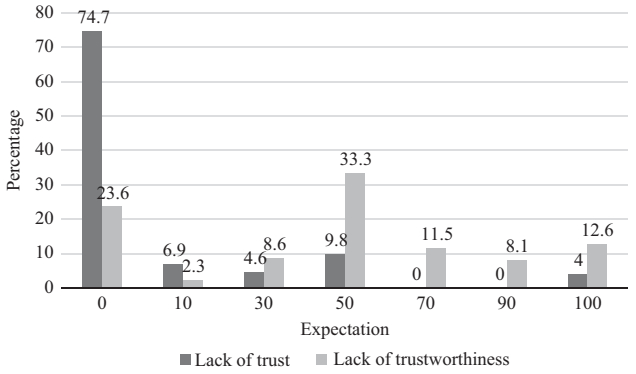

Figure 2.
Distribution of expectations regarding the most frequently chosen fine amount imposed (in percent of the decision-maker's payoff)

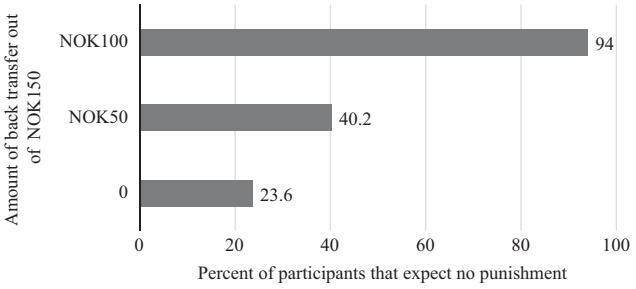

Figure 3.
Percent of participants that expect no punishment is the most frequent choice

Our data reveal that participants believe that most others will punish an untrustworthy trustee. What would our respondents do themselves? Are they prepared to follow a norm of trustworthiness? Participants were asked to consider the following situation: "You (B) receive NOK100, and the same does A. A sends you all his money, which is tripled in value. You thus have a total of NOK400. How much would you like to return to A?" The rational thing to do for a person with self-regarding preferences who does not care about norms is to keep everything. However, only 3.7 percent of participants state that they will follow this strategy (see Figure 4).

As much as 64.2 percent state that they will return NOK200. That is, they want to return an amount such that the total sum (NOK400) is equally distributed between A and B. Nearly 10 percent state they will return NOK150 to A, which secures B a payoff of NOK250. In total, 19.7 percent will return NOK100 which leaves A with his initial amount only.

These results show that a large majority wants to behave trustworthy and that trustworthiness implies that A should at least get what he can secure by not trusting B. The results also provide support for the observation made by Ben-Ner and Halldorsson (2010), who note that trustworthiness is grounded in social norms, but these norms "[…] do not entail a precise course of action in a variety of situations, providing B considerable room for discretion as to what constitutes trustworthy behavior" (p. 66).

Previous experiments with the trust game have documented that the amount sent by the trustor affects the proportion returned by the trustee. The share of money returned is higher the more money is sent (Fehr and Rockenbach, 2003; Johnson and Mislin, 2011). We observe the same pattern in our experiment (see Figure 5).

Participants were asked to decide how much they will return to A if A sends NOK20 and NOK100, respectively (see the last part of the questionnaire). In the first case, a large majority (66.7 percent) wants to return an amount that leaves B with a higher payoff than A. In the second case, 64.2 percent state that they will return an amount that will leave A and B

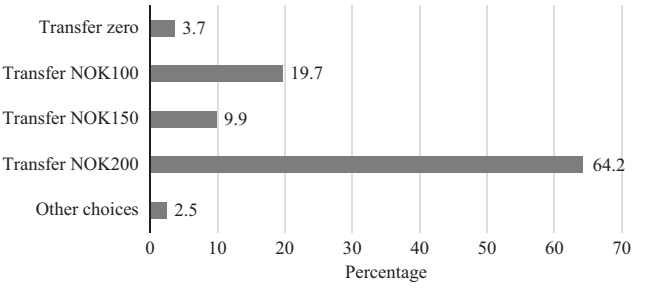

Figure 4.
The amount of back-transfer out of NOK400 (in percent)

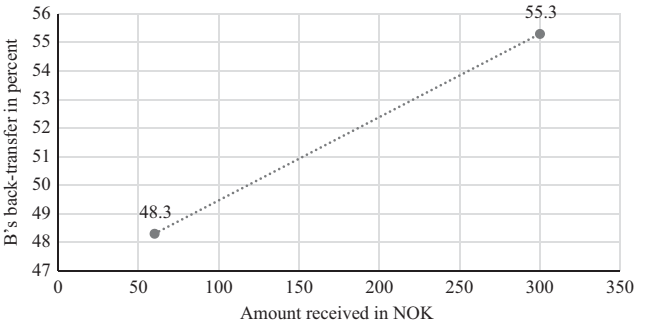

Figure 5.
B's back-transfer as a percentage of the amount received

with the same payoff. The average amount returned of the tripled amount received is 48.3 percent when A sends NOK20, while it is 55.3 percent when A sends NOK100 (the difference is statistically significant, one-sided t-test, $p < 0.01$). It seems that trustworthiness is always there, but that more trust generates more trustworthiness.

5. Conclusion

There is now a large literature stressing the importance of trust for economic interaction and outcomes. However, an understanding of the importance of trust and its consequences requires an understanding of the related yet different concept of trustworthiness. As noted at the beginning of the paper, it makes little sense to trust others if they are untrustworthy. Trusting can even bring harm if others are untrustworthy. Therefore, trust is only sustainable in the presence of widespread trustworthiness. The goal of this paper is to explore the concept of trustworthiness and its main driver.

This paper contributes to the understanding of why people act trustworthily in non-repeated anonymous meetings – a behavior that seems surprising, given that trustworthiness benefits the trustor and runs against the trustee's self-interest. Our results provide evidence for a norm of trustworthiness. Our participants strongly believe that most people will punish untrustworthy behavior, lending support to the idea that trustworthiness is norm driven. We found no evidence for the existence of a norm of trust. Our participants believe that most people would not punish an untrusting trustor. These results are highly consistent with previous results provided by Bicchieri *et al.* (2011).

Why do people choose to trust others, who they do not know, in a non-ongoing situation? Our results, together with the results provided by Bicchieri *et al.* (2011), offer an explanation. When trustors are aware that trustworthiness is rooted in norms, they have reason to expect trustees to act trustworthily. Thus, it makes sense to trust since trustors will benefit from their trusting.

Notes

1. See also Simpson and Eriksson (2009), Chaudhuri *et al.* (2003), Chaudhuri and Gangadharan (2007), Ben-Ner and Halldorsson (2010), Dasgupta and Menon (2011) and Cox *et al.* (2017) who argue along similar lines.
2. See Henrich *et al.* (2004) for a more thorough discussion of this research strategy. They have performed three different experiments (the Ultimatum Game, The Dictator Game and the Public Good Game) in 15 small-scale societies around the world. Henrich *et al.* (2004) document substantial variations in behavior among societies and this variation relates strongly with differences in social and cultural conditions.
3. In the World Value Survey (WVS), the variable "social trust" is based on a simple dichotomous question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" with the response alternatives "Most people can be trusted" and "Need to be very careful".
4. See Gibbons (2001) for a more formal proof of this result.
5. Two classical studies outside the field of economics that illustrate the importance of repeated interaction for the formation of trust are Coleman's (1988) discussion of the market for diamonds in New York and Kollock's (1994) discussion of raw rubber markets in Thailand.
6. Russel Hardin's (2002, 2006) famous theory of trust through encapsulated interests is very close to economists' understanding of trust. The encapsulated-interest conception of trust is based on the idea that A trusts B because A thinks it is in B's interest to take A's interests seriously. The most important reason for B to take A's interests seriously, according to Hardin, is that B wants the interaction with A to continue. B thus takes A's interests into account out of self-interest.

7. Although the trust game is popular and widely used in experiments, Cox (2004, 2009) argues that it has one important limitation. The trust game design does not make it possible to differentiate between transfers and returns resulting from trust and reciprocity and transfers resulting from other sources (e.g. altruism or inequality aversion). Cox (2004, 2009) offers a method for how this problem can be solved.
8. If norms are completely internalized, this conclusion does not hold. If norms are internalized, people follow them even when violation would be undetected and therefore unsanctioned. The main motivating force is the desire to avoid feelings of shame, guilt, regret and loss of self-respect if the norm in question is violated. See Elster (1989) and Horne (2009) for a more complete discussion of the difference and importance of external and internal sanctioning of norms.
9. NOK = Norwegian kroner. NOK100 equals approximately EURO10 (1 EUR = 10 NOK).
10. The numbers reported in Figure 1 are very close to the numbers reported in Bicchieri *et al.* (2011, Figure 1(a)).
11. The numbers reported in Figure 2 are also close to the numbers reported in Bicchieri *et al.* (2011, Figure 1(b)).

References

Alesina, A. and Glaeser, E.L. (2004), *Fighting Poverty in the US and Europe: A World of Difference*, Oxford University Press, Oxford.

Algan, Y. and Cahuc, P. (2013), "Trust, growth and well-being: new evidence and policy implications", in Aghion, P. and Durlauf, S. (Eds), *Handbook of Economic Growth*, Vol. 2A, Elsevier, Amsterdam, pp. 49-120.

Ashraf, N., Bohnet, I. and Piankov, N. (2006), "Decomposing trust and trustworthiness", *Experimental Economics*, Vol. 9 No. 3, pp. 193-220.

Ben-Ner, A. and Halldorsson, F. (2010), "Trusting and trustworthiness: what are they, how to measure them, and what affects them?", *Journal of Economic Psychology*, Vol. 31 No. 1, pp. 64-79.

Berg, J., Dickhaut, J. and McCabe, K. (1995), "Trust, reciprocity, and social history", *Games and Economic Behavior*, Vol. 10 No. 1, pp. 122-142.

Bicchieri, C. (2017), *Norms in the Wild*, Oxford University Press, Oxford.

Bicchieri, C., Xiao, E. and Muldoon, R. (2011), "Trustworthiness is a social norm, but trusting is not", *Politics, Philosophy & Economics*, Vol. 10 No. 2, pp. 170-187.

Bowles, S. and Gintis, H. (2006), "Prosocial emotions", in Blume, L.E. and Durlauf, S. (Eds), *The Economy as an Evolving Complex System, III*, Oxford University Press, New York, NY, pp. 339-366.

Brennan, G., Eriksson, L., Goodin, R.E. and Southwood, N. (2015), *Explaining Norms*, Oxford University Press, Oxford.

Chaudhuri, A. and Gangadharan, L. (2007), "An experimental analysis of trust and trustworthiness", *Southern Economic Journal*, Vol. 73 No. 4, pp. 959-985.

Chaudhuri, A., Ali Khan, S., Lakshmiratan, A., Py, A.-L. and Shah, L. (2003), "Trust and trustworthiness in a sequential bargaining game", *Journal of Behavioral Decision Making*, Vol. 16 No. 5, pp. 331-340.

Coleman, J.C. (1988), "Social capital in the creation of human capital", *American Journal of Sociology*, Vol. 94, Supplement, pp. 95-120.

Cox, J.C. (2004), "How to identify trust and reciprocity", *Games and Economic Behavior*, Vol. 46 No. 2, pp. 260-281.

Cox, J.C. (2009), "Trust and reciprocity: implications of game triads and social contexts", *New Zealand Economic Papers*, Vol. 43 No. 2, pp. 89-104.

Cox, J.C., Servátka, M. and Vadovič, R. (2017), "Status quo effects in fairness games: reciprocal responses to acts of commission versus acts of omission", *Experimental Economics*, Vol. 10 No. 1, pp. 1-18.

Dasgupta, U. and Menon, A. (2011), "Trust and trustworthiness among economics majors", *Economics Bulletin*, Vol. 31 No. 4, pp. 2799-2815.

Elster, J. (1989), "Social norms and economic theory", *Journal of Economic Perspective*, Vol. 3 No. 4, pp. 99-117.

Elster, J. (2009), "Norms", in Hedström, P. and Bearman, P. (Eds), *The Oxford Handbook of Analytical Sociology*, Oxford University Press, Oxford, pp. 195-217.

Fehr, E. (2009), "The economics and biology of trust", *Journal of European Economics Association*, Vol. 7 Nos 2-3, pp. 235-266.

Fehr, E. and Fischbacher, U. (2004), "Social norms and human cooperation", *Trends in Cognitive Sciences*, Vol. 8 No. 4, pp. 185-190.

Fehr, E. and Gächter, S. (2002), "Altruistic punishment in humans", *Nature*, Vol. 415 No. 10, pp. 137-140.

Fehr, E. and Rockenbach, B. (2003), "Detrimental effects of sanctions on human altruism", *Nature*, Vol. 422 No. 13, pp. 137-140.

Fukuyama, F. (1995), *Trust. The Social Virtues and the Creation of Prosperity*, The Free Press, New York, NY.

Gibbons, R. (2001), "Trust in social structures: Hobbes and Coase meet repeated games", in Cook, K.S. (Ed.), *Trust in Society*, Russel Sage Foundation, New York, NY, pp. 332-353.

Gintis, H. (2017), *Individuality and Entanglement: The Moral Bases of Social Life*, Princeton University Press, Princeton, NJ.

Granovetter, M. (2017), *Society and Economy: Framework and Principles*, The Belknap Press of Harvard University Press, London.

Gunnthorsdottir, A., McCabe, K. and Smith, V. (2002), "Using the Machiavellianism instrument to predict trustworthiness in a bargaining game", *Journal of Economic Psychology*, Vol. 23 No. 1, pp. 49-66.

Hardin, R. (2002), *Trust & Trustworthiness*, Russel Sage Foundation, New York, NY.

Hardin, R. (2006), *Trust*, Polity Press, Cambridge.

Hechter, M. and Opp, K.D. (Eds) (2001), *Social Norms*, Russel Sage Foundation, New York, NY.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. and Gintis, H. (2004), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford University Press, Oxford.

Horne, C. (2009), *The Rewards of Punishment: A Relation Theory of Norm Enforcement*, Stanford University Press, Stanford, CA.

James, H.S. (2002), "The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness", *Journal of Economic Behavior & Organization*, Vol. 47 No. 3, pp. 291-307.

Johnson, N.D. and Mislin, A.A. (2011), "Trust games: a meta-analysis", *Journal of Economic Psychology*, Vol. 32 No. 5, pp. 865-889.

Kollock, P. (1994), "The emergence of exchange structures: an experimental study of uncertainty, commitment, and trust", *American Journal of Sociology*, Vol. 100 No. 2, pp. 313-345.

Krupka, E.L. and Weber, R.A. (2013), "Identifying social norms using coordination games: why does dictator game sharing vary?", *Journal of the European Economic Association*, Vol. 11 No. 3, pp. 495-524.

Ostrom, E. (2000), "Collective action and the evolution of social norms", *Journal of Economics Perspectives*, Vol. 14 No. 3, pp. 137-158.

Pontusson, J. (2005), *Inequality and prosperity: Social Europe vs Liberal America*, Cornell University Press, Ithaca, NY.

Reiersen, J. (2017), "Trust as belief or behavior?", *Review of Social Economy*, Vol. 75 No. 4, pp. 434-453.

Rothstein, B. and Uslaner, E.M. (2005), "All for all: equality, corruption, and social trust", *World Politics*, Vol. 58 No. 1, pp. 41-72.

Rousseau, D., Sitkin, S.B., Burt, R.S. and Camerer, C. (1998), "Not so different after all: a cross-discipline view of trust", *Academy of Management Review*, Vol. 23 No. 3, pp. 393-403.

Sapienza, P., Toldra-Simats, A. and Zingales, L. (2013), "Understanding trust", *Economic Journal*, Vol. 123 No. 573, pp. 1313-1332.

Schelling, T. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

Simpson, B. and Eriksson, K. (2009), "The dynamics of trust and generalized trustworthiness", *Rationality and Society*, Vol. 21 No. 1, pp. 59-80.

Torsvik, G. (2000), "Social capital and economic development: a plea for the mechanisms", *Rationality and Society*, Vol. 12 No. 4, pp. 451-476.

Uslaner, E.M. (2002), *The Moral Foundation of Trust*, Cambridge University Press, Cambridge.

Further reading

Cox, J.C., Kerchbamer, R. and Neururer, D. (2016), "What is trustworthiness and what drives it?", *Games and Economic Behavior*, Vol. 98, July, pp. 197-218.

Kreps, D.M. (1990), "Corporate culture and economic theory", in Alt, J. and Shepsle, K. (Eds), *Perspective on Positive Political Economy*, Cambridge University Press, Cambridge, pp. 90-143.

Appendix

In the following we provide an English translation of our Norwegian instructions and questionnaire. As noted, the questionnaire is originally developed by Bicchieri *et al.* (2011). We have modified the questionnaire slightly. Some questions were added while others were removed. Norwegian instructions and questionnaire are available on request.

Instructions

In this experiment, two participants are randomly and anonymously paired up. One plays as Actor A and the other plays as Actor B. At the beginning of the experiment both actors receive an initial endowment of NOK 100. First, A can transfer, from his endowment, any amount from NOK 0 to NOK 100 to B. The experimenters will triple this transferred amount, so that B receives three times the amount that A transferred. Then, after A's decision, B can return to A any percentage of the tripled amount he received.

Final payoff for A	**Final payoff for B**
A receives NOK100	B receives NOK100
– Transfer to B	+ 3 *x* transfer from A
+ Back-transfer from B	– Back-transfer to A
= A's payoff	= B's payoff

In order to answer the following questions, you need to know there are _____ participants in total in today's session.

Part 1

In the following question you will decide whether to impose a payoff cut to actor A's final payoff. The payoff cut amount does not go to either B or you.

Scenario 1: Imagine that actor A transfers NOK 0 (out of NOK 100) to actor B.

I would choose: ___ (Fill in one of the alternatives below).

a) No payoff cut to actor A.
b) To deduct 10% from actor A's earnings.
c) To deduct 30% from actor A's earnings.
d) To deduct 50% from actor A's earnings.
e) To deduct 70% from actor A's earnings.
f) To deduct 90% from actor A's earnings.
g) To deduct all the earning from actor A.
h) Please briefly explain your decision here:

How many participants in today's session do you think chose a) no payoff cut to actor A? Answer: _____

What is the option that you think most participants chose today? Answer: ___

Part 2

In the following question you will decide whether to impose a payoff cut to actor B's final payoff. The payoff cut amount does not go to either A or you.

Scenario 1: Imagine that actor A transfers NOK50. B then receives NOK150. Actor B transfers back NOK0 (keeps everything for himself). Therefore, at the end of the experiment, actor A receives NOK50 and actor B receives NOK 250.

I would choose: _____ (Fill in one of the alternatives below).

a) No payoff cut to actor B.
b) To deduct 10% from actor B's earnings.
c) To deduct 30% from actor B's earnings.
d) To deduct 50% from actor B's earnings.
e) To deduct 70% from actor B's earnings.
f) To deduct 90% from actor B's earnings.
g) To deduct all the earning from actor B.

Please briefly explain your decision here:

How many participants in today's session do you think chose a) no payoff cut to actor B? Answer: _____

Which payoff cut in percent (if any) do you think most participants chose today? Answer: ____

Scenario 2: Imagine that actor A transfers NOK 50. B then receives NOK150. Actor B transfers back NOK 50. Therefore, at the end of the experiment, actor A receives NOK100 and actor B receives NOK 200.

How many participants in today's session do you think chose "No payoff cut to actor B"? Answer: ____

Which payoff cut in percent (if any) do you think most participants chose today? Answer: _____

Scenario 3: Imagine that actor A transfers NOK 50. B then receives NOK 150. Actor B transfers back NOK 100. Therefore, at the end of the experiment, actor A receives NOK150 and actor B receives NOK 150.

How many participants in today's session do you think chose "No payoff cut to actor B"? Answer: ____

Which payoff cut in percent (if any) do you think most participants chose today? Answer: ____

Part 3

Scenario 1
Imagine that you are actor B in the situation explained above. You have received NOK100, the same as A. Actor A decides first. He is free to transfer any amount between NOK0 and 100. You receive three times the amount that A has transferred. Then, after A's decision, you can return to A any amount of what you received.

Suppose that A transfers all his initial endowment, that is NOK100. You receive NOK300. In addition, you have your initial NOK100 (in total NOK 400). How much would you like to transfer back to A? Answer: NOK _____

Scenario 2
Imagine the same situation as in scenario 1 above.
Suppose that A transfers NOK 20. You receive NOK 60 from A. In addition, you have your initial NOK100 (in total NOK160). How much would you like to transfer back to A? Answer: NOK _____

Scenario 3
Imagine that you are actor A in the situation explained above. How much would you like to transfer to B? (Remember that B receives three times the amount that you transfer, and B has the possibility to return money.)
Answer: NOK _____

We thank you for your participation!

Corresponding author
Jon Reiersen can be contacted at: jon.reiersen@usn.no