Analele Universității din București. Filosofie
Vol. LXXII – Nr. 1, p. 85-106
ISSN 0068-3175; e-ISSN 2537-4044
DOI: 10.62229/aubpslxxi/1_23/5

Maria BANU[1]

WHOM DO WE TRUST?
ON HOW WE ASSESS OTHERS' TRUSTWORTHINESS

Abstract. The aim of this paper is to integrate competing notions of trustworthiness in the literature on trust under a common framework. I defend a notion of trustworthiness around three criteria: competence, predictability, and responsiveness. These are both necessary and sufficient conditions for trustworthiness assessments. Competence means having the required abilities to fulfill one's trust in a certain matter, but also the capacity to understand, assess, and choose the best way to apply those abilities in that matter. Predictability means both reliability and having the relevant reasons to fulfill one's trust in that respective matter. Responsiveness captures the trustee's intentionality about or directed at the trustor, her willingness to fulfill her trust. The three criteria are not fixed characteristics of the trustee; assessments of each will reflect aspects that are relevant under specific trust situations. The criteria seek to establish a common ground for trust research to reconcile different perspectives, while recognizing that trust is a highly contextual and relational notion.

Keywords: Trust, trustworthiness, reliance, competence, predictability, responsiveness

Introduction

Few topics stirred such an interest in the social sciences like trust has. Trust has been studied in economics, sociology, psychology, political sciences, behavioral economics, neurosciences, and evolutionary biology. All produced remarkable results for our understanding of trust. Yet, trust

[1] University of Bucharest, Faculty of Philosophy. PhD Candidate, Email: <maria.banu@drd.unibuc.ro>.

EDITURA UNIVERSITĂȚII DIN BUCUREȘTI
BUCHAREST UNIVERSITY PRESS

research now deals with great conceptual and theoretical fragmentation. "There are at least as many conceptualizations of trust as there are disciplines in the social sciences" (Cook and Santana 2018, 253). Trust research has become a label for (sometimes radically) different definitions and measurement methodologies.

One cause is the clash of different intuitions about trustworthiness. Most accounts of trust take "I trust you" as equivalent to "I consider you trustworthy" (*e.g.*, Coleman 1990; Baier 1996; Jones 1996; Gambetta 1988; Hardin 2002). Trustworthiness is hardly treated in the literature, although "the complexity of the problem of trust derives primarily from the complexity of the problem of trustworthiness" (Hardin 2002, 31). Philosophers discuss it only to the extent it helps them separate trust from reliance. Reliance can happen without trust, and some think the difference lies with the motivations of the trustee to fulfill the trustor's trust (*e.g.*, Baier 1986). Yet, they explain trustworthiness in terms of opposing motivations, like interests (Hardin 2002) versus goodwill (Baier 1986). Other philosophers discount motivations altogether: one is trustworthy when she fulfills her commitments (Hawley 2014), obligations (Hertzberg 1988; Hollis 1998), or, they argue, trust itself elicits trustworthiness (Holton 1994).

My objective in this paper is to build an integrated concept of trustworthiness that will reconcile different perspectives on trust and build common ground among trust researchers. I defend an account of trustworthiness around three criteria: competence, predictability, responsiveness. They are necessary and sufficient conditions to consider one trustworthy. I take trustworthiness as a context-dependent notion which describes a three-part relationship: *B* is trustworthy to *A* with *X* (in context *C*, at moment *t*). From this perspective, competence means having the required abilities to fulfill one's trust with respect to *X*, but also the capacity to understand, assess, and choose the best way to apply those abilities and perform the task one is entrusted with. Predictability means both reliability and having the relevant reasons to fulfill one's trust with respect to *X*. Responsiveness captures the trustee's intentionality *about* or *directed at* the trustor, her willingness to meet the trustor's needs with respect to *X*.

Section I discusses the link between trust and trustworthiness and accounts of the latter in the literature. Section II deals with the three criteria, what they mean for trust and how we assess them. Section III addresses an immediate objection to my account of trustworthiness in the context of instant trust. I then conclude on my accounts' added value to trust research.

I. From trust to trustworthiness

Trust caught the attention of social scientists for its benefits both to personal relationships and society in a broad sense. Social trust, the kind we have in strangers, is a great and cheap vehicle for social coordination and collective action. High trust in a society fosters economic prosperity (Knack & Keefer 1997; Zack & Knack 2001) and supports democracy (Putnam 1993). The puzzle social scientists endeavor to solve is how trust can rationally emerge if we assume, like neoclassical economic theory does, that individuals are self-interested. Social capital theorists (Coleman 1990; Putnam 1993), who enshrined trust as an 'it' topic of social research, argue that trust is upheld by the perception that those around us are trustworthy. Trust is rational when potential gains from our interactions with others are higher than potential losses, and others are trustworthy (Coleman 1990, 99).

Trustworthiness in early accounts of trust[2] is unsophisticated. It means that the trustee "fulfill[s] his part of the agreement" (Dasgupta 1988, 51) or "perform[s] an action that is beneficial or at least not detrimental to us" (Gambetta 1988, 217). Yet, beliefs about others' trustworthiness are constitutive of trust. It makes no sense to discuss about trust unless it implies at least a minimal degree of belief that the other person is trustworthy. Otherwise, what we call trust may be simply reliance or cooperation. We may decide to rely on or cooperate with others for various reasons, even when we *dis*trust them. What separates trust from reliance or cooperation is then the belief that the other is trustworthy. This belief may be unwarranted; it may not rely on good evidence, but it must exist to be able to talk about trust.

[2] See, *e.g.*, Coleman (1990), Yamagishi (1998), Dasgupta (1988), Gambetta (1988).

Later theories of trust ask why we trust others, and, in this context, they address what motivates trustworthiness. If selfishness motivates you to fulfill my trust, can I trust you? In his theory of trust as encapsulated interest, Hardin (2002) thinks self-interest is a valid motivation for trustworthiness. We trust others if we think it is in their interest to consider ('encapsulate') our interests in the relevant matter into their own. To others, trust is to depend on others' goodwill or, at a minimum, lack of ill will toward us (Baier 1986, 234; Jones 1996). Later, Jones (2012) argues that a trustworthy trustee must be competent in the domain of our interaction with her and must be motivated by the fact that we are counting on her. Other philosophers think motivations are not at all important. To them, trust has a normative dimension. We trust those who fulfill their obligations (Hertzberg 1988; Hollis 1998) or commitments to us (Hawley 2014) or those who take the fact that we trust them as reason enough to answer positively to our trust (Holton 1994), irrespective of their motivations.

The different intuitions about trustworthiness led to competing understandings of trust. Hardin (2002, 10) argues that trust is cognitive; it is a belief about trustworthiness. It is not purposive (Baier 1986, 235), nor behavioral (Hardin 2002, 10). It is not a decision, nor an action. Baier (1986) thinks theories like Hardin's leave out many forms of discretionary trust, where there is high asymmetry in power. Jones (2012) argues that interests are not a stable motivation for trustworthiness, but neither goodwill is a necessary or sufficient condition. In contrast, those who discount motivations imply that we can trust without believing those we trust are trustworthy (Hawley 2014, 2030). This means trust is non-cognitive and that we can even trust at will (Holton 1994; Faulkner 2014, 1979). Jones (1996; 2019) thinks we cannot trust at will, but trust is non-cognitive in that it is an affective attitude that elicits trustworthiness.[3]

The account I discuss further does not assume beliefs about trustworthiness paint the whole picture of trust. They are necessary to trust, but trust is not a simple reflection of such beliefs. Two people might differ in their assessment of the same person's trustworthiness.

[3] See Adler (1994), Fricker (2006), Hieronymi (2008) on cognitive trust and Becker (1996), Jones (1996; 2019), Lahno (2001; 2020) on affective trust.

Yamagishi (1998, 42) explains that these are not "random errors," nor an "inability to discern that person's trustworthiness precisely," but a matter of differences in trustfulness. We all have a disposition to trust in general. Some are high trustors, others are not. Trustfulness provides a lens through which we perceive others, especially when information is scarce. High trustors will be inclined to look for signs of trustworthiness in others[4]; low trustors will focus on cues that will confirm their already low general expectation. Assessments of trustworthiness are fully subjective (Yamagishi 1998, 60). Had they been objective, we would only discuss about trustworthiness.

This is why I adopt the trustor's stance in defining trustworthiness. I see trustworthiness as specific to the trust relationship between a trustor and a trustee. Trust is context dependent and captures a three-place relation between the one that trusts (*A*), the recipient of trust (*B*), and the object of trust (*X*). This relation happens in a certain context (*C*) and at a certain moment in time (*t*). So, '*A* trusts *B* with *X* in *C*, at *t*.' Conversely, '*B* is trustworthy to *A* with *X* in *C*, at *t*."

II. Three criteria for trustworthiness

In defining the three criteria, my question is: how do we assess trustworthiness in others? The criteria of competence, predictability, and responsiveness are not general characteristics of the trustee. They reflect the trustor's perception of the trustee with respect to a certain matter, in a certain context, and at a certain moment in time. Each criterion is necessary to perceive someone as trustworthy. I will not trust you if I think you are incompetent in the relevant matter, regardless of your good intentions. I will not trust you if I find you unpredictable; I must be

[4] The fact that *A* is a high trustor in general does not mean *A* will trust *B* more. It simply means that she will be more inclined to enter a trust relationship with *B*. Yamagishi (1998) shoes that high trustors are not gullible; they possess 'social intelligence' which allows them to detect cues of (un)trustworthiness in others. In the end, *A*'s trust in *B* will depend on how trustworthy *A* thinks *B* is. Compared to low trustors, high trustors are more sensitive to positive and negative information about others' trustworthiness (Yamagishi 1998, 24).

able to believe, with a certain degree of certainty, that you will fulfill my trust, based on my assessment of your previous behaviors and your reasons to fulfill my trust. I will not trust one that is unresponsive, that is, unwilling to respond positively to my trust. One is trustworthy with respect to some matter in a certain context if she meets all three criteria.

II.1. Competence

Competence is a default requirement of trustworthiness. It is particularly important when we deal with professionals because we expect them to possess a certain level of technical ability. The issue of competence rarely emerges in personal relationships or casual interactions because we usually take for granted that people are competent in these contexts. For instance, I assume my friend understands the norms of friendship and the expectations that derive from my calling her 'my friend.' It doesn't mean competence is not important in these contexts. Yet, when we fail in such matters, the odds are that those who trusted us would think we did it with bad intention rather than a lack of competence (who doesn't know what honesty is, right?). Trust scholars rarely discuss competence in their inquiry into trustworthiness and rush to motivations. Some exceptions are Hardin (2002), Jones (1996; 2012), and Mayer *et al.* (1995). There may be yet another reason why competence is less controversial: incompetence is usually not as morally blameworthy as ill intention is.

As trust varies with context, competence is not a fixed trait of the trustee; it can refer to various types of abilities that are relevant for the specific trust situation. As Jones (1996, 7) points out, the competence we expect from a professional is technical, the one we expect from strangers amounts to an "understanding of the norms for interactions between strangers," the one we expect from friends "is a kind of *moral* competence." Nevertheless, a trustworthy trustee must possess the necessary abilities to fulfill our trust in the relevant matter, whatever that may be. This is the first meaning of competence I advocate for here. Competence may be specific to a domain, comprising a set of technical skills applicable to a particular profession or field. Or it may refer to a broader suite of abilities which may apply in different contexts. Mayer and colleagues (1995, 717) define it as a group of abilities and attributes that enable one to exercise influence in a certain domain. They point out, though, that specialized skills are insufficient to perceive one as competent. Technical excellence must be accompanied by other skills (like communication skills) to ensure successful performance on a task we are entrusted with.

There is yet a second, more subtle meaning of competence. Imagine you work with a person with a lot of experience and an impressive record of delivered projects. These provide you with good reasons to trust her. Yet, you are wary of trusting her. You noticed several times that she was unable to adequately tailor her expertise to *your* project's needs. It is not that she lacks the technical skills, but the ability to apprehend what you need and to customize her response to your needs. Indeed, trust is not instrumental. We do not trust others with the aim of getting something from them; it is *because* we trust them that we expect something from their part (Hardin 2002, 10). There is, though, an instrumental element to trust. When we trust, we delegate things we cannot or will not do ourselves to others. In doing so, we expect others not only to 'get the job done' but to 'get it done right.' So, competence, I argue, is not just about having the relevant abilities to perform a task, but also the capacity to exercise one's own judgement and come up with the appropriate ways to apply those abilities.

Assessing competence can be challenging. How can I assess professionals' competence if I do not possess such competence myself? First, we may judge competence in relative terms: I will trust more a lawyer with 20 years of experience than a recent law graduate. Second, we may use cues that are unrelated or remotely linked to competence, like one's appearance, demeanor, decisiveness, language used. Such assessments can be flawed and exploited, but we use them every day. Finally, our societies are generally designed to help us assess others' competence, professionals in particular. There are formal and informal norms and institutions in place that signal competence. We have "agencies that assess the competence of such professionals as doctors, lawyers, and even mountain-climbing guides" (Hardin 2002, 8-9). Diplomas, certificates, letters of recommendation, reputation – they all work for this purpose. They provide indirect means to judge specialists' competence.

The latter is mediated trust, though. It emerges from our underlying trust in institutions as well as in the wider, complex web of social and personal relations that we are part of. The question is whether these indicators can in fact amount to good evidence for competence remains open. Can I trust a professional services company with good reviews on Google? In practical terms, the rate of success is good enough; scarce, indirect information is better than no information. This is why agents (individuals, public or private organizations) invest so much in branding and reputation. Yet, it should not surprise us if one lost confidence in an entire system from just one bad experience. Uncertainty is an integral part of trust and people often rely on feeble and fast generalizations to assess trustworthiness in others and decide what to do next.

II.2. Predictability

Competence alone does not render one trustworthy. We must also believe, with a sufficient degree of certainty, that the trustee will in fact fulfill our trust. A trustee is predictable in the sense that we can predict her positive response in specific trust situations. The concept of predictability borrows from rational choice and game theory, where it refers to the probability with which one player can anticipate the choice of another and adjust her actions accordingly. Such predictions build on beliefs about the other player's motivation to choose one course of action rather than another. In the context of trust, an assessment of the trustee's predictability must result in an estimate of the probability with which we believe that she will fulfill our trust in a certain matter. This estimate is inherently subjective and typically informed by the trustee's past behaviors, as well as assessments of her reasons to fulfill our trust.

Predictability, as I define it here, captures two elements: reliability and one's reasons to fulfill another's trust. As philosophers separate trust from reliance, they focus on the motivations that render one trustworthy rather than just reliable (*e.g.*, Baier 1986; Jones 1996; Petit 2002). My notion of predictability, though, emphasizes both reliability and the reasons for fulfilling another's trust. A trustworthy person must be generally reliable; she must show consistency in the way she speaks and acts. We do not typically trust the unreliable, those who swiftly change their minds or moods or who never fulfill their commitments. Despite their potentially good intentions, we avoid putting things we care about in the hands of the unreliable.

This is why, I argue, the separation operated by philosophers between trust and reliance adds little to our understanding of trust in real life. If trust is a species a reliance (*e.g.*, Baier 1986, 234; Pettit 2002, 364), every time we trust we rely. We may rely without trust, and we may trust without acting on it (that is, again, *relying*). Outside these conceptual prerequisites, though, trust means little if we do not act on it, namely if we do not rely on those we trust. You'd be right to tell me I don't trust you if I never rely on you for anything. If we *must* separate trust from mere reliance, the distinction lies not in the motivations of the trustee, but in the overall belief that the trustee is trustworthy. This separates reliance on one we trust from reliance on one we don't trust (or distrust). This belief is more complex than the mere assessment of the trustee's motivations; it includes criteria like competence and responsiveness.

One is generally reliable if her actions and speech are typically consistent over time. The most relevant indicators of consistency in the context of trust are sincerity and promise-keeping. They can turn trust from a matter of degree to an all-or-nothing game because, once we lie or fail to keep our promise, we expose ourselves to the risk of never being trusted again. Sincerity captures trustworthiness in speech, while it promises to link speech to action. Sincerity means not just telling the truth but telling it *completely* and *accurately* (Williams 2002, 124). As Williams explains, utterances yield sets of implicit or explicit beliefs, based on which different audiences may form different beliefs. Beliefs we acquire via those we trust further guide our actions, so untrustworthiness in speech can have, to say the least, unpleasant consequences. On the other hand, promises are a "mysterious and incomprehensible" (Hume 1986 [1739-40], 3.2.5) vehicle for trust, because they seem to *create* trust and do so at the will of the trustor (Baier 1986, 245). Although, we cannot decide to trust just as we cannot decide to believe something at will.

Now, a generally reliable person may not welcome my trust in certain matters, which means I cannot trust her in those matters. To trust her, I must know she has the relevant reasons to fulfill my trust. This will allow me to predict, with a certain degree of certainty, if she will fulfill my trust in the relevant matter. This is the second meaning of predictability I advance in this paper. The trustee's motivations to fulfill others' trust is where philosophers usually clash. Competing views on the importance and nature of motivations for trustworthiness have led to opposing theories of trust in literature. While the trustee's motivations seem key to explain trust, the debate around interests (Hardin 2002) versus goodwill (Baier 1986) is difficult to reconcile. In contrast, non-motives-based theories stress that motivations do not and should not matter to trust. Commitments (Hawley 2014), obligations (Hertzberg 1998), or the act of trust itself (Holton 1994; Jones 2012) may elicit trustworthiness, irrespective of the trustee's motivations.

Some philosophers think it is important to know what motivates those we trust because there are motivations that are incompatible with trustworthiness, like ill will (Baier 1986), fear, or hatred (Jones 2012; 2019). Meeting another's expectation out of fear does not render one trustworthy. Yet, non-motives-based theories do have a point. There may be contexts where we decide to act on reasons that go against our motivations. For example, I may fulfill your trust because it is the right thing to do, although I might hate you. As complex beings acting in complex environments, it is strange to imply that people have or should have unique motivations to act. A mixture of factors usually determines action (Hausman 2012, 36). As Hausman would argue, the choice of an action does not consider just the outcome of that action, like the fact that I will gain something from you, but also the meaning of that action, like the fact that it is wrong to betray others for personal gains.

One solution to this problem is motivational pluralism, whereby individuals are driven by a variety of motivations rather than a single dominant one. This perspective recognizes the complexity and diversity of human motivation, asserting that people have multiple motivations that can vary in strength and importance depending on the context and individual differences. Motivational pluralism is less interesting when our reasons to act converge toward the same action (Sober and Wilson 1998). It may happen you trust me with something that is compatible with both my interests and moral values. When reasons conflict, though, we tend to assume individuals will choose an outcome at the expense of another. Yet, Hausman (2012) argues that it may be more reasonable to assume our preferences over different courses of action will influence each other in shaping our final choice of an outcome. The value I place on fulfilling the promise I made to you may soften my temptation to betray you for a higher personal profit.

From this angle, endeavoring to identify the one motivation that is specific to trustworthiness is misfocused. Both Hardin's and Baier's theories face similar issues. First, neither interests, nor goodwill is necessary for trustworthiness. I may have no interest in fulfilling your trust, yet I may do it out of sympathy, friendship, commitment. Similarly, I may have enough goodwill to fulfill your trust, but other interests might trump on that goodwill and determine me to betray you. Second, both interests and goodwill are rather loose concepts. Hardin (2002, 4) thinks trustworthiness may be motivated by material interest only in a minimal sense; to him, interests have a larger, deeper sense. Yet, he expands them to the point where they accommodate almost anything. You may be trustworthy to me because you "may enjoy doing various things with me or you might value my friendship or my love, and your desire to keep my friendship or love will motivate you to be careful of my trust." Baier refrains from clearly defining goodwill; she allows us to infer its meaning from examples she discusses. Jones (2012, 67) argues that if we identify goodwill with feelings of friendship, then Baier's theory becomes too restrictive. If we enlarge the notion enough to include a broad meaning of goodwill, honesty, awareness, it becomes meaningless. It would mean that a trustworthy person simply needs to have a positive reason to fulfill our trust, whatever it may be, and it would be enough to be trustworthy. So, neither Hardin, nor Baier can find a motivation for trustworthiness specific enough not to have its meaning slip away and large enough to account for the different shapes and forms of trust.

The account I propose in this paper focuses on reasons rather than motivations when assessing trustworthiness. The two notions are related and often used interchangeably, but there are a few important differences. Reasons are circumstances that can cause a certain action and explain or justify why we acted like that. They are often based on factual premises. For instance, the reason I study is to pass the exam. Motivation, on the other hand, refers to internal or external driving factors behind a person's actions or behaviors. Motivation also answers the question of 'why' we act a certain way but, unlike reasons, it links the answer to desires, needs, aspirations, goals. For example, the motivation to study could be the desire to achieve good grades or the ambition to excel in a domain. Reasons are often more objective and subject to logical scrutiny, motivations are more subjective, personal, and they explain the psychology behind our actions.

We shouldn't disregard motivation in trustworthiness assessments. Understanding motivations can provide valuable insights into a person's character, long-term reliability, and commitments. Yet, a reasons-based assessment of trustworthiness is more aligned with the idea of trust being highly context dependent. Motivations like interest and goodwill transcend context, they may apply in several situations whereas, in others, people may even decide to act against them and in favor of other, stronger reasons arising from circumstances. Reasons may be (Alvarez 2017): (i) normative, "which, very roughly, favor or justify an action, as judged by a well-informed, impartial observer," and (ii) motivating, that is "reasons the 'agent' (that is, the person acting) takes to favor and justify her action and that guide her in acting."

Normative reasons become motivating when we act on them (Parfit 1997). This allows us to integrate non-motives-based theories into a comprehensive account of trustworthiness. While non-motives-based theories discount motivations as important for trustworthiness, they discuss normative reasons based on which trustworthy people act and should act. They base the belief about the trustee's trustworthiness in her ability to fulfill a commitment (Hawley 2014) or obligation (Hertzberg 1988; Hollis 1998). Hawley argues that we may choose to fulfill our commitments even when we do not want to. Her argument goes like this. To trust someone to perform a task is to think that she has a commitment to do that task. Similarly, to not trust means to think that the other person has a commitment to do something, but that she has no reason to do it. Borrowing on Hawley's example, I promised to attend your birthday party, but I decided not to; my commitment stands, but I have no intention to deliver on it. Now, I might not abide by my commitment, by if I still decide to act on it then it becomes a motivating reason. If I fail to meet my commitment, this is not because commitments do not offer reasons for action, but simply because other reasons trumped on my commitment. I need not wish to fulfill a commitment for that commitment to constitute a good reason to do it anyway.

So, apart from interests and goodwill, other reasons may elicit trustworthiness and trust: moral commitments, obligations, norms. I would add here feelings, like sympathy, affection, love, friendship, and even external constraints, like sanctions. We may even choose to fulfill others' trust because we like to think of ourselves as trustworthy individuals. Or, as Jones (1996; 2012) argues, the simple fact that one is counting on us is a sufficient reason to respond positively to that trust. In everyday interactions, there is usually a mix of reasons to fulfill others' trust. Sometimes, reflection on such reasons may reveal some tension. In her reply to Hardin, Jones (2012, 70) argues that interests are a compatible but unstable motivation for trustworthiness. One must have an additional incentive to interests to uphold trustworthiness, she thinks. Indeed, but it is not necessarily so. Reasons to fulfill trust depend on the context of that trust and the stakes. Goodwill may be a good enough reason in certain situations but not in others.

As reasons to fulfill one's trust will vary with the specificity of the trust situation, the trustor's task to assess the trustee's reasons is not easy. Epistemic access to another's reasons to act may be difficult to get, especially if those reasons reflect on their internal psychological life. Sometimes we are lucky and those we trust make their reasons known to us (assuming they are honest). Other times, it is up to us to determine their reasons. While it may not be simple, we do such assessments every day. We form beliefs about others' intentions, goals, motivations, beliefs, wishes. We rely on all sorts of information, beyond what people tell us. We rely on third-party information, observations of past behaviors, non-verbal and paraverbal cues, like face expressions, tone, the use of certain words or phrases. This helps us "read" others and explain their behaviors and actions. Competing theories in the philosophy of mind[5] explain how

[5] See, for instance, Sellars (1956) and Lewis (1970; 1972) on folk psychology as a theory of mind and Gordon (1986) and Heal (1994) on simulation theory.

this is possible and how it works. Most of "mind reading" is automatic, unconscious, and biased, indeed, but the fact is that we use it every day.

II.3. Responsiveness

I may think of you as a competent and reliable person. I may think you could have the relevant reasons to perform a task I entrust with and do it well. Yet, if I feel that you are rather unwilling, or indifferent, or you do not care to answer positively to *my* relying on you, then I will find it difficult to believe that you could be indeed trustworthy *to me*. Responsiveness is a special criterion for trustworthiness. Trust and trustworthiness are relational concepts; they happen within the relationship between the trustor and the trustee. They describe that relationship, not general characteristics of the trustor or the trustee (Hardin 2002, 88). Responsiveness, in this sense, captures the trustee's intentionality *about* or *directed at* the trustor, her willingness to meet the trustor's expectations, irrespective if she knows if she is being trusted or relied upon.

I borrow the concept of responsiveness from Jones (2012). Yet, the meaning I convey to it differs significantly from hers. Jones assimilates responsiveness to motivations. She argues that trustworthiness, in a certain domain and toward a certain trustor, results from "competence together with direct responsiveness to the fact that the other is counting on you" (2012, 62). This builds on her older definition of trust as an attitude of optimism about another's goodwill and competence, together with the expectation that the trustee "will be directly and favorably moved by the thought that someone is counting on her" (Jones 1996, 8). In her later work, she criticizes Baier's notion of goodwill and argues that "[t]here is a minimal sense in which the trustworthy can indeed be said to have goodwill toward the trustor: just in virtue of being positively responsive to the fact of someone's dependency" (Jones 2012, 68-69). Jones' responsiveness mixes the motivational and the normative dimensions of trustworthiness. The fact that someone relies on us must be a good enough reason to fulfill that person's trust.

In my account, responsiveness does not amount to a motivation or a reason to fulfill one's trust. It reflects the attitude of the trustee toward the trustor, her willingness to account for the trustor's needs and, if properly motivated, to abide by them. Acknowledgment of the fact that someone might count on me may motivate me to respond positively to their trust (in the matter at hand). Yet, I may still show responsiveness in the absence of this acknowledgement. Jones' responsiveness means that the trustee must acquiesce to the thing she is being trusted with. The trustee must be "directly and favorably moved" by someone else's reliance on her. This is where I fully break away from Jones'. Suppose your friend calls you in the middle of the night and asks you to rescue her from a messy situation. She knows you do not want to do that, yet she knows she can trust you. This is not a situation where you, the trustee, acquiesce to what she asks of you. You are responsive to her trust, in that you acknowledge her need, you are there for her, although you dislike being called on to do that. Responding positively to someone else's trust does not mean that we acquiesce to what they are asking of us. Many times, trustworthiness means being there for those that trust us even though we disagree or disapprove.

Responsiveness, as I define it here, is closely linked to competence. It reflects on our capacity to assess and choose the right way to fulfill someone else's trust. Trust is an expectation that others will answer appropriately to it, not more, not less. Sometimes we define exactly what we need. Most times, we expect the trustee to decide what are the appropriate ways to answer to our trust. This means that they must be able to understand, assess, and act in virtue of their understanding of our needs, desires, and beliefs. In turn, this requires that the trustee is sensitive and attentive to our needs, that is, she is responsive in relation to us.

At a minimum, we are responsive when we exhibit at least some consideration for others' needs, goals, beliefs. To Maister *et al*. (2000, 91), the more self-centered we are, the less others will trust us; typically, a responsive trustee promptly acknowledges others' needs, communicates timely and effectively, keeps her promises, takes responsibility for mistakes, and seeks to address misunderstandings. She is open and honest, even on difficult or uncomfortable topics, and offers her support when needed or when she is being called for. It is hard to exhaust all the behavioral cues that could determine us to assess others as responsive to us, since trust is deeply contextual. Sometimes, the simple fact that you take the time to listen to me may determine me to trust you. Other times, too much concern for my well-being may even trigger suspicion.

III. Instant trust

I address here an immediate concern my account of trustworthiness raises. One might argue that the notion I propose is too burdensome and that people rarely perform such complex assessments of others' trustworthiness in everyday trust situations. Plus, we rarely have access to relevant information that could feed such assessments. To counterargue, I will discuss the case of instant trust. Instant trust happens when we quickly and instinctively trust another person although we just met her, often without any prior information or proof of her trustworthiness. This is the type of trust we have in strangers; it is immediate and implicit. The literature on trust focuses a great deal on it because of its potential in driving collective action and social coordination. It is also known as spontaneous, swift, or thin trust.

The concern I discuss here rests on two erroneous assumptions. First, there are forms of trust, like instant trust, which do not require assessments of trustworthiness, at least not such complex assessments. In such cases, one may argue trust is determined by other factors, like context, stakes, the trustor's availability to trust or other individual characteristics of hers. In fact, behavioral economists focus a great deal on identifying what determines trust and cooperation in situations where we know nothing about the other person[6]. The second assumption is that trustworthiness assessments in the sense I discuss in this paper require some sophisticated cognitive abilities, whereas trust is often borne by affects or emotions.

In reply, my paper starts off from the notion that there is no trust in the absence of at least a feeble perception or belief that the other

[6] See, for instance, Berg *et al*. (1995), Camerer (2003).

person is trustworthy. As I already discussed, this is what separates trust from reliance or cooperation. This belief rests on the assessment of the three criteria. Now, the content that feeds the assessment under each criterion will differ from one context to another and even from one trustor to another. One criterion may even prevail over the others, like in the case of trust in professionals, where competence is the first thing we care about. Or, in other contexts, I may be inclined, for instance, to trust more those that abide by strong moral principles and discount other reasons.[7]

Trustworthiness assessments may, thus, come in different shapes or forms. They are not always well-grounded or justified. They may rely on little information. We may perform them without being aware. In real life they rarely follow this exact conceptual model or use these exact terms, like competence, predictability, responsiveness. Yet, the things we look for in other people, in situations where we trust or trust could emerge, must amount to these three criteria. Each is a necessary condition for trustworthiness, but they must all be met to form the belief that the other person is trustworthy. The processes by which we form beliefs about other people's trustworthiness are not always conscious, but they are complex, even in cases of swift trust. Trust may happen in an instant, but that does not mean that it does not rely on any kind of assessment of trustworthiness or that it is simplistic.

Instant trust emerges from fast, intuitive judgments ("I have a great feeling about this person!"). They result from observations of body language, facial expressions, tone of voice, demeanor. Studies show that physiognomy, facial expressions, and emotions have a great impact on how trustworthy we perceive others to be (*e.g.*, Todorov *et al*. 2008; 2009). In new encounters, we also judge others' trustworthiness based on how familiar they feel to us (FeldmanHall *et al*. 2018). Dunn and Schweitzer (2005) show that we are more prone to trust happy or grateful people rather than angry or sad individuals. From an evolutionary point of view, fast judgements of this kind helped us distinguish friends from foes (Kahneman 2011, 25).

[7] Everett and colleagues (2016) show that we perceive those that tend to make deontological moral judgments as more trustworthy compared to consequentialists.

So, we are hard-wired to use such cues in our belief-forming and deliberative processes. We are prone to seek patterns and use simplistic cognitive shortcuts (Kahneman & Tversky 1979) to navigate complex situations. Heuristics are very useful in this sense, they reduce cognitive burden, and their outcomes are often "good enough" (Simon 1997). They do not help with estimations of statistical probabilities and can lead to systematic errors of judgment (Kahneman and Tversky 1979, 1124-1129). Yet, when it comes to trust, it seems we are quite well equipped to 'read' signs of (un)trustworthiness in others. According to Yamagishi (1998), this capacity is a byproduct of social intelligence, and we can invest cognitive resources to train it. Importantly though, we are often successful with quick judgments, but proof of their accuracy is very weak (Uddenburg *et al*. 2020) and they also pose a moral question because of implicit biases.

Trustworthiness assessments may or may not be warranted[8] and their sources may be cognitive, affective, or both.[9] This is, indeed, important since acting on trust involves risk. My account of trustworthiness offers a conceptual model for the study of trust and trustworthiness. Assuming our assessments of others' trustworthiness follow the three criteria and they rely on good evidence, then we can consider our trust warranted. There is, however, a question of what counts as good evidence, and I argue that 'good evidence' will differ depending on the context and stakes of the trust situation. Trusting a stranger to give you correct indications to reach your destination is an entirely different thing than trusting her with the keys to your house. What may warrant trust in the

[8] Warranted in the context of trust means that they rely on "good evidence" or that they "successfully target a trustworthy person" (McLeod 2022, para 2).

[9] There is a big debate in the literature on whether trust amounts to a belief or an emotion, given the similarity of trust to both beliefs (Keren 2020) and emotions (Lahno 2020). In my view, the debate is wrongly focusing on the cognitive or affective nature of trust. Rather, the focus should be on the sources of the belief that one is trustworthy; these sources can be both cognitive and affective. As Jones (1996, 5-12) explains, emotions are not themselves beliefs, but they can generate beliefs, or they can make certain evidence look more convincing. Trust is an affective loop: if I trust you, I will find reasons to continue to trust you and vice versa (Jones 2019, 396). Empirical studies confirm the affective dimension of trust (*e.g.*, Fehr *et al*. 2005; Bohnet & Zeckhauser 2004; Kosfeld *et al*. 2005).

first situation will not do so in the second, when the stakes are much higher. The amount of time we need to perform trustworthiness assessments will typically depend on the context and will be directly proportional to the stakes of the trust situation.

Conclusions

This paper provides a comprehensive and nuanced understanding of trustworthiness, grounded in three fundamental criteria: competence, predictability, and responsiveness. These criteria serve as necessary and sufficient conditions for trustworthiness assessments in any situations that involve trusting others. These are not, however, fixed characteristics of the trustee. The trustor's subjective assessment of each criterion will differ depending on the specifics of the trust situation and the context.

The three criteria build on previous research and integrate the competing intuitions about trustworthiness one will find in the literature. The aim is to establish a common, integrated framework for trust research, thereby addressing the conceptual and theoretical fragmentation that has long challenged this field. This approach also offers a framework to investigate trust in different settings, from interpersonal relationships to organizational and institutional settings. While the three criteria are broad enough to accommodate a range of trust-related information, they are also sufficiently distinct to differentiate trustworthiness from related concepts like reliability. This balance allows for a multifaceted exploration of trust, enhancing our ability to understand and compare trust dynamics across various cases.

References

Adler, J. (1994). "Testimony, Trust, Knowing." In *The Journal of Philosophy* 91(5): 264-275.

Alvarez, M. (2017). "Reasons for Action: Justification, Motivation, Explanation." In *Stanford Encyclopedia of Philosophy*, available online at <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/.> last time accessed at 3rd March 2023.

Baier, A. (1986). "Trust and Antitrust." In *Ethics* 96: 231-260.

Becker, L. (1996). "Trust as Noncognitive Security about Motives." In *Ethics* 107: 43-61.

Berg, J., *et al*. (1995). "Trust, Reciprocity, and Social History." In *Games and Economic Behavior* 10: 122-142.

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

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

Coleman, J. (1990). *Foundations of Social Theory.* Cambridge (MA), London: The Belknap Press of Harvard University Press.

Cook, K., J. Santana (2018). "Trust and Rational Choice." In *The Oxford Handbook of Social and Political Trust*, edited by E. Uslaner. New York: Oxford University Press.

Dasgupta, P. (1988). "Trust as a Commodity." In *Trust: Making and Breaking Cooperative Relations*, edited by D. Gambetta. Oxford, Cambridge: Basil Blackwell, 51-72.

Dunn, J., M.E. Schweitzer (2005). "Feeling and Believing: The Influence of Emotion on Trust." In *Journal of Personality and Social Psychology* 88(5): 736-748.

Everett, J.A.C., M.J. Crockett, D.A. Pizzaro (2016). "Inference of Trustworthiness from Intuitive Moral Judgments." In *Journal of Experimental Psychology* 145(6): 772-787.

Faulkner, P. (2014). "The practical rationality of trust." In *Synthese* 191(9): 1975-1989.

Fehr, E., U. Fischbacher, M. Kosfeld (2005). "Neuroeconomic Foundations of Trust and Social Preferences." In *IZA Discussion Papers*, No. 1641. Bonn: Institute for the Study of Labor (IZA).

FeldmanHall, O., *et al*. (2018). "Stimulus generalization as a mechanism for learning to trust." In *Proceedings of the National Academy of Sciences* 115(7): E1690-E1697.

Fricker, E. (2006). "Second-Hand Knowledge." In *Philosophy and Phenomenological Research* 73(3): 592-618.

Gambetta, D. (1988). "Can We Trust *Trust*?" In *Trust: Making and Breaking Cooperative Relations*, edited by D. Gambetta. Oxford, Cambridge: Basil Blackwell, 213-37.

Gordon, R.M. (1896). "Folk Psychology as Simulation." In *Mind and Language* 1(2): 158-171.

Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Hausman, D.M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.

Hawley, K. (2014). "Trust, Distrust and Commitment." In *Noûs* 48(1): 1-20.

Heal, J. (1994). "Simulation vs Theory-Theory: What is at Issue?" In *Objectivity, Simulation, and the Unity of Consciousness: Current Issues in the Philosophy of Mind*, edited by Ch. Peacocke. Oxford: Oxford University Press, 129-144.

Hertzberg, L. (1988). "On the attitude of trust." *Inquiry: An Interdisciplinary Journal of Philosophy* 31(3): 307-322.

Hieronymi, P. (2008). "The Reasons of Trust." In *Australasian Journal of Philosophy* 86(2): 213-236.

Hollis, M. (1998). *Trust Within Reason*. Cambridge: Cambridge University Press.

Holton, R. (1994). "Deciding to Trust, Coming to Believe." In *Australasian Journal of Philosophy* 72(1): 63-76.

Hume, D.A. (1986) [1739-40]. *Treatise of Human Nature*, edited by L.A. Selby-Bigge. London: Oxford University Press.

Jones, K. (1996). "Trust as an Affective Attitude." In *Ethics* 107(1): 4-25.

Jones, K. (2012). "Trustworthiness." In *Ethics* 123(1): 61-85.

Jones, K. (2019). "Trust, distrust, and affective looping." In *Philosophical Studies* 176(4): 955-968.

Kahneman, D., and A. Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk." In *Econometrica* 47(2): 263-291.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss and Giroux.

Keren, A. (2020). "Trust and Belief." In *The Routledge Handbook of Trust and Philosophy*, edited by J. Simon. New York, London: Routledge, 109-120.

Knack, S., Ph. Keefer (1997). "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." In *The Quarterly Journal of Economics* 112 (4): 1251-1288.

Kosfeld, M. *et al*. (2005). "Oxytocin increases trust in humans." In *Nature* 435: 673-676.

Lahno, B. (2001). "On the Emotional Character of Trust." In *Ethical Theory and Moral Practice* 4: 171-189.

Lahno, B. (2020). "Trust and Emotion." In *The Routledge Handbook of Trust and Philosophy*, edited by J. Simon, 147-159. New York, London: Routledge.

Lewis, D. (1970). "How to Define Theoretical Terms." In *The Journal of Philosophy* 67(13): 427-446.

Lewis, D. (1972). "Psychophysical and Theoretical Identification." In *Australasian Journal of Philosophy* 8(4): 249-274.

Maister, D.H., *et al*. (2000). *The Trusted Advisor.* New York, London, Toronto, Sydney: The Free Press.

Mayer, R.C., J.H. Davis, and F.D. Schoorman (1995). "An Integrative Model of Organizational Trust." In *The Academy of Management Review* 20(3): 709-734.

McLeod, C. (2022). "Trust." In *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/archives/fall2021/entries/trust/.

Parfit, D. (1997). "Reason and Motivation." In *Proceedings of the Aristotelian Society*, Vol. 71, 99-130.

Pettit, Ph. (2002). *Rules, Reasons, and Norms*. Oxford: Clarendon Press.

Putnam, R.D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Sellars, W. (1956). *Empiricism and the Philosophy of Mind.* Vol. I, in *Minnesota Studies in the Philosophy of Science*, edited by H. Feigl and M. Scriven. Minneapolis, MN: University of Minnesota Press, 253-329.

Simon, H.A. (1997). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 4th edition. The Free Press.

Sober, E., Wilson, D.S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge (MA), London: Harvard University Press.

Todorov, A., *et al*. (2009). "Evaluating Faces on Trustworthiness After Minimal Time Exposure." In *Social Cognition* 27(6): 813-833.

Todorov, A., *et al*. (2008). "Evaluating face trustworthiness: a model-based approach." In *Social Cognitive and Affective Neuroscience* 3: 119-127.

Uddenburg, S., *et al*. (2020). "A face you can trust: Iterated learning reveals how stereotypes of facial trustworthiness may propagate in the absence of evidence." In *Journal of Vision* 20(11): 1735.

Williams, B. (2002). *Truth and Truthfulness*. Princeton, Oxford: Princeton University Press.

Yamagishi, T. (1998). *The Structure of Trust: An Evolutionary Game of Mind and Society*. Tokyo: Tokyo University Press.

Zak, P.J., S. Knack. (2001). "Trust and Growth." In *The Economic Journal* 111: 295-321.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

DECLARATION OF CONFLICTING INTERESTS
The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING
The author received no financial support for the research, authorship, and/or publication of this article.