On Defining "Reliance" and "Trust": Purposes, Conditions of Adequacy, and New Definitions

Downloaded from: https://research.chalmers.se, 2026-10-17 03:15 UTC

Citation for the original published paper (version of record):
de Fine Licht, K., Brülde, B. (2021). On Defining "Reliance" and "Trust": Purposes, Conditions of Adequacy, and New Definitions. Philosophia, 49(5): 1981-2001.
http://dx.doi.org/10.1007/s11406-021-00339-1

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Philosophia
https://doi.org/10.1007/s11406-021-00339-1

On Defining "Reliance" and "Trust": Purposes, Conditions of Adequacy, and New Definitions

Karl de Fine Licht[1] · **Bengt Brülde**[2]

Received: 19 June 2019 / Revised: 21 January 2021 / Accepted: 24 January 2021
© The Author(s) 2021

Abstract
Trust is often perceived as having great value. For example, there is a strong belief that trust will bring different sorts of public goods and help us preserve common resources. A related concept which is just as important, but perhaps not explicitly discussed to the same extent as "trust", is "reliance" or "confidence". To be able to rely on some agent is often seen as a prerequisite for being able to trust this agent. Up to now, the conceptual discussion about the definition of trust and reliance has been rational in the sense that most people involved have offered arguments for their respective views, or against competing views. While these arguments rely on some criterion or other, these criteria are rarely explicitly stated, and to our knowledge, no systematic account of such criteria has been offered. In this paper we give an account of what criteria we should use to assess tentative definitions of "trust" and "reliance". We will also offer our own well-founded definitions of "trust" and "reliance". Trust should be regarded as a kind of reliance and we defend what we call "the accountability view" of trust, by appealing to the desiderata we identify in the first parts of the paper.

Keywords Trust · Two-place trust · Three-place trust · Reliance · The accountability view on trust

✉ Karl de Fine Licht
karl.definelicht@chalmers.se

[1] Technology, Managment and Economics, Division Science, Technology and Society, Chalmers University of Technology, Gothenburg, Sweden

[2] The Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg, Gothenburg, Sweden

Published online: 25 March 2021

1 Introduction

Trust and reliance (confidence) is often perceived as having great value when invested properly.[1] For example, trusting ourselves tends to make our lives better, trusting our close friends and partners makes our relationships better, and trusting people in general makes it easier to fulfill our goals in life, at least when our trust is well-founded.[2] Similarly, a population of high trusters also make our economies work more smoothly, generating larger profits, and it is necessary for solving the new and potentially devastating evils of climate change and antibiotic resistance.[3] Reliance is often seen as a prerequisite for trust (e.g. Baier 1986, Holton 1994, Jones 1996) and is often used by, for instance, political scientists when measuring the perceived quality of government among the general public.[4] Thus, trust and reliance are central for our lives to go well and this is reflected in the vast amount of research being made during the last two decades.[5]

However, even though these claims about trust and reliance are frequently made, there is a great controversy about how to define these concepts (e.g. Baier 1986; Hardin 2002; Nickel 2017), and there is no clearly articulated account of what makes such a definition a good one. For example, some people defending a certain definition of "trust" appeal to how we use the term in ordinary language (e.g. Holton 1994), whereas others attacking a definition of trust appeal to its negative methodological implications, e.g., that it gives us a more complex theory without explaining more actions (e.g. Hardin 2002, 2006). While these discussions have given us many great insights, no systematic account of criteria of adequancy has been offered, and with no such account we will have no clear way of deciding the validity of the definition in a clear and transparent way.[6]

[1] For an empirical discussion about these issues, see e.g. Rotter 1967, Luhmann 1979, Yamagishi and Yamagishi 1994, Fukuyama 1995, Hardin 2002, Uslaner 2002, Rothstein 2005, and Colquitt et al. 2007.

[2] Trustworthiness is also perceived as valuable (e.g. Hardin 2002, Colquitt et al. 2007, and Jordan et al. 2016). For example, Jordan et al. (2016) argues that people sometimes punish each other to appear more trustworthy themselves, even though the act of punishing has a large cost for them personally. However, trust is not just based on judging someone as trustworthy (e.g. Jones 2012). For example, we might not trust foreign doctors as much as domestic even though we judge them just as trustworthy.

[3] Trust is also by some people perceived as having final value (e.g. Pettit 1995, Lahno 2001).

[4] For example, when the Swedish Society Opinion Media institute (SOM institute) asks questions related to trust these are most often formulated in terms of "reliance" or "confidence". Question 8 in the questionnaire reads "How much confidence do you have in the way the following institutions and groups are doing their work?", and question 11 reads "Generally speaking, how big is your confidence in Swedish politicians?". Only two questions are phrased explicitly in terms of trust (in question 27): "To what extent, according to you, can we trust people in general?" and "To what extent do you think people in your neighborhood can be trusted?" The questions can be found in Swedish here: http://som.gu.se/digitalAssets/1581/1581065_riks-1-2015.pdf).

[5] For references, e.g. footnotes 1 and 2.

[6] Of course, there are some exceptions. For example, Jones (2012: 62–66) argues that "trust" and "trustworthiness" should be defined in such a way that makes them play a certain functional role. In a similar way, Nickel (2017: 197–202) argues that a definition of "trust" can only be acceptable if it gives trust a certain explanatory role, and Hardin (2002: 1, 7f, 11, 54, 87f) agrees with Nickel. However, no complete list of the conditions of adequacy is offered, even if e.g. Nickel insists that there are more conditions than explanatory power.

The purpose of with this paper is to give an account of what criteria we should use to assess tentative definitions of "trust" and "reliance" ("confidence") and then show what we belive follows in terms of definitions of these concepts when using this account. In order to know what we want from a definition in the first place we need to identify the possible purpose(s) of defining "trust" and "reliance". There are both theoretical and practical purposes where the most important one is theoretical (explanatory) while the other is practical (normative). To fulfill these theoretical and practical purposes we need at least eight conditions of adequacy ranging from our definitions being compatible with ordinary language and the values we ordinarily believe the definitions are imbued with, to be simple and precise enough so that we can apply these definitions in practice. Using these eight conditions we will argue in favor of a "non-restrictive" view on "reliance" and a "restrictive view" ("a moralized account") on trust, where it is argued that "trust" should be perceived as a kind of reliance.

The structure of the paper is as follows. First, we discuss the purposes and aims of defining "trust" and "reliance" and what conditions of adequacy that follows from these purposes and aims. Then, we argue in favor of a non-restrictive view on "reliance" and in favor of a restrictive view on "trust" called the "accountability view". Last, we discuss how our view could answer different possible objections before we conclude the paper with a few reflections on how our framework can be used in the future.

2 Purposes and Desiderata for a Good Definition of Trust and Reliance

In this section, we first identify the aims of defining "trust" and "reliance" in the first place. We then use these purposes to identify the conditions of adequacy that are most in line with these aims.

2.1 Purposes and Aims

There are both theoretical and practical purposes for wanting well-founded definitions of "trust" and "reliance", and for distinguishing between trust and reliance. In our view, the most important purpose is theoretical (explanatory), since the notions of trust, distrust and reliance sometimes play an important role in the explanations we give, both in the social sciences and in everyday life.

For example, it seems that a mature and reasonably complete explanatory theory about "trust" and "reliance" should aim to explain (1) cooperation, (2) failure in cooperation, and (3) the emergence of cooperation-enabling institutions,[7] since an agent's willingness to cooperate is (in part) dependent on such factors. But to determine how important these factors are for cooperation; we

[7] Nickel 2017, p. 197.

also need to be able to measure different kinds of trust and reliance. To have clear definitions of trust and reliance increase the explanatory power of the theories which include these concepts, and makes these theories serve their purposes better.[8]

There are also practical purposes for wanting a well-founded definition of trust. For example, many of us believe that (justified) trust is a good thing that ought to be promoted, and that this holds for interpersonal trust as well as trusting our societal institutions. However, it is easier to know what exactly we ought to promote if the concept has been clearly defined, i.e. a definition has at least some practical importance.

It seems reasonable to assume that there is a difference in value between "trust" and "reliance",[9] and that this gives us a prima facie reason to distinguish between "trust" and "mere reliance". For example, while reliance might help us solve coordination problems just as well as trust, trust might be more conducive to respect. So, when Coleman and Coleman argue that "trust" should be defined in terms of having a need for or interest in cooperation, this definition might be criticized for not taking into account everything we value about "trust".[10] Another reason for distinguishing between trust and reliance is that some forms of reliance seem compatible with the use of force, which does not hold for trust. Moreover, a breach of trust might spur different reactions than a breach of reliance, e.g., blame rather than disappointment, which suggests that it might also have explanatory value to separate the two.

But again, the main purpose for distinguishing between trust and reliance is probably evaluative (i.e. practical), even though we think that they are both good things that ought to be promoted, e.g., because they help us solve coordination problems. We seem to value rather disparate things with regard to trust and reliance, and that the distinction between reliance and trust proper can help us become more aware of these differences, as well as the more practical issue of when we ought to promote trust rather than reliance, and when we ought to promote other forms of reliance rather than trust. (There are also evaluative reasons for distinguishing between different forms of trust and reliance, e.g. between "genuine trust" and "therapeutic trust", trust that is spurred on by a hope on creating trustworthiness.)

How and why we value trust and reliance is also related to the theoretical purpose outlined above. The idea that we value trust and reliance for different reasons can in part explain why people react differently when trust and reliance are breached, and that trustees might be subjected to pressures that would not otherwise be an issue. For example, physicians might subscribe antibiotics more liberally when patients trust them instead of relying on them in some other way.

[8] Of course, this is not to say that cooperation is the only context in which reliance and trust are important, but it might be the most important contexts.

[9] Pettit 1995, Lahno 2001.

[10] Coleman and Coleman 1994.

2.2 Eight Desiderata

There seems to be at least eight desiderata, or conditions of adequacy, that a good definition of "trust" and "reliance" should ideally satisfy.[11] It is likely that some readers want to exclude some of these desiderata right away (e.g. the ordinary language condition), but if this happens, it is most probably because they rely on some other desideratum, explicitly or implicitly. It is worth noting that all the requirements can perhaps not be met at the same time. It is also worth noting that desiderata 7 and 8 are closely related to the purposes listed above, whereas conditions 1–6 are independent of these purposes.

The Ordinary Language Condition There are two aspects of this condition. Firstly, it is always desirable that a term is defined in a way that is *consistent with* how these terms are ordinarily used by the relevant groups, and secondly, a definition should (ideally) *explain* our intuitive linguistic judgments of how the term is used.

The idea that a definition should be consistent with ordinary usage has many facets when it comes to "reliance" and "trust". Firstly, a definition should include those mental states (attitudes etc.) we intuitively regard as cases of trust, and it should exclude related phenomena that we intuitively regard as non-trust, like certain types of reliance. Secondly, a definition should be consistent with the fact that attributions of trust and distrust are sometimes invoked to (partially) explain people's behavior, e.g. their willingness to cooperate. Thirdly, a definition of "trust" should be consistent with the fact that we tend to regard the category of trust as a sub-category of the more general category of reliance. It is also desirable, however, that a definition of "trust" should help us draw a line between trust and other forms of reliance.

The idea that a definition should *explain* why we think and talk about trust and reliance the way we do also has several facets. For example, a definition should (ideally) explain why almost all of us tend to agree about some cases, whereas we tend to disagree about other cases. A definition should also explain our comparative judgements about trust, e.g. that I am more trusting now than I was two years ago, or that one agent is more trusting than another.

It should be noted that for a definition to be in accordance with the ordinary language condition, it must allow for the possibility of trusting institutions and people in general. Trust in the government and our fellow human beings is quite entrenched in how we talk and experience the world, and it does not just figure in our explanations of why collective action problems occur and how they might be solved,[12] but also in how we explain why certain individuals have fewer friends or do not shop online.[13]

If we denounce a notion like "generalized trust," like some want to do, we can either try to show that it is not applicable to our purposes, e.g. since it lacks

[11] Cf. e.g. Brülde 1998, 2000, 2007 with regard to "health" and "mental disorder".

[12] E.g., Putnam 1993, Uslaner 2002, Rothstein 2005.

[13] E.g., Putnam 1995, Lahno 2004.

explanatory power, or that it really refers to something else, e.g., reliance.[14] Unless this is shown, it is desirable that a definition of trust is compatible with the idea of generalized trust and trust in large institutions.[15]

A definition of trust should also be able to explain the particular reactions of blame when our trust is breached, as opposed to the disappointment we tend to feel if we relied on someone or something. This is something that is often forgotten when explanatory power is discussed even though it is hugely important for policymakers and others to know how we might react to different breaches.

The ordinary language condition is of course not the only criterion, since our everyday thinking might be conceptually confused, and to the extent that this is so, we need to *explicate* the relevant terms, i.e. provide *stipulations* that are in line with ordinary language. The reason why we need this can be explained by a second criterion, namely the precision condition.

The Precision Condition A definition should (ideally) be sufficiently clear and precise so that there is*, in principle*, little or no doubt whether a certain thing is denoted by the term. The focus is normally on the conceptual boundaries of the concept (the categorical approach), but when the term denotes a variable that admits of different degrees one might also adopt a dimensional approach.

Since e.g. trust is a mental state that comes in degrees, we can use both these approaches. If we conceive of trust as a category, i.e. if we focus on a certain level of trust, a definition should draw a sharp conceptual boundary between trust and related phenomena, like other kinds of reliance. If we adopt a "dimensional approach" rather than (or in addition to) a "categorical model", the condition requires that we define "trust" in such a way so that we, for every attitudinal (etc.) change, can determine whether this change constitutes an increase in trust or a reduction in trust. If we adopt the dimensional approach only, it is not necessary to draw a sharp conceptual boundary between trust and non-trust, but we might still need to draw a sharp line between trust and other forms of reliance.

Again, it is worth pointing out that the precision condition should always be considered in combination with the ordinary language condition, i.e. that mere precision is not necessarily a virtue (compare with the critiques against different "New Public Management" regimes).[16] If the two conditions are jointly satisfied, we can talk about "descriptive adequacy",[17] "extensional adequacy",[18] or a successful explication. It is also worth noting that precision condition is strongly connected to the theoretical or explanatory purpose described above, e.g. the desire for a theory that

[14] There is an ongoing discussion about whether trust is a one-, two-, three- or even a five- or a six-part relation (e.g., Uslaner 2002, Castelfranchi and Rino 2010, Baier 2013, Faulkner 2015, Domenicucci and Holton 2016). Many people think that "generalized trust" should be conceived of as a "one part relationship" ("A trusts"). In this study, we have chosen to focus on the three-part relations ("A trusts B to C" or "A relies on B to C"), mainly because these statements seem to be most in line with the theoretical and practical purposes of a definition.

[15] For a defense of the idea that, in general, we cannot trust intuitions, see Hardin 2002. For a thorough critique of this view, see e.g. de Fine Licht 2017.

[16] E.g. Power 1997, Bevan and Hood 2006.

[17] Sumner 1996.

[18] Bradley 2012, p. 394.

tells us to what extent a person's cooperative behavior can be predicted by her level of trust.

The Value Condition If a term we want to define denotes something that is generally regarded as valuable (like health or democracy), a definition should capture this fact. The particular attitudes that we classify as examples of rational and well-founded trust or reliance are typically regarded as desirable, both for the trusting agent and the trustee. A definition of "trust" should not just be consistent with this fact, however, but also with the fact that we sometimes regard trust as a bad thing, e.g. as harmful for the trusting agent, and that it is not always better to trust more, i.e. that there seems to be an optimal level of trust. Moreover, it is also desirable that the different ways in which we value trust and reliance are also reflected in our categories (see above). This condition is strictly speaking a part of the ordinary language condition, but it is worth emphasizing because of its practical importance.

The Reliability Condition A definition should be practically applicable; it should make it relatively easy *in practice* (not just in principle) to determine whether something is denoted by the term or not. This makes it possible for different observers to apply the concept in the same way, e.g. to agree on what should be excluded from or included in the concept. The reliability condition is easier to satisfy if the concept is defined in observational terms, or if it can at least be easily operationalized.

When it comes to "trust" and "reliance" we want to be able to easily determine whether some agent trusts another agent in some respect, or whether some agent is more trusting than another agent. Practical applicability is obviously important for communication purposes, i.e. a definition that satisfies this condition can (if commonly accepted) improve or facilitate communication between different groups and individuals, both across different settings and cultures.[19]

The Measurability Condition Whenever a term denotes something that admits of degrees, a definition should ideally make it possible to make comparisons between different degrees, i.e. it should make the relevant levels measurable in the ordinal (and interpersonal) sense.[20] This also holds for "trust" and "reliance". However, in our view, it is not really possible to find a plausible concept of trust that allows for cardinal measurability, since no attitude or emotion is measurable in such a strong sense (at least not interpersonally). It is important to note that this condition is

[19] Of course, we could operationalize mental states and other phenomena without having a clear definition in the first place and the question then becomes if we need a definition at all. Why not just ask people straight-forwardly about what they trust and to what extent, without providing a definition? This relates to the aims of the purposes of having a definition in the first place, which was discussed in Section 2.1.

[20] The well-renowned professor in political science, Uslaner 2009, argues in favor of dichotomous measures against 5, 7, or 11-point scales, so this is a quite common and accepted stance in that field even though it has become more common to measure trust on more finely grained scales. According to Colquitt 2007 the three most common ways of measuring trust is (1) asking people directly, (2) asking them indirectly about their attitude to others, and (3) asking them how they feel about taking risks, being vulnerable before others, and their job performance Colquitt et al. 2007, p. 910. Interestingly, these different ways of questioning did not vary in their results Colquitt et al. 2007: 920.

strongly connected to the theoretical or explanatory purpose described above, e.g. the desire for a theory that tells us to what extent a person's cooperative behavior can be predicted by her level of different forms of reliance such as trust.[21]

The Simplicity Conditions A definition should ideally be as simple as possible, e.g., to be easy to use in practice. For this to be possible, at least two things are needed. Firstly, to identify a conceptual core of the concept instead of merely providing a list of ad hoc features.[22] Secondly, to use one criterion only if possible, or a conjunction of different criteria, instead of using a disjunction of different criteria. (This is a desire for monism or parsimony.) This condition can be used in conjunction with the condition of explanatory power (see below). If we do this, we get an idea that it is very similar to Occam's razor, namely that if we have two definitions with the same explanatory power that differ in simplicity, we should, ceteris paribus, choose the simpler one.

These six desiderata are all conditions of adequacy in the strict sense, i.e. conditions that are (in part) derived from certain purposes, but which do not themselves constitute such purposes. Note also that most of the desiderata above are standard for most purposes. The last two conditions explicitly appeal to the two purposes of a definition mentioned above, and they imply that a definition can be criticized for not fulfilling these purposes.

The Condition of Normative Adequacy This is the idea that a definition of "trust" or "reliance" should help us make better decisions, or create better institutions in different areas, e.g. by promoting trust. This condition is strongly related to the practical purpose of a definition.

The Condition of Explanatory Power A definition has (strictly speaking) no explanatory power in itself, but how a concept is defined can, as we have seen, affect the explanatory power of the theories which contain this concept. Here, the idea is that if we have two definitions and one "explains" more than the other, we should, ceteris paribus, choose the one that explains more.[23] For example, if we have no restrictions on motive when trying to explain cooperative behaviour in terms of trust, it might be easier to satisfy this condition than if we have such restrictions.[24] This condition is strongly related to the theoretical or explanatory purpose of a definition. Again, this is to suggest that cooperation is a necessary condition for "trust" or "reliance", or that cooperation is the only thing they need to explain. The point is being instead that having a definition of "trust" and "reliance" which explains why people more easily can cooperate when they trust or rely on each other is to be preferred over defintions which cannot give such an explanation.

[21] Again, it is worth noting that other actions besides cooperation can be (in part) explained in terms of trust and reliance.

[22] This is a desire for conceptual unity; cf. Bradley 2012, p 395, in the case of "harm".

[23] Cf. Hardin 2002, 2006, Nickel 2017.

[24] Nickel 2017, pp. 197—203. This is one reason why we do not discuss the ontological question regarding the nature of trust, e.g. whether trust is ultimately cognitive or emotional: it makes no difference from an explanatory perspective.

This concludes our list of tentative desiderata. However, it is far from certain that there is one single definition that satisfies all these conditions. If some of these conditions happen to pull in different directions, we simply must take a stand on which desiderata are most important, and if possible, provide arguments for our choice.

Before we provide our own definitions of "trust" and "reliance", let us first note that these terms have both contrary opposites – e.g. distrust – as well as contradictory opposites, e.g. non-trust. Distrust is not a mere absence of trust or reliance,[25] and there is also a kind of "non-trust" where the potential trustor neither trusts nor distrusts the potential trustee.[26] However, it is an open question whether we can talk about reliance, "dis-reliance", and non-reliance in a similar way.

3 Our Own Definitions of "Reliance" and "Trust"

Reliance has been called the neutral core of trust.[27] We agree with this, and in this section, we argue that "trust" should be regarded as a special kind of "reliance" ("confidence", or "assurance"), namely a "moralized" kind of reliance.

3.1 Reliance

There is a controversy about how to define "trust", but how "reliance" should be defined is not as controversial. We conceive of reliance as a three-place relation, where one agent (A) relies on an agent or some other object (B) to do something, to maintain some state, or the like (C).[28] For example, I (A) rely on the government (B) to support me if I am in need (C), I (A) rely on my rope (B) to stay safe (C) when I am climbing a mountain, the FBI (A) relies on the local police force (B) to give them all the information they need (C), or I (A) rely on people in general (B) to do their duty (C).

One might believe that reliance on B to C consists in believing that B will actually make C happen. This is too strong, however, since B might not be in total control of C. For this reason, we believe that the core of "A relies on B to C" is that A judges B to make C, ceteris paribus, more probable (in many cases highly probable). Thus, A relies on B to C when,

[25] Hawley 2014.

[26] For example, according to Hawley, A trusts B to C when he relies on B to C and believes that B is committed to C. For us to distrust someone, Hawley says, "is to believe that she has a commitment to doing it, and yet not rely upon her to meet that commitment" Hawley 2014, p. 10. This suggests that distrust is not the mere absence of trust or reliance. Moreover, if I distrust someone whom I believe afterwards I should have trusted, I feel remorse, apologize, and ask for forgiveness Hawley 2014, p. 3.

[27] Blackburn 1998: 32.

[28] E.g. Baier 1986, Hawley 2014, Domenicucci and Holton 2016, p. 5.

(i) A judges B to increase the probability for C (the probability condition)[29]

According to this view on reliance, we can rely on both inanimate objects (like ropes) and agents, but here we are only interested in the latter. The most plausible (and widely held) conditions of "agential reliance" are competence, motivation, and opportunity. Thus, A relies on some individual or collective agent B to perform an act or ensure a state C, if A believes that:

(ii) B is able to do (achieve, ensure, etc.) C (the competence condition)
(iii) B is motivated to do (ensure, etc.) C (the motivation condition)
(iv) B has a reasonable opportunity to do (etc.) C (the opportunity condition).[30]

Conditions (ii)-(iv) are closely related to (i), and they can even be empirically derived from (i). Moreover, when an agent has the right competence and have the opportunity to C, that agent's motivational structure is typically perceived as a significant explanation of C.[31]

It is worth noting that reliance can be both voluntary and involuntary. That is, A can either *choose* to rely on B to C or "come to realize" (even against her own will) that B can be relied upon to C. There is some disagreement about whether we can choose to trust or whether we just happen to trust regardless of what we decide to do,[32] but in the case of reliance, it seems that that A can choose to rely on B to C if there are no compelling reason not to do so.[33]

Now, when A chooses to rely on B to C it is reasonable to assume that A regards C as valuable, but A can also regard the very relying on B as valuable.[34] For example, I might rely on my partner to make plans for the holidays because that will make the holiday more successful, but I might also think that the mere relying on her would be good for our relationship. However, when A relies on B to C willy-nilly, A does not necessarily attribute positive value to C. For example, we can rely on someone "behaving like a pig" whenever he drinks, which is of negative value. But in most cases, we can safely assume that when A relies on B to C, then.

(xxii) A values C positively, and when A chooses to rely on B to C, A might also value the vary act of reliance (the value condition).

[29] This suggests that "reliance" ("confidence") refers to the same thing as "trust" in a non-restrictive, rational, or predictive sense (e.g. Luhmann 1979, Gambetta 1988, Coleman and Coleman 1994, Hardin 2002, pp. 1—10, Levi and Stocker 2000, p. 476, Delhey and Newton 2005, p. 311, Colquitt et al. 2007).

[30] Since we conceive of trust as a kind of agential reliance, these criteria are of course necessary for trust as well (cf. McLeod 2015).

[31] Cf. e.g. Björnsson and Persson 2012, 2013, Björnsson 2011, 2014.

[32] E.g. Holton 1994, Hardin 2002, Faulkner 2015.

[33] It might seem counterintuitive to say that we rely on someone when the chances of them coming through are really slim. However, it seems to us that we can choose to rely on someone else just because there is no one else to rely on and the values at stake are great. For example, I might rely on a Nazi guard who dislikes me to smuggle me out from the camp. In this case I might deliberately focus on the successful chains of events leading up to my freedom. If I am unable to do this, I start to second guess my judgment of reliance. The reason for my deliberate focus on success here is, of course, that the alternative is to horrible to imagine.

[34] Cf. Pettit 1995, and Lahno 2001, in the case of trust.

This definition of "reliance" satisfies the ordinary language condition, and it is sufficiently simple and clear to satisfy the simplicity conditions as well. The precision condition and the reliability condition are satisfied to the extent that key notions like competence, motivation, and opportunity are precise (and can be reliably applied), something that we judge to be the case. We also think (partly for this reason) that the definition also satisfies the criterion of explanatory power.

It is also worth noting that we often rely on agents *not to interfere* when we are trying to achieve our goals. Just think about driving a car in ordinary traffic, where we constantly rely on others not breaking the rules. In these cases, our reliance is rarely the result of a conscious choice.

Our definition of "reliance" can also easily account for cases of non-reliance, the contradictory opposite of reliance. According to our view, there are several possible reasons why A might not rely on B to C, e.g. because he thinks that B lacks the competence, the motivation, or the opportunity to C. If A believes that B is motivated to not-C, we can talk about dis-reliance or distrust proper (the contrary opposite). For example, a common view among African Americans in the United States is that you should try to keep out of the way of government officials, such as the police, because they want to make your life worse and has the ability and opportunity to do so. This case of ill will should be distinguished from the case where the motivation to do good is simply lacking. The distinction between non-reliance and dis-reliance is quite clear in the cases where we choose to rely on agents or objects because we think we need to.[35] If we believed that we could climb the mountain safely and with no problem without the rope, then we would not rely on it, but this does not mean that we distrust it.

3.2 Trust

How we should define "trust" is a more controversial issue than how we should define "reliance". Our own "accountability view" consists of two steps. We first propose that trust is a special form of agential reliance, i.e. that conditions (i)-(iv) above are also necessary for trust. We then suggest that the difference between trust and other forms of reliance is that a trusting agent attributes a certain kind of moral motivation to the trustee, namely a perceived normative responsibility to care about something. Moreover, the judgement that the other perceives herself as responsible is often accompanied by a judgement that she is in fact responsible, which neatly explains why we tend to hold the trustee accountable (e.g., blame him or her) if our trust is breached. Thus, A trusts B to C if and only if

(1) A relies on B to C, i.e. A believes that B is likely to C because B is competent, motivated, has the opportunity to C, *and*
(2) A judges that B has a belief that B has a moral obligation to do C, *or* a forward-looking normative responsibility to care for C. This is the reason why A relies on B to C.

[35] Cf. Holton 1994, p. 67, Blackburn 1998, p. 32.

Normative responsibilities are here understood as requirements to care about the object of responsibility (in this case C), and the requirement to care gives rise to moral requirements to act in a way that caring would yield under normal circumstances.[36] We tend to view the requirement to care as a moral requirement, but it is worth noting that there are several possible sources of normative responsibilities, such as capacity and retrospective causal responsibility, as well promises and agreements.

Now, that A attributes a certain moral belief to B – that B perceives himself as normatively responsible for C – is not the same as attributing responsibility to B, e.g., I can regard B as responsible for C even if B herself fails to realize this, and I can attribute a perceived normative responsibility to B even if I do not regard her as responsible. But in most cases of trust, A does not just believe that B *perceives* himself as responsible for C, she also judges B to be normative responsible for C. This is not part of our definition, but it is still worth emphasizing that in most cases of trust, A herself tend to hold the belief that she attributes to B in (2). In these cases,

(3) A judges that B has a moral obligation to do C, or that B is normatively responsible to care for C.

Thus, when trusting B to C, A typically attributes normative responsibility to B for C. This makes B a central part of the significant and straightforward explanation of C.[37] When I trust the baby-sitter to care for my son, I do not just attribute a perceived responsibility to her, I also regard her as responsible for him in the sense that she should look out for his well-being by keeping him safe, fed, and happy. In this case, her responsibility to care for my son serves as a straightforward, significant, and normal explanation of why he will stay safe.[38]

Moreover, when I trust the baby-sitter, I also believe that she is going to live up to the responsibilities I attribute to her, and I tend to blame her if she lets me down. That is, when B does not do C (or does not care for C), then B is prima facie blameworthy for not C-ing, but only if it was in B's power to C (if the voluntariness condition of moral responsibility is satisfied) and if B knew (or should have known) that he was normatively responsible for C (if the epistemic condition of moral responsibility is also satisfied). This explains why we tend to blame people who breach our trust.

[36] Björnsson and Brülde 201, pp. 18–21.

[37] It should be noted that this idea about explanation is compatible with many of the common views on reliance, such as Holton's idea that A relies on B to C when A plans for C on the assumption that B will do C for the reason that B is motivated to do C, and where B does not just happen to do C (Holton 1994, p. 66), or Barber's view that reliance (which Barber calls trust) is based on perceived regularities (Barber 1983).

[38] That we think B will C is an explanation of C is in line with the widely held view that A needs to be vulnerable towards B but also optimistic that B will C (for references, see Colquitt et al. 2007, p. 909). We are transferring responsibility, which makes us vulnerable, and we assume that the agent to which we transfer responsibility will make C happen.

Many philosophers in the field seem to assume that the natural reaction to a breach of reliance is disappointment,[39] whereas the natural response to a breach of trust is blame.[40] Since we tend to blame people, ceteris paribus, when they do not do fulfill their responsibilities, we have a straightforward explanation of why betrayed trust often leads to blame. When B betrays the trust of A, B does not fulfill her normative responsibilities; and hence she deserves blame, ceteris paribus. As we will see, this is a major strength of our account: it can easily explain the perceived differences between trust and other forms of reliance, regardless of whether these attitudes are justified or not.

There are several reasons to prefer our account of trust to its competitors. Firstly, it should be preferred to the non-restrictive views on "trust", i.e. the views that do not clearly distinguish trust from reliance in general (as we have defined it). For example, it seems that we can rely on people who we threat or force to do something, but to say that we trust them is in conflict with ordinary language. Our account can also explain why we do not trust these people. When we force someone to do C against their will, we do not think that they perceive themselves as responsible for C. This is why we do not trust them to C. Our definition does not only capture the ordinary usage of the word "trust" better than the non-restrictive views, it also captures how and why we regard trust as valuable, and that we tend to regard it as more valuable than other forms of reliance, at least in some contexts.

It is true that less restrictive accounts of trust will often satisfy e.g. the precision condition, the reliability condition and the measurability condition, to a higher degree than more restrictive accounts, such as our own.[41] However, this is not a problem for our view, since we have access to a concept of reliance that shares the same features as non-restrictive trust. It is worth noting that politicians and social scientists are probably (and perhaps rightly[42]) more interested in reliance than trust, and our definition of reliance can explain why this is, e.g., why it is primarily reliance that is needed to solve different kinds of coordination problems and make the market function smoothly. What we need is really a conceptual scheme where trust is distinguished from other forms of reliance, and where different forms of reliance are given somewhat different roles. For example, it may well be the case that reliance has more explanatory power, whereas trust may well be more normatively relevant.

The most severe competitors to our account are not the non-restrictive views, however, but the "good will" accounts, on the one hand, and other moral obligation accounts, on the other. What makes our account preferable to these two accounts?

Good will accounts define trust (roughly) as reliance with good will as motivation: A trusts B to C if A thinks B makes C more likely, and this reliance is in

[39] Baier 1986, Holton 1994, Jones 1996, Potter 2002, Hieronymi 2008, Hawley 2014.

[40] E.g., Baier 1986, pp. 234f, Jones 1996, pp. 14f, Hawley 2014, pp. 1f.

[41] E.g. Hardin 2006, p. 16, Nickel 2017.

[42] See Hawley 2017.

part based on a belief that B has good will towards A, that B likes A and wishes B well. These views capture the trust involved in love or friendship quite well: we often rely on our loved ones because we believe that they have a good will toward us (assuming that the competence and opportunity is there). It also seems that we might sometimes trust strangers or institutions for similar reasons. But our view can also explain this. If I trust B to C because I think B loves me, then (it seems) I rely on B to C even if she does not feel like it, i.e., because she will then be motivated by a sense of responsibility. Another strength of the good will accounts is that they can explain why we sometimes distrust people rather than simply not trust them: I distrust someone if I think he has ill will towards me, at least ceteris paribus.

The main objection to these accounts is that they do not satisfy the ordinary language condition very well, since we sometimes trust people without believing that they have good intentions toward us.[43] It seems that we can even trust someone who is negatively disposed towards us. I can trust the cab driver to drive me to the right location even though I might think that she does not like me. If this is possible – and not just a case of reliance based on perceived self-interest – it is easily explained by our view. I assume that the cab driver cares about doing a good job and that she perceives herself as morally responsible to take me to my location. Furthermore, caring about doing her job well explains in a normal way the actions that take me to the right place, and if she does not act in this way, she deserves blame because of it. This objection is closely related to the objection that we can trust people for other reasons than perceived good will,[44] e.g., because we think they have a good work ethic and are motivated to do a good job. Of course, this also holds for occupations like doctors, police officers, and teachers.

Since our own account is a kind of moral obligation account, we need to ask why it should be preferred to similar accounts. According to one such account, A trusts B to C when (1) A judges B to be committed to C in the sense that B has a strong intention to C and (2) A relies on the fact that B will C.[45] This view is similar to ours if we take the commitment (motivation) that A attributes to B to be of a moral nature, but our idea that A attributes perceived normative responsibility to B (rather than just commitment) can better explain the strong connection between blame and not C-ing (see above).

The moral obligation accounts provided by Philip Nickels and Lars Herzberg do not suffer from this weakness. According to Nickels, A trusts B to C when (1) A ascribes a moral obligation for B (and no other agent) to C, and (2) A in the absence of exonerating circumstances is open to blame B for not

[43] E.g., Holton 1994, pp. 68f, Blackburn 1998, p. 37, Hawley 2014, p. 17.
[44] Blackburn 1998, p. 37.
[45] Hawley 2014, p. 10.

C-ing,[46] whereas Hertzberg argues that A trusts B to do C when A correctly judges B to be prima facie blameworthy for not C-ing.[47]

The main problem with these accounts is that they fail to attribute any moral motivation to B, e.g., a perceived normative responsibility to C, which also means that they fail to regard trust as a special kind of reliance, with a certain motivational content. Another problem is that there are cases of trust where the trusting agent do not attribute normative responsibility (or any other moral obligation) to the trustee. It is true that this is often the case, but the reason for this is that our attributions of responsibility (e.g., to the cab driver above) typically coincide with his own motivation. If the two come apart, e.g., if I regard the cab driver as normatively responsible to C but do not believe that he himself shares this view, then I do not really trust him, even if I would most probably blame him for not fulfilling his responsibilities.

The account provided by Karen Jones is more plausible than the three accounts above. Jones suggests that A trusts B to C when (1) A takes B to care about C, (2) A does not try to secure that C, and (3) A has normative expectations on B to C.[48] This definition captures the strong connection between blame and not C-ing, but as we have already seen, it is not always the case that A attributes a moral obligation to B, i.e. her third condition is too strong. Another point to consider is that Jones explicitly refers to A's beliefs about B's motivation to C, but the reference to "care" makes her first condition too loosely connected to the third condition. Our suggestion does not suffer from this weakness, since we take the trusting agent to believe that the trustee regards herself as responsible for C, which is closely connected to the fact that the trusting agent regards the trustee as responsible for C.

Another advantage of our view is that it is formulated in terms of normative responsibility for C, a notion that is highly context dependent,[49] just like the notion of trust.[50] For example, when we trust the cab driver we do not just trust her to take us from A to B, we also trust her to do this safely and with a certain degree of comfort, to be on time, not to be rude, and so on. The reason why we do this is that we are familiar with the special responsibilities that are attached to the role as taxi driver, and the fact that our expectations are determined by this explains why we tend to blame taxi drivers if they do not arrive in time.

Depending on the situation, different aspects will be more or less relevant, and their importance will vary in strength. Our normative responsibilities are context dependent in the same way, since the object of responsibility is an object of care, and caring can be expressed by a multitude of actions that are ordinarily explained by the agent caring. Our account does not just satisfy the ordinary language condition better than the competing views, but also, e.g., the simplicity condition – because

[46] Nickel 2007, p. 311.
[47] Hertzberg 1988, pp. 319f.
[48] Jones 2004, p. 6.
[49] Björnsson and Brülde 2017, pp. 15–18.
[50] Cf. Faulkner 2015: 424f, 428f, Domenicucci and Holton 2016, pp. 155–157.

we regard trust as a special form of reliance and uses a rather simple criterion to distinguish it from other forms of reliance – and the precision condition, e.g., since we specify the normative expectations that Jones refer to as expectations of care.[51]

Our account has other attractive features, some of which are shared by other moral obligation accounts of trust. For example, they seem to satisfy both the condition of explanatory power and the condition of normative adequacy. As Nickel suggests, these accounts have "explanatory potential insofar as the idea of ascribing and applying norms to others is a fundamental mechanism among humans for changing the balance of reasons and improving compliance with expectations in cooperative exchanges; and also because when these normative expectations fail, we sometimes take other steps to encourage better compliance in the future, such as the establishment of sanctions and institutions".[52]

Another attractive feature of our definition of "trust" is that it is not just compatible with the notion of generalized trust, but that it also gives a plausible analysis of this notion. (As we have already seen, a good definition of "trust" must either make sense of this notion or give a good argument for disregarding it.) We can think of generalized trust both as a three-place relation – with regard to a specific C, like paying taxes, obeying the law, or not stealing my stuff – and as a two-place relation (see below). In the case of two-place generalized trust, we think that people in general are morally decent and responsible agents who care about each other in the way that is morally required. In the case of three-place generalized trust, we think that people in general are motivated to fulfil, e.g., their role responsibilities, and that they have the necessary competence to do this.

Our account also allows for the possibility of trusting collective agents, like states, authorities, or corporations, i.e. the trustee need not be an individual. Again, this is in line with ordinary language. We often talk about trusting institutions ("institutional trust"), and our definition of "trust" can account for this. Firstly, we tend to think that corporations and other structured collectives have a wide range of responsibilities, and that they are blameworthy if they fail to live up to them.[53] For example, when it came to the attention of the public that H&M used subcontractors that exploited its workers, it was blamed for not having acted in a responsible way.[54] But we also tend to view corporations and governments as the sort of entities that can regard themselves as normatively responsible, and if we believe that their perception of their responsibilities coincide with our own, we tend to trust them, but only on assumption that we regard them as competent. (That is, we can distrust a government by virtue of its motivation, e.g. because we think of it as corrupt, or by virtue of its incompetence.).

Lastly?, since trust is, in our view just one form of reliance, depending on the content of the motivation attributed in condition (iii), we will have different forms of reliance. One of the basic forms of reliance is based on beliefs about other people's

[51] Cf. Björnsson and Brülde 2017, which also contains a list of different possible sources of our normative (forward-looking) responsibilities (on pp. 22—30).

[52] Nickel 2017, p. 208.

[53] Cf. Björnsson and Brülde 2017, pp. 21f.

[54] Not everyone agrees that such collective entities are can be regarded as proper agents, and that it is thus not appropriate to blame them. If this is true, collectives cannot really be trusted, only their representatives.

self-interest. For example, if A judges that B will C because B fears A or because B has something to gain by C-ing, then A might rely on B to C if B is competent and has a reasonable opportunity to C. (The definition can also be used to distinguish between different kinds of dis-reliance or non-reliance, e.g. I can either distrust you as a baby-sitter because you do not have the competence or because you do not seem to have the proper motivation.).

Another relevant kind of motivation is that of "good will" or "good intention", which is typically included in "will-based accounts" of trust. On these views, A trusts B to C when A judges B to have a good will toward A.[55] For example, according to Baier, A (genuinely) trusts B to C when A judges B to have a good will toward A, and accepts his own vulnerability in relation to B's C-ing.[56] And according to Jones, A trusts B to C if A judges B to have good intentions toward A, and to be directly and favorably moved by knowing that A counts on B to C.[57] Toshio Yamagishi makes a similar point when he distinguishes between two different kinds of "expectations of intention", namely "assurance of security" – where A relies on B doing C because it is in B's self-interest, e.g. because of fear – and trust, where A expects B (because of the information he has about B) not to act selfishly – not to deceive or exploit or take advantage of A – but to act e.g. ethically.[58]

3.3 Two-place Trust and Three-place Trust, Distrust and Non-trust

There are possible objections to our view on trust and reliance. For example, it has been argued that trust at its core is a two-place relation whereas reliance is a three-place relation. This suggests that trust cannot really be regarded as a form of reliance.[59] Now, trust can obviously be regarded both as a one-place relation or property ("A is trusting"), as a two-place relation ("A trusts B"), and as a three-place relation ("A trust B to C"), but the three-place relation is probably the most common one.[60] Some seem to believe that the two-place relation is most fundamental,[61] whereas others, including ourselves, tend to regard the three-place relation as

[55] E.g. Baier 1986, p. 235, Yamagishi and Yamagishi 1994: Jones 1996, pp. 4, 8f, Lahno 2001.

[56] Baier 1986, p. 235.

[57] Jones 1996, pp. 4, 8f.

[58] Alternatively, A trusts B to C because she believes that B shares her own norms or values Lahno 2001. "Virtue-based accounts" are a subset of goodwill accounts. According to such accounts, A trusts B to be x sort of person with regard to y, where x is (from A's perspective) a positive quality of character or way of performing an action and where y is some good that A values (Potter 2002, p. 17, see also Yamagishi and Yamagishi 1994 on character based trust). One problem with this view is that it is difficult to construct a coherent account of trustworthiness as a virtue from which A's trust is produced e.g. Jones 2012, pp. 82—85.

[59] Domenicucci and Holton 2016, pp. 153—155.

[60] E.g. Hawley 2014.

[61] E.g. Uslaner 2002, pp. 22—24, Faulkner 2015, p. 424, Domenicucci and Holton 2016, pp. 150—154.

most fundamental.[62] But this is of little importance, since our view can be modified to account for two-place trust as well.

We suggest that two-place trust involves attributions of responsibility, just like three-place trust, but a different kind of responsibility, namely virtue responsibility. In the case of three-place trust – when A trusts B to C – A judges B to regard himself as normatively responsible for C, and this is typically accompanied by a belief that B is in fact morally required to care about C. In the case of two-place trust – when A trusts B (without any explicit or implicit reference to any C) – A judges B to be a responsible or good person. It is not clear exactly what this involves in a context of trust, however, e.g., whether A simply believes that B is morally motivated, or whether A also has to believe that B is generally competent, or at least that he is honest about his own incompetence. This view of two-place trust seems to equate trust with what Toshio Yamagishi calls character-based trust, where A trusts B because of what he knows about B's general character.[63] However, Yamagishi also suggests that two-place trust can be relational, i.e., that A's trust in B can also be based on information about B's feelings and attitudes toward A. We want to allow for this possibility as well, for ordinary language reasons, even if it makes our own view somewhat less precise and more complex.

Now, our expanded view of trust – that takes both two-place and three-place trust into consideration – might be regarded as less unified than some of the alternatives, but is satisfies the ordinary language condition quite well, since the term "trust" is used in both ways in everyday speech. In addition to this, our expanded view is unified in one important sense, since both virtuous and loving people tend to live up to their normative responsibilities. This suggests that if we trust someone in the two-place sense, i.e. if we perceive her as a responsible or caring person, we also trust her in the three-place sense, i.e. with regard to some particular C or other, on the assumption that we believe she is competent with regard to C and has the opportunity to C. If I perceive someone as having a virtuous character or as really caring about me, I do not just perceive her as normatively responsible for particular Cs, I also tend to believe that she regards herself as responsible, which is pretty much the same as trusting her in the three-place sense, with regard to some particular C (assuming that the competence and opportunity conditions are both satisfied.) In short, it seems that two-place trust is tightly connected to three-place trust (in our sense) regardless of whether it is "character-based" or "relational" in Yamagishi's sense.

Distrust and its implications can also be explained by our account.[64] In the case of two-place distrust, we think of the other as an irresponsible person, who is not motivated to live up to his normative responsibilities and who do not care sufficiently about what they should care about. This explains why I tend to feel remorse,

[62] Baier 1986, Holton 1994, Hardin 2002. The one-place relation can also be conceived of as fundamental, in some sense. For example, me being a trusting person can, according to some, explain why I trust the government to do what it has told us that it will do. When three-place trust cannot be explained in this way, it is, according to some, *strategic* or *contractual* (Uslaner 2002, Faulkner 2015, p. 424), which is regarded as less valuable than trust proper.

[63] Yamagishi 2011.

[64] Note that trust and distrust are contradictories and not just contraries (see above, as well as Faulkner 2015, p. 426; Hawley 2014: 3).

apologize, or ask for forgiveness if I distrusted someone whom I later realize I should have trusted. Is there also such a thing as three-place distrust, and if so, how should it be analyzed? It is rather common that A does not trust B to C, or does not rely on B to C, but not necessarily because she distrusts B (even though this is one possible explanation). A might also fail to trust B to C, or rely on B to C, because she regards B as weak-willed or incompetent regarding C.

However, we suggest that these are not really cases of distrust, but of non-trust or non-reliance. One reason for this is that we would not blame B for not C-ing if this is due to weak will or incompetence, we would at most be disappointed. On this view, distrust is always based on beliefs about character and motivation, but this is not to say that there is no such thing as three-place distrust, since I can regard someone as responsible in one area (like work) without regarding him as responsible in another area (like close personal relationships). If we regard someone as well-meaning but incompetent, it is a case of non-trust or non-reliance rather than a case of distrust. However, we tend to blame incompetent people if they make us believe they have the required competence, but in this kind of case, we do not distrust the person because she lacks the ability to C, but because it is irresponsible to pretend to have an ability she does not have.

4 Conclusions

In this paper, we have argued that there are eight conditions of adequacy that a definition of "trust" and "reliance" should ideally satisfy, and that can be used to assess competing definitions. These conditions are in part derived from the most important practical and theoretical purposes of a definition. We have also proposed our own definitions and argued that these definitions satisfy the relevant conditions of adequacy to a higher degree than the alternative definitions.

In our view, trust is a form of reliance, which can be regarded as the "neutral core of trust". We (roughly) suggest that we rely on someone to do or ensure something when we judge them to have the relevant competence, motivation, and opportunity. According to our account of three-place trust, which we have dubbed the Accountability view, we trust someone to do or ensure something when we also judge that they perceive themselves as having a forward-looking normative responsibility to care for whatever it is we trust them to care about. In most cases, we also think that they are in fact normatively responsible for the object of trust, and we tend to blame them if they do not fulfill their responsibilities.

Two-place trust can take two different forms: we either trust other people because we regard them as virtuous or responsible persons (character-based trust) or because we think they care for us (relational trust). In both cases, two-place trust can neatly explain three-place trust as we have defined it (given that the competence and opportunity are both in place). We have also suggested that the most interesting opposite of trust is distrust (in both the two-place and three-place cases), whereas the most interesting opposite of reliance is non-reliance. We have not yet investigated when our trust or reliance is *justified*, however, but in our next paper we will use the presented account to address this question.

Funding Open Access funding provided by Chalmers University of Technology

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Baier, A. (1986). Trust and antitrust. *Ethics, 96*(2), 231–260.

Baier, A. (2013) What is trust? In Ed. Archard, D., Deveaux, M., Manson, N., & Weinstock, D. (Eds.), *Reading Onora O'Neill*. Routledge.

Barber, B. (1983). *The logic and limits of trust*. New Brunswick: Rutgers University Press.

Bevan, G., & Hood, C. (2006). What's measured is what matters: targets and gaming in the English public health care system. *Public Administration, 84*(3), 517–538.

Björnsson, G. (2011). Joint responsibility without individual control: Applying the explanation Hypothesis pp. 181—199. In *Moral Responsibility*. Dordrecht: Springer.

Björnsson, Gunnar. (2014). Incompatibilism and 'bypassed' agency 95—122. In *Surrounding free will*.

Björnsson, Gunnar., & Brülde, Bengt. (2017). Normative responsibilities: structure and sources pp. 13—33. In *Parental Responsibility in the Context of Neuroscience and Genetics*. Springer International Publishing.

Björnsson, G., & Persson, K. (2012). The explanatory component of moral responsibility. *Noûs, 46*(2), 326–354.

Björnsson, G., & Persson, K. (2013). A unified empirical account of responsibility judgments. *Philosophy and Phenomenological Research, 87*(3), 611–639.

Blackburn, Simon. (1998). Trust, cooperation, and human psychology. In ed.. Braithwaite, Valerie. and Levi, Margaret. *Trust and Governance*. Russel Sage Foundation.

Bradley, B. (2012). Doing away with harm. *Philosophy and Phenomenological Research, 85*(2), 390–412.

Brülde, Bengt. (1998). *The human good*. Acta Universitatis Gothoburgensis.

Brülde, B. (2000). On how to define the concept of health: A loose comparative approach. *Medicine, Health Care and Philosophy, 3*(3), 303–306.

Brülde, B. (2007). Happiness theories of the good life. *Journal of Happiness Studies, 8*(1), 15–49.

Castelfranchi, Christiano., and Rino, Falcone. (2010). *Trust theory: A socio-cognitive and computational model*. Vol. 18. Wiley.

Coleman, James S., & Coleman, James Samuel. (1994). *Foundations of social theory*. Harvard university press.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology, 92*(4), 909.

de Fine Licht, Karl. (2017). Trust and trustworthiness –En bok med oförtjänt gott rykte ("Trust and trustworthiness –A book with a too good of a reputation"). Tidskrift för politisk filosofi ("Journal of political philosophy"). Available here: http://www.politiskfilosofi.se/fulltext/2017-1/pdf/TPF_2017-1_052-061_de_fine_licht.pdf.

Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust: global pattern or Nordic exceptionalism? *European Sociological Review, 21*(4), 311–327.

Domenicucci, J., & Holton, R. (2016). Trust as a Two-Place Relation. *The Philosophy of Trust*.

Faulkner, P. (2015). The attitude of trust is basic. *Analysis, 75*(3), 424–429.

Fukuyama, Francis. (1995). *Trust: The social virtues and the creation of prosperity* (No. D10 301 c. 1/c. 2). Free Press Paperbacks.

Gambetta, Diego. (1988). *Trust: Making and breaking cooperative relations*. Blackwell Publishing.

Hardin, Russel. (2002). *Trust and trustworthiness*. Russell Sage Foundation.
Hardin, R. (2006). *Trust*. Cambridge: Polity.
Hawley, K. (2014). Trust, distrust, and commitment. *Noûs, 48*(1), 1–20.
Hawley, Katherine. (2017). Trustworthy Groups and Organizations. In *The philosophy of trust* ed. Faulkner, P. and Simpson, T. Oxford U P.
Hertzberg, L. (1988). On the attitude of trust. *Inquiry, 31*(3), 307–322.
Hieronymi, P. (2008). Responsibility for believing. *Synthese, 161*(3), 357–373.
Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy, 72*(1), 63–76.
Jones, K. (1996). Trust as an Affective Attitude. *Ethics, 107,* 4–25.
Jones, Karen. (2004). Trust and terror pp. 3—18. In Peggy DesAutels & Margaret Urban Walker (eds.), *Moral Psychology: Feminist Ethics and Social Theory*. Rowman & Littlefield.
Jones, K. (2012). Trustworthiness. *Ethics, 123*(1), 61–85.
Jordan, J. J., et al. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473–476.
Lahno, B. (2001). On the emotional character of trust. *Ethical theory and moral practice, 4*(2), 171–189.
Lahno, B. (2004). Three aspects of interpersonal trust. *Analyse & Kritik, 26*(1), 30–47.
Levi, M., & Stoker L. (2000). Political trust and trustworthiness. *Annual Review of Political Science, 3*, 475–507.
Luhmann, Niklas. (1979). Trust and power/two works by Niklas Luhmann; with introduction by Gianfranco Poggi.
McLeod, Carolyn. (2015). Trust. *The Stanford Encyclopedia of Philosophy. Ed.* Edward N. Zalta, http://plato.stanford.edu/archives/fall2015/entries/trust/.
Nickel, P. J. (2007). Trust and obligation-ascription. *Ethical Theory and Moral Practice, 10*(3), 309–319.
Nickel, Phillip J. (2017). Being pragmatic about trust pp. 196—213. In Faulkner, Paula & Simpson, Thomas, *The Philosophy of Trust*. Oxford University Press.
Pettit, P. (1995). The cunning of trust. *Philosophy & Public Affairs, 24*(3), 202–225.
Potter, N. N. (2002). *How Can I be Trusted? A Virtue Theory of Trustworthiness*. Lanham: Rowman & Littlefield.
Power, M. (1997). *The audit society: Rituals of verification*. Oxford: O U P.
Putnam, R. D. (1993). The prosperous community. *The American Prospect, 4*(13), 35–42.
Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy, 6*(1), 65–78.
Rothstein, Bo. (2005). *Social traps and the problem of trust*. Cambridge University Press.
Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of personality, 35*(4), 651–665.
Sumner, Leonard Wayne. (1996). *Welfare, happiness, and ethics*. Clarendon Press.
Uslaner, Eric M. (2002). *The moral foundations of trust*. Cambridge University Press.
Uslaner, Eric. (2009). *Is eleven really a lucky number? Measuring trust and the problem of clumping*. working paper available at http://www.bsos.umd.edu/gvpt/uslaner/uslanertrustclumping.pdf.
Yamagishi, T. (2011). *Trust The evolutionary game of mind and society*. Springer.
Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*(2), 129–166.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.