

2 The Concept of Trust

“... trust is a term with many meanings” (Williamson 1993: 453).

To begin a scientific conceptualization of trust, we can ask, “What does the term *to trust* really mean?” In answering this question, we reveal the term’s diverse and equivocal use in vernacular language. We may “trust” other people with respect to their future actions, or “trust” organizations with respect to the promised quality of their products. We “trust” a doctor when we see her to cure us, as well as with respect to her abilities and intentions to heal us. When driving in traffic, we “trust” others to abide by the rules, just as we do. Some people “trust” the government, while others only “trust” in god or in themselves. Obviously, in each example, the term “to trust” refers to a different situation and to a different object, and connotes a different meaning. In fact, it is impossible to uncover a consistent and universal notion of trust based on the everyday usage of the term (McKnight & Chervany 1996). Some researchers argue that the analysis of ordinary language is a futile tool in trust research, not only because it cannot promote one meaning of “trust” above all other candidates, but also because the use of the term—and even its very existence—varies greatly between different languages (Hardin 2002: 58). On the other hand, if scientific definitions are too distant from their everyday counterparts, they run the danger of missing important dimensions of the concept under scrutiny, and such definitions should therefore be informed by common-sense understanding (Kelley 1992). This is especially true for the concept of trust, which has a breadth of meaning in everyday life.

Unsurprisingly, trust researchers commonly relate scientific conceptualizations of trust to their everyday counterparts, for example by comparing dictionary and scientific definitions (Barber 1983, McKnight & Chervany 1996), by explicitly assessing lay theories of trust and individual-qualitative experiences (Henslin 1968, Gabarro 1978, Kramer 1996, Weber & Carter 2003), and by analyzing the meaning of the term in everyday language (Baier 1986, Lahno 2002). Most importantly, these studies suggest that trust must be conceived of as a multidimensional concept which (1) develops only under certain structural conditions and (2) merges different cognitive, affective, and behavioral dimensions into a unitary social experience (Lewis & Weigert 1985b, Rempel et al. 1985, McAllister 1995). We will therefore organize our introductory study of the concept along these two fundamental ingredients of trust: the objective structure and the subjective experience of trust. As it turns out, this course of action is most helpful in delineating the core problems and contradictions that mark the current state of trust research.

To begin with, a reasonable level of consensus exists on the structural prerequisites which form the set of conditions necessary for trust to arise (Rousseau et al. 1998: 395). We will refer to these as the *objective structure* of trust. Most importantly, the situation must involve a risk which stems from the trusting parties' uncertainty about the preferences of the trusted party. Secondly, the situation must be marked by social interdependence, meaning that the interests of one party cannot be achieved without reliance on another party. As a corollary, the trusting party will have to transfer control over certain resources or events to the trusted party (Coleman 1990) and thereby become objectively vulnerable (Heimer 2001). Lastly, trust has to be future-oriented, in the sense that the outcome of trust cannot readily be observed, but will be determined at a more or less specified point of time in the future (Luhmann 1988).

In contrast, the *subjective experience* of trust—that is, the internal mental state associated with trust—seems to be one prime reason for the “confusing potpourri” (Shapiro 1987: 625) of trust definitions in the literature. Although there appears to be a substantial consensus on defining trust primarily as a mental state, there is much less agreement when it comes to the precise definition thereof. The propositions that have been offered by researchers are contradictory (Bigley & Pearce 1998). In a crossdisciplinary review of 60 articles on trust, McKnight and Chervany (1996) find that 50% of scientific definitions include cognitive elements, such as expectations, beliefs, and intentions, while 37% include affective elements, such as feelings of confidence and security. Almost all relate trust to some form of action. About 60% of all definitions locate trust on more than one dimension. What is more, some authors conceive of trust of as a state that is explicitly *not* perceived until it is broken, as something “non-cognitive” (Becker 1996), related to aspects of automatic decision making (Hill & O'Hara 2006, Schul et al. 2008) and preconscious processes which shape perception (Zucker 1986, Luhmann 2000, Lahno 2002). In a nutshell, while trust has clear structural antecedents, its subjective experience is a source of disagreement among trust researchers, because different experiential phenomena come into focus. As a result, the acknowledged phenomenology of trust and its resulting definition differ remarkably between research traditions and disciplines.

This poses a challenge to the development of a comprehensive theory. If the phenomenological foundation of trust is variable, then it is no surprise that its subsequent conceptualization and definition remain contradictory. Different authors focus on different phenomenological aspects of the same explanandum. As a consequence, trust definitions have historically become too narrow and “homonymous,” preventing theoretical formulations and empirical results from becoming comparable and accumulating (McKnight & Chervany 1996). Although everyday language cannot provide a unitary precise definition, it does

help trust researchers to sharpen their conceptualizations, because it suggests how the phenomenology of trust should be conceived of. As it is, analysis of ordinary language indicates that the subjective experience of trust is in fact multifaceted (for example, consider the use of the term in idioms such as “I trusted you blindly!” versus “Trust, but verify!”). Ultimately, in moving towards a broad theoretical framework, we will have to absorb this wide phenomenological foundation of trust, and to explain when and why the subjective experience of trust differs so vastly between situations, as well as which factors (internal and external) promote, constrain, and shape the subjective experience of trust.

The following chapter presents a comprehensive review of the core concepts of trust research. Moving from objective structure to subjective experience, the “ingredients” of trust are systematically explored, thereby also revealing the challenging diversity in the trust definitions present in the literature. Even though the amount of research reviewed may initially appear confusing and contradictory, this exercise is nonetheless most rewarding, as it enables the general lines of conflict, the stumbling blocks, and bones of contention in trust research to be carved out. The chapter includes a delineation of the conceptual boundaries between trust and related concepts, such as confidence, system trust, distrust, and the like. As it is, the concept of trust is often used only in conjunction with a host of related concepts. Some of these are antecedents to trust; others only seem to be related, and are frequently confused with trust. The discussion of the conceptual boundaries completes the introduction of the trust concept, and rounds up the terminology that will be used throughout the book. The chapter closes with a discussion of the relation between objective structure and subjective experience. It is argued that a prime reason for the diversity and contradictions among trust definitions is rooted in the fact that trust researchers rarely pay attention to the process of interpretation—the subjective definition of the situation—and how it relates to the objective structure of the trust problem. The central question that this chapter opens up is, “how does the objective structure of the trust problem translate into the subjective experience of trust?”

2.1 Objective Structure

2.1.1 *Constituents of Interpersonal Trust Relations*

Trust is a phenomenon that we usually ascribe to other persons or to ourselves, but not to “things” or inanimate objects. Although we have not yet dealt with the question of the subjective experience of trust, it is clear that we can always find *ex post* reasons which have motivated a trusting act, even if that trust was misplaced. In other words, individual and purposeful *actors* are the primary *subjects of trust*. This raises the question of whether collective actors can be a subject of

trust as well. The “US government” might trust the “North Korean Regime” with respect to nuclear policies, for example. We could even extend the concept to include collective “trust systems” (Coleman 1990) and the variations of collective action they enable. As Coleman notes, “the analysis of these phenomena requires going beyond the two- or three-actor systems, but it can be done through the use of these components as building blocks” (ibid. 188). The analysis of collective actors, collective trust systems and collective action is, however, beyond the scope of this book. The present work seeks to conceptualize and understand trust as an outcome of individual framing and decision-making processes. If trust is conceived of as a mental phenomenon, then the individual microlevel represents its sole level of emergence. Hence, the concept will be limited to individual actors and their dyadic relations in the following. As a subject of trust, these actors will be called *trustors*.

Trust generally denotes a special relation between two actors, but it can additionally refer to a relation between a trustor and other groups, organizations, or more abstract social institutions. Other actors, groups, organizations, institutions, and the like are examples of primary *objects of trust*. Many researchers have developed ideal-type trust classifications to refine these basic distinctions, for example by varying the degrees of social distance and generalization of the objects of trust (Bigley & Pearce 1998, Sztompka 1999). But given the broad experiential basis that trust can assume, any attempt to identify, classify, and validate various ideal-type objects of trust must remain incomplete and arbitrary (Möllerling 2001). The case of *interpersonal* trust between two actors can nevertheless be regarded as a prototypical case. Dyadic trust relations form the micro social building blocks of larger systems of trust—this makes their comprehension a premise for understanding a wide range of social phenomena, including social integration at large. Focusing on these building blocks, the concept of trust will in the following be limited to a relation between two individual actors. Trust in groups and collective actors, in more abstract institutions (such as expert knowledge systems), and trust in the political or economic system as a whole will not be subject to detailed analysis. These types of trust will be delimited from the concept of interpersonal trust under the rubric of “system trust.” They are relevant insofar as they can become the basis for certain types of interpersonal trust that use system trust as a starting point (a topic which will be more fully explored in chapter 3.2).

The special relation between two actors to which trust refers will be called a *trust relation*. Its emergence is the prime explanandum of this work. Trust relations are always a three-part relation in the form of “A trusts B with respect to X” (Hardin 2002: 9). In a trust relation, the second actor B—the object of trust—will be called the *trustee*. The item X is called the *content of the trust relation*.

Trustor, trustee, and the content of the trust relation are the main constituents of interpersonal trust relations. Note that, for the moment, we restrict the trust relation to a one-way relation. The notion of trust is often connected to the idea of a situation of mutual trust and iterated exchange. These are aspects of “social embeddedness” and of corresponding two-way trust relations in a social environment, which will be introduced later. Interpersonal trust, even when it is conceptualized as a psychological state of the trustor, is always relational and social in the sense that the trust relation necessarily extends to the dyad and “transcends” the boundaries of the individual (Lewis & Weigert 1985b).

2.1.2 *The Basic Trust Problem*

The emergence of a trust relation is contingent upon certain structural prerequisites which confront the trustor with a particular decision-making problem, when he has to decide *whether or not to trust*. In the *basic trust problem*, a trustor faces two different sets of actions:

(1) Actions from the first set make him vulnerable with respect to the actions of the trustee. The trustor transfers control over resources or events to the trustee, and the trustee’s future actions determine whether the trustor will experience a loss or a gain. While making his decision, the trustor cannot foretell with certainty how the trustee will decide, and he cannot rely on sanctions or any other form of external enforcement to induce the desired outcome.

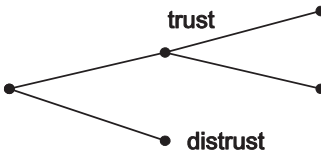
(2) In contrast, actions from the second set allow the trustor to eliminate potential damage, with certainty and from the beginning. In this case, the trustor does not put himself into a position where the trustee can determine the loss or gain. By refraining from a transfer of resources or control, the trustor is not taking a risk, and he prevents getting into a vulnerable position; he can maintain the *status quo* with certainty.

If a trustor chooses actions from the first set, we say that “*A trusts B*,” and the observable action will be called a *trusting act*. The choice of a trusting act constitutes the trust relation between the actors. If A chooses actions from the second set, then “*A distrusts B*,” and the observable action is distrust.

It is now possible to spell out more precisely the content X of a trust relation: If “*A trusts B with respect to X*,” then all classes of actions which (1) do not harm A, and (2) serve to realize the prospective gain and utility increase for A, belong to the content of the trust relation X. If B chooses his actions accordingly, he is *trustworthy*, he *fulfills* A’s trust, or he *honors* A’s trust. Note that a trustworthy response often leaves some “latitude of judgment” to the trustee. That is, the content X may be very specific, and may demand a unique course of action, or it may be more general, defining the desired outcome state, but not the precise actions necessary to achieve it. Mutual gains can be achieved through trust and

trustworthy response, but there are also incentives for the trustee to choose the untrustworthy option and to fail A's trust (Messick & Kramer 2001: 91). Principally, a trustworthy response requires some form of effort (time, energy, or other resources) and thus has a cost to the trustee which he can save by simply not fulfilling the content of the trust relation. Deutsch (1958) emphasizes that trustworthiness implies that the trustee will fulfill the content of the trust relation, even if violating trust is more immediately advantageous. The interests of the trustor are violated if the trustee disregards the content of the trust relation. In that case, "B *fails* A's trust" and violates the trust relation. Coleman emphasizes that a breach of trust must put the trustor in a worse situation than if he had not trusted (Coleman 1990: 98f.). The following picture summarizes the basic trust problem (figure 1):

Figure 1: The basic trust problem

Trustor A	Trustee B	Payoff A	Payoff B
	honor	1	1
	fail	$a < 0$	$b > 1$
	distrust	0	0

In this case, the *status quo* payoffs are zero for both actors, the successful establishment of a trust relation yields payoffs (1|1), respectively. A failure of trust puts the trustor into a worse position as compared to *status quo*, while the trustee experiences some gain that puts him in a better position than a trustworthy response. To conclude, the trustor has to decide whether to trust the trustee with respect to X. The content of the trust relation comprises all actions which improve A's utility with respect to the status quo and "realize" the content of the trust relation. These prospective gains present a basic motivation to engage in the trust relation. However, the transfer of control over resources or events involves the risk of incurring a loss if the trustee disregards the content of the trust relation and fails A's trust. These prospective losses present a basic risk. While mutual gains can be achieved from trust and trustworthy response, there is also an incentive for the trustee to fail the trust. Taken together, these structural conditions constitute the basic trust problem.

2.1.3 Trust and Action

As we have seen, if “A trusts B with respect to X,” then the trustor chooses an action which makes him vulnerable with respect to the actions of the trustee. On the level of overt behavior, we can observe the choice of a trusting act, manifested as a transfer of control over resources or events to the trustee. In other words, the behavioral content of trust is a risky course of action through which trust is demonstrated (Lewis & Weigert 1985b). The choice of a trusting act can be interpreted as a “risky investment” (Luhmann 2000: 27), because the trustor must transfer control over certain resources or events to the trustee, and at the same time is incapable of determining the final outcome (the potential gain or loss) with certainty. So far, we have used the notion of a “choice” that the trustor “decides” on without further explication. But the issue is not trivial, and warrants closer inspection: how are trust and choice related? Is “trusting” equal to “making a choice”?

In the “behavioral approach” to trust (Kramer 1999), researchers define trust solely in terms of cooperative choices in an interpersonal context (e.g. Deutsch 1958, 1960, Loomis 1959). Trust is defined as a behavioral outcome based on sufficiently positive expectations which allow the trustor to choose a risky course of action (Gambetta 1988a, Coleman 1990). Essentially, trust is regarded as a rational choice among actions. For instance, Gambetta suggests that trusting someone means that “the probability that he will perform an action that is beneficial to us ... is high enough for us to consider *engaging in some form of cooperation* with him” (1988a: 217, emphasis added). One advantage of this approach is that it opens up the toolbox of rational choice theory, which then can be applied to the study of trust decisions (Messick & Kramer 2001: 91). Defining trust as choice behavior also warrants that it can be examined with the help of experimental devices (Fehr 2009: 238).

Most authors maintain, however, that trust fundamentally differs from the choice of a trusting act. The trusting act merely displays the observable behavioral outcome of trust. In essence, “the fundamental difference between trust and trusting behaviors is between a ‘willingness’ to assume risk and actually ‘assuming’ that risk” (Mayer et al. 1995: 724). Pointing to the question of choice, Hardin notes, “Trust is in the category of knowledge, acting on trust is in the category of action ... I do not, in an immediate instance, choose to trust, I do not take a risk. Only actions are chosen” (2001: 11ff.). In this line, authors often conceive of trust as a “mental phenomenon” (Lahno 2002: 37) and a “psychological state” (Kramer 1999: 571). It is intrinsically tied to the subjective categories of knowledge and affect. When trusting someone, the trustor—in some way—accepts the vulnerability inherent in a trust problem. The choice of a trusting act merely displays an observable outcome.

Do we “choose” to trust, then, or not? This far from trivial question will be subject to analysis throughout the book. As we will see, trust can be understood as a two-step process, which begins with interpretation—that is, with the subjective definition of the situation—and leads to (and prepares for) a subsequent choice of action. The choice of the trusting act can approximate a rational decision-making process, but it need not. As such, “choosing to trust” can be an appropriate empirical description, but it can also be misleading if automatic processes prevail. To keep the terminology simple, we will use the notion of choice (“if A chooses to trust B,” “if A trusts B,” “the choice of a trusting act” etc.) to denote the fact that a trustor has opted for a transfer of control and is in a vulnerable position—ignoring, for the moment, the way in which this outcome has been internally achieved by the trustor.

A closely related question is whether trust and cooperation are the same (as suggested, for example, by Gambetta’s definition, cited above). Although observable cooperative choices and trust are intimately related—cooperation can be a manifestation of trust—it is problematic and confusing to simply equate the two concepts (Good 1988, Mayer et al. 1995, Hardin 2001). Cooperation may occur for many reasons, even when there is no risk taken, no potential loss at stake, or no choice available. The most extreme case imaginable may be a situation in which cooperation is enforced by deterrence and the threat of punishment, against the will of the actors involved. While this, to an alien observer, might look like some form of cooperation, it is clearly not an outcome of trust. Therefore, whenever we observe cooperation, we must carefully ascertain whether or not we can ascribe trust to the actors involved in the interaction. Conceptually, these are not the same. As it is, cooperation must be conceived of as one indicator, among many others, of the latent construct of trust (McKnight & Chervany 1996: 32). In fact, researchers have recently devised means of experimentally separating trust from cooperation, showing that an upward spiral of benign attributions and increased cooperation is involved in the mutual build-up of trust (Yamagishi et al. 2005, Ferrin et al. 2008).

2.1.4 Social Uncertainty

Assume a state of perfect information. If a trustor knew the trustee’s preferences, corresponding motivations, and intentions with certainty, he could predict whether the trustee would honor or fail his trust. In this deterministic setting, there would be no “need” for trust, although it would nonetheless be possible to observe trusting acts. Since the trustor can determine the outcome and knows all preferences and incentives, his actions can more adequately be described as *reliance* (Nooteboom 2002). Imagine, on the other hand, that no information at all was available; then the trustor could just as well roll a dice to make the decision

in a trust problem. In this setting, there would be no “opportunity” for trust, although again we could observe trusting acts. In this case, it is more adequate to speak of *hope* (Lewis & Weigert 1985b, Luhmann 2000: 28). Trust is limited to instances where specific knowledge structures play a crucial role in the solution of a trust problem (Endress 2001: 175). In short, the concept of trust addresses a state of knowledge that is neither perfect, nor completely ignorant (Simmel 1992: 392).

In a trust problem, information is asymmetric and imperfect. While certain characteristics of the trustee (his preferences, motivation, and intentions) are hidden to the trustor, they are perfectly well-known to the trustee. At the same time, the trustor usually has at least some information that he can use in a given trust problem—but this information is typically imperfect. In the case of imperfect information, the likelihood of an event (for example, “trustworthy response to trust”) can be assessed as, at best, some probability. We will henceforth refer to the subjective assessment of the probability of an event as an *expectation*. Depending on the “precision” of his expectations, a trustor subjectively faces either a situation of *risk* or of *ambiguity*. The terms expectation, risk, and ambiguity will be treated as subjective categories, and they describe the internal representations of the objective uncertainty involved in the trust problem (see chapter 2.2.2).²

What is the source of uncertainty for the trustor? In a trust problem, uncertainty is based on the fact that the trustor cannot control the outcome of the trusting act once he has chosen to trust, and also on his imperfect information about the trustee. Since the outcome genuinely rests upon his interaction with the trustee, the trustor faces a fundamental *social uncertainty* (Kramer 2006: 68). Social uncertainty is endogenous to interactions. It results from the contingent decisions of other actors, and becomes relevant whenever the utility of an actor is directly or indirectly influenced by the decisions of others. As we have seen, this is exactly the case in trust problems. Unlike exogenous, environmental uncertainty, an actor can to some degree influence endogenous, social uncertainty. For example, a trustor can mitigate social uncertainty by opting for the safe alternative of distrust. However, he can never avoid it when he chooses a trusting act. Social uncertainty is a constitutive element of the basic trust problem.

All in all, trust problems are characterized by social uncertainty and asymmetric, imperfect information. The trustor has limited information about the

² Note that some authors (e.g. Knight 1965) have used the term “uncertainty” to refer to a subjective state of “ambiguity” (Camerer, Weber 1992: 326). The terminology adopted here follows Camerer & Weber (1992) and minimizes the risk of confusion: Uncertainty is *objective*, while risk and ambiguity are *subjective*, as discussed in chapter 2.2.2 below.

trustee's preferences, motivations, and intentions and, when confronted with the choice of a trusting act, cannot be sure of the future actions of the trustee. As noted by many researchers, the mix of social uncertainty and imperfect information, paired with the possibility of opportunistic action, is a core element of trust problems (Dasgupta 1988, Gambetta 1988a, Mayer et al. 1995, Luhmann 2000, Hardin 2001, Heimer 2001, Kramer 2006). The fact that these elements are structural prerequisites and objective "facts" does not, however, give an answer to the question of how trustors subjectively handle the uncertainties involved in a trust problem. This question is, as we will see, the crux of trust research, and it will be dealt with throughout the remainder of the book. Most trust researchers agree that trust is a special way of dealing with social uncertainty and imperfect information. In choosing to trust, the trustor—somehow—bypasses the social uncertainty inherent in the trust problem and takes a "leap beyond the expectations that reason and experience alone would warrant" (Lewis & Weigert 1985b: 970). We can conclude at this point that trust hints at the particular nature of the expectations involved, a particular way that they emerge and form, and a particular way of dealing with the uncertainty and the subjective risk or ambiguity involved in a trust problem.

2.1.5 Vulnerability

The concept of trust is almost routinely linked to the aspect of vulnerability (Hosmer 1995, Bigley & Pearce 1998). In fact, almost all research on the topic of trust rests on the idea that actors, in some way or other, become vulnerable to each other during their interaction. Many authors specify this vulnerability by referring to the objective structure of trust. From this perspective, vulnerability simply means that something must be "at stake" for the trustor. Trust always includes a transfer of control, and therefore results in the objective vulnerability of the trustor (Deutsch 1960, 1973). In the absence of vulnerability, the concept of trust is not necessary, as outcomes become irrelevant to the trustor (Mishra 1996). Vice versa, vulnerability increases with the proportion of total wealth that is at stake in an interaction (Heimer 2001). In short, vulnerability originates from the interaction in a basic trust problem, and it mirrors its incentive structure.

In contrast, many authors emphasize the importance of the subjective perception of vulnerability for trust, and use the term only with reference to the trustor's subjective experience. In contrast to a structural prerequisite or mere consequence thereof, the term vulnerability then describes a qualitative element of individual subjective experience—an internal response to the incentive structure. In this perspective, the most commonly emphasized elements are favorable expectations and the willingness and intention to be vulnerable. For example, Rousseau et al. propose that "trust is a psychological state comprising the *inten-*

Trust and Rationality

An Integrative Framework for Trust Research

Rompf, S.A.

2015, XIV, 415 p. 48 illus., Softcover

ISBN: 978-3-658-07326-8