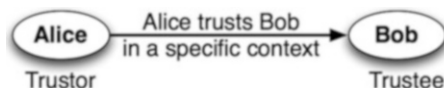


Chapter 2

Trust as a Computational Concept

Trust is a frequently used concept in many disciplines. In social sciences, the emphasis is on understanding how trust impacts the way people make decisions. The computing literature concentrates on the design of tools that can assist people in various tasks. Often these tools operate on a given model of trust and provide methods to measure trust in a specific application context. Before we go into the details of any such model or discuss the differences between different models, we first provide a broad definition of trust.

Fig. 2.1 Trust relationship between trustor and trustee



Trust is a relationship between a **trustor**, who we call Alice, and a **trustee**, who we call Bob (Fig. 2.1). Trust represents the trustor's willingness to be vulnerable under conditions of risk and interdependence [16]. Trust is not a behavior (e.g., cooperation) or a choice (e.g., taking a risk), but an underlying psychological belief that allows the trustor to put herself in a situation where she can be hurt or harmed by the trustee if her trust turns out to be misplaced. Trust is not a property of the trustor such as her propensity to trust. Nor is it a property of the trustee, such as his trustworthiness. Trust is a relationship between the trustor and the trustee that can be symmetric or asymmetric. Generally, the trustor trusts the trustee. The value of trust is either a quantitative or a qualitative measure that is used to check whether the trust relationship exists or to compare trust for different choices.

Regardless of the underlying model of trust, its value depends on the context. How much Alice trusts Bob depends on the underlying goal. For example, Alice trusts Siri on her iPhone to recommend her a good mechanic but not to choose a doctor. Why is this? The trust evaluation involves two different contexts: choosing a mechanic vs. choosing a doctor, which leads to two different evaluations. Furthermore, Alice's trust for Bob for the same goal at different times may vary based on external factors such as the existence of another

information source. Hence, trust is not a fixed value between two entities. It is a function of the trustor, the trustee and the context. What is in this context?

To understand the elements of context, we need to examine what factors impact the trust relationship between two entities. Going back to the example of whether Alice trusts Siri to find a doctor, Alice's goal is to find a good doctor. The criterion considered in determining trust could be the expertise of Siri on the given topic, i.e., health. The reason that Alice does not trust Siri for choosing a doctor could be that she does not yet think of Siri as an expert in the topic of health. However, Alice's needs in this topic may go beyond expertise. Does Siri know enough about what Alice looks for when searching for a good doctor? Maybe Alice cares greatly about the doctor's bedside manner. Can Siri account for this appropriately?

An important factor in deciding to choose Siri for such a recommendation could be the underlying intent behind Siri's operation. Is Siri built to help people or to promote certain information sources based on monetary considerations? In other words, does Alice believe that Siri has good intentions towards its user? Alice may not have firsthand information to arrive at a decision on this topic and may instead consider it as an Apple product. She might love Apple and trust that Apple products are built with users' needs in mind. As a result, she might believe in Siri's good intentions as an Apple product.

Alice's goal may be more complex than just getting a doctor recommendation. She might not have a problem with getting doctor recommendations from Siri as a way of generating an initial list and then sifting through them herself. She also thinks that she can send some names to her friends and get their opinion. The contextual component of a decision involving trust incorporates both the goals and the dependencies inherent in these goals: Alice's dependence on the intelligent agent, on the entity/company that produced it, on herself and on her friends.

Most trust modeling work does not explicitly mention the context of a trust evaluation because it is generally implied by the specific domain the work is originating from. While some complex models implicitly incorporate some elements of context, there is no explicit effort in this area. As we will see, different theories may apply when considering dependence on different entities and different goals.

2.1 An Example of Different Trust Contexts

The issue of context is quickly becoming very important due to increasingly sophisticated networks that combine social systems with computational systems. As a running example, we consider the newly emerging field of **co-robotics** (Fig. 2.2), which is dedicated to the development and use of robots that work beside, or cooperatively with, people. This field provides us with many interesting examples of trust from very different contexts. All of these different types of trust must be addressed in the design of the new generation of robots.

Consider a robot that is designed to help people construct buildings by providing physical support and help. How can the person trust the robot?

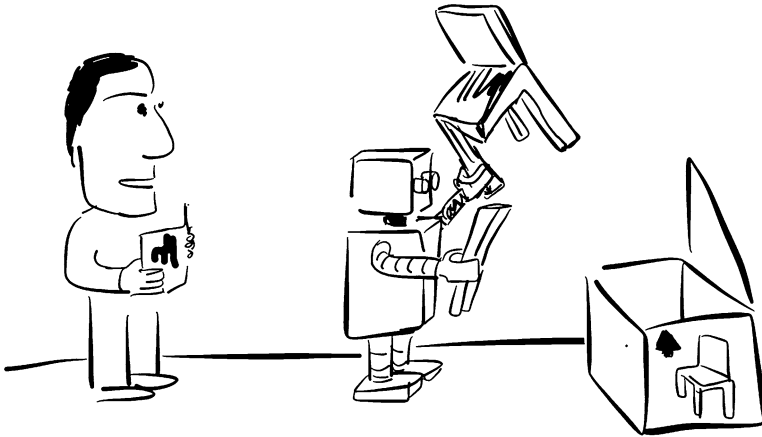


Fig. 2.2 Trust in co-robotics, humans and robots working side by side

- First, if the robot is meant to understand natural language and take commands from a human operator, its capacity to understand language correctly is an important component. The person needs to be able to trust that the robot will understand the commands given to it. The robot's ability to assess its current state by sensing its environment are crucial considerations as well.
- Second, if the robot is giving information to the person, the person has to trust this information to be able to use it in decision making. Hence, the robot's ability to use information gathered from different sensors in problem solving are relevant to deciding whether to trust the robot or not.
- Third, the robot should be able to accomplish tasks, such as grabbing objects properly, lifting and positioning as needed.
- The first aspect describes the sensing ability of the robot, the second aspect involves its problem solving ability and the third aspect refers to the physical ability of the robot. These aspects of capability are not sufficient to "trust" the robot for a specific action. The person needs to believe that the robot is designed in such a way that it will not cause bodily harm (by accidentally bumping into people, for example). This involves not only the robot's capability, but also a person's beliefs about the intentions behind its design.

Let us now look at the situation from the robot's perspective. The robot also has to trust the human team member in various ways.

- When should the robot follow the person's commands? Does the person know what he is talking about? What if the command will hurt the person or is based on an unsound move?
- If the robot is cooperating with the person, it has to take into account that people are not always as careful as robots. What if the person is not paying attention? What if he is getting tired? What if he will not be able to hold a heavy part in place?

Some of these issues are explored as part of trust in automation. What if the person trusts the robot blindly and stops monitoring it? It may happen that the robot can signal that it is experiencing a serious malfunction, but this may not register with the human team member. Unfortunately, grave consequences of this problem have been encountered in aviation.

As the human and robot work together, they evaluate their trust and calibrate their actions accordingly. This is called learning to work together. In a social context, this is how each team member learns the intentions of the others and builds relationships. For effective teamwork, the person needs to trust the robot for accomplishing tasks, as well as for providing correct and timely information. The robot needs to trust the person to be capable of completing the tasks he is taking on, giving correct information and reacting to problems in a timely way so as not to endanger others. These definitions correspond to many different trust contexts. They imply different goals, dependencies and in some cases, cognitive processes. The aim of this brief is to explain and categorize these differences. In the remainder of this chapter, we provide an overview of the distinctions before going into a review of the literature in this area. First, we describe networks and the new interdependencies introduced by them.

2.2 Trust Context in Networks

Networks, especially socio-technological networks, bring the additional element of interdependence. When the human and robot are working together, there is no additional network context. However, in many realistic scenarios, the actions of a person are constrained by the network environment in which they are embedded. In many cases, the network provides new resources and enables people to take actions that they cannot take alone. For example, social constructs help explain how people in the same network can benefit from the existing social relationships in that network. Social networks allow people to come together and accomplish bigger tasks than they can accomplish alone. The individuals trust one another within the context of a specific network that they are part of (see Fig. 2.3).

In networks, the social relationships are only part of the story. Technology allows people to interact with many other people that are not part of their social network, but still contribute to create value for themselves and others. Many tools and services continuously depend on human input in one way or another. Wikipedia uses human contributors and editors to create and curate data. Amazon Turk and other crowdsourcing systems allow people to come together to solve problems. In these systems, the participation is either voluntary or paid, changing the motivation of the participants. Any service that relies on human input has to incorporate methods to assess when and if such input can be trusted and how to process the input depending on the level of trust. In addition to the motivation of the participants, one has to consider the issues of bias and noise in the human input.

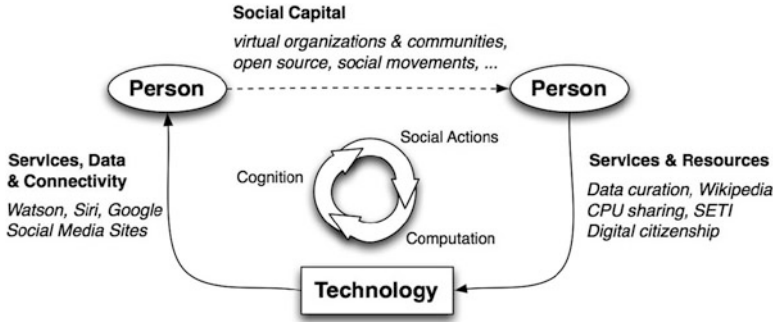


Fig. 2.3 Socio-technological networks introduce dependencies between people, dependencies of technology on human input, and human dependencies on tools provided by technology. All of these dependencies co-exist simultaneously in networks

Technology also relies on people to supply services to the network. In the case of CPU sharing systems like SETI@Home, the accessibility and reliability of computer systems are important concerns. Furthermore, systems and network security rely heavily on people abiding by protocols and implementing the necessary protection mechanisms. Similar to public health, networks remain healthy if the majority of its citizens are diligent about patching systems against vulnerabilities. In short, technology provides many sophisticated systems by trusting people to provide quality data, and secure and available systems.

These systems and services provided by technology are in turn used by people as resources in many daily activities. Many such services are of the type we just described: they rely on information created by people. IBM's Watson system used many sophisticated algorithms, analysis of past contestant behavior, and data sources like Wikipedia to answer Jeopardy questions better than any human contestant. Google relies heavily on the link structure of Web pages created by human input as well as the click behavior (pages people choose to click from those returned as answers to their queries). Social media sites depend heavily on user activity to decide which pieces of information and which users are prominent at a point in time. All these resources provide new capabilities to people, allowing them to find information and to connect with other people. People are also constrained by what these tools allow them to do. People trust these tools to get information and to accomplish tasks.

In summary, the networks provide context for trust relationships. People trust each other within a specific network. Socio-technological networks allow for new types of communities to form by online interactions. As a multi-faceted concept, trust is studied from many perspectives, such as social psychology, cognitive psychology, and computer science, among many others.

Some research addresses how communication mediated by technology is different than face to face communication. It is important to understand how people trust other people online. An equally important problem is how people trust information

they find online. Finally, researchers study how people trust systems that they use for information or assistance.

Computer science studies how systems can make use of human input for many different services. There are many algorithms that incorporate a notion of trust, either implicitly or explicitly, to accomplish this. These algorithms often need to consider both social and cognitive aspects that impact people's decisions. As a result, understanding trust from these aspects is crucial to the development of principled algorithms, which is what we tackle in the next chapter. We then follow it with a review of trust in computing. However, we note that the study of trust in networks must incorporate not only an understanding of trust in each field, but also the dependencies between all the different components.

2.3 Defining Trust in Its Context

In the introduction, we defined the trust context as the system level description of trust evaluation. In particular, the trust context incorporates a number of variables that are crucial in determining how the trust evaluation unfolds. We will consider these variables as the **elements of trust context** that must be defined to indicate clearly what the context is.

So far, we have discussed a number of contextual differences in trust evaluation. First of all, the trust evaluation may refer to either a human or a computational agent trusting another entity (**cognitive trust vs. algorithmic trust**). In this brief, we use the term **cognition** to refer to human cognition only, in other words cases in which the trustor is a human. Clearly, for cognitive trust, one has to consider the impact of cognitive and social processing on trust evaluation. In the case of algorithmic trust, the trustor is a program and the algorithmic properties of the agent are a concern. Ultimately, cognitive trust uses a different system than algorithmic trust. While some algorithms are designed to mimic cognitive trust, others are not. Furthermore, the trustor is part of a network that provides her with different institutions that help or limit her actions in various ways.

Another important element of the trust context is the **trustor's goals** and the **trustee(s) she depends on** for these goals. The difference in the goals is apparent when the trust is for an action vs. when it is for information. For example, trusting Bob to perform a task is different than trusting the information given by Bob (**trust for actions vs. trust for information**). When discussing information trust, we note that it is different than information credibility. In fact, information credibility is a factor that may change the processing of information trust: credible information may not be trusted and trusted information may not be credible. For example, a highly trusted source may tell us that the building is on fire. The information is not credible, but the source is trustworthy. When evaluating the information trust, the trustor typically considers two things: the trust for the source of information and her own evaluation of the credibility of the message. Both of these are part of

her goal. The trust evaluation may also take into account the network context. For example, the social network can be trusted to punish bad behavior and make it more likely that Bob will perform a specific action. In this example, both Bob and the network are trusted to some degree as part of the overall goal of the trustor. In short, trust evaluations involve complex goals with possibly multiple trustees.

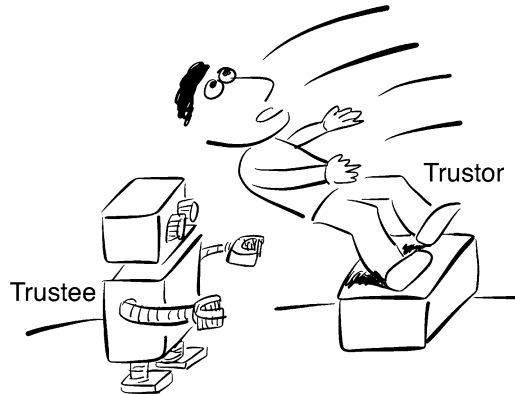
Different trust goals lead to different concerns and processing. While many factors can be considered in evaluating trust goals, we will concentrate on two main classes: the **competence** and the **trustworthiness** of the trustee. Most trust goals can be mapped to either one or both of these classes of goals, which we will refer to as **trust constructs**. When relying on Bob to hold a secret, his trustworthiness is the most important concern. When relying on Bob to hold a heavy object, his physical capability is more relevant. Informativeness of signals varies depending on which trust construct is being considered. For example, trustworthiness is impacted by first impressions, especially by those obtained from facial attributes. For competence, first impressions may come from information related to the education and the training of an individual. Information obtained from the network and the trustor's firsthand experience with the trustee play a role for both constructs, but the trustor may pay more attention to positive evidence for competence and negative evidence for trustworthiness. Furthermore, one expects that competence is specific to a topic, while trustworthiness is a more general belief that applies to a large number of goals. In short, the goals and the underlying trust constructs determine which signals are important for trust evaluation and how evidence is evaluated.

The trust evaluation differs for different types of trustees in various ways; judging the trustworthiness of a person versus the trustworthiness of a robot may be quite different in terms of human cognition. As a result, the trustee may have an impact on how trust is evaluated.

Finally, external conditions such as the ordering of events and availability of different signals may also change the result of trust evaluation for the same trustee(s) and the same goal(s). This is especially true for human trustors. These dynamic considerations are part of the **trust evaluation environment**, which describes the conditions that impact the trust evaluation beyond the trustor, the trustee and the trustor's goals. They can be considered as a part of the dynamic properties of trust context. Note that these elements of context are not orthogonal categories; they are inter-related in many ways.

In summary, the trust context is a system level description of trust evaluation that takes as input the following variables/tunable parameters: (a) the trustor and the network she is operating in, (b) her goal(s) in making a trust evaluation and the underlying trust constructs, (c) the trustee(s) that she depends on, and (d) the environmental conditions that impact the trust evaluation. Next, we investigate the impact of each variable in trust evaluation in detail. However, we first describe the preconditions of trust that are common to all trust contexts.

Fig. 2.4 Trustor and trustee in trust relationships



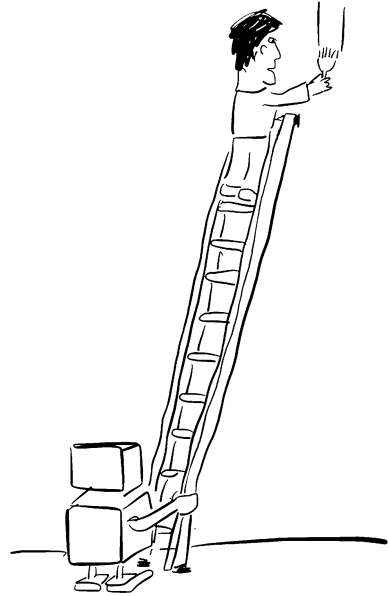
2.3.1 Preconditions of Trust

While the contexts we consider here are quite disparate, there are a number of preconditions of trust that uniformly apply across all the different contexts we consider. It is generally agreed that trust is a relationship involving two entities, a trustor and a trustee. The strength of the relationship is used as a measure of trust. The trustor trusts the trustee to accomplish a specific goal in a specific context. Trust allows the trustor to take actions and make decisions under conditions of **uncertainty** and from a position of **vulnerability** (Fig. 2.4) by **depending on the trustee**. In a trust relationship, the trustor encounters uncertainty at the decision time due to a lack of information or an inability to verify the integrity, the competence, the positive intentions, and other characteristics of the trustee. The trustor is vulnerable to suffering a loss if expectations of future outcomes turn out to be incorrect, i.e., if her trust is misplaced. Vulnerability and uncertainty together can be summed up as **risk**, the possibility that an action may lead to a loss.

2.3.1.1 Choices and Dependence

When making a trust evaluation, **dependence** describes to which degree the trustor relies on the trustee for her specific goals (Fig. 2.5). Many outside factors may determine the dependence. When judging trust for information, the ability of the trustor to determine the credibility of the information is crucial. If the trustor is not familiar with a specific topic, she may rely solely on the trustee's trustworthiness and competence in determining information trust. When dealing with an unknown trustee, the trustor may rely on the trustworthiness of a recommendation for that trustee. The trustor depends on the trustee for a specific choice, sometimes called as the **trusting choice**. For the decision to involve trust, there must be at least one other choice in which the trustor does not depend on the trustee.

Fig. 2.5 The dependence on the trustee



Some definitions of trust also emphasize that the trustor should be able to reject a choice. Even if there is more than one choice to be made, it is possible that some of these choices are not desirable at all. According to our formulation, as long as there is one other choice in which the trustor does not depend on the trustee, then the dependence on the trustee involves trust. One can argue that for any action, there is always another choice, i.e., not to take that action. Some would argue the opposite, that our decisions are predetermined by our mental predispositions, especially in cases where we have not considered the pros and cons of different choices consciously. Without going into a deep discussion about free will, we would like to instead emphasize that dependence relations can vary greatly depending on the perception of the trustor as to what her choices are.

As an example, suppose Alan is painting the house with the help from his robot helper, Chip. Alan trusts Chip to hold the ladder steady so that he can paint high up. Thus, the trustor is Alan and the trustee is Chip. By trusting Chip, Alan is able to complete the task. If Alan did not trust Chip, he cannot use the ladder and cannot accomplish his task. Alan might consider other ways to paint these hard-to-reach locations, but he is convinced that they will produce lower-quality workmanship.

2.3.1.2 Risk

Emphasized in the trust definition is the possibility to suffer negative consequences or to be disappointed if trust is misplaced. In other words, the trustor is made vulnerable as a result of trusting. The evaluation of consequences is generally

encapsulated by a utility function, the perceived desirability of a situation according to the trustor. A positive consequence is considered to have positive utility. Similarly, negative consequences are determined by their negative utility. The desirability of different situations is subjective and may change from trustor to trustor. Various methods to formalize these preconditions of trust have been introduced. We summarize these conditions following conventions chosen by Wagner et al. [19]:

1. The trustor has more than one option to choose from. The trust decision is made before the trustee acts.
2. The outcome in the trusting option t depends on the trustee x . In the non-trusting option nt , the outcome does not depend on the trustee.
3. The trustor is better off if she chooses t instead of nt and the trustee ends up being trustworthy, given by a utility function U .
4. The trustor is worse off if she chooses t instead of nt and the trustee ends up being not trustworthy.

The first condition corresponds to our definition of options. The second condition describes the notion of dependence. The final two conditions describe the notion of risk.

In our example, if Chip fails in his duty, Alan will fall and badly injure himself. This situation clearly has low or negative utility. If Chip does as he is expected, Alan will finish his job successfully, a desirable condition with positive utility. If Alan does not trust Chip, other options will not produce a nice coat of paint and lead to lower utility than trusting Chip.

We sum up the above options by the equation of expected utility U :

$$U(\text{trust} \wedge x = \text{bad}) < U(\text{not trust}) < U(\text{trust} \wedge x = \text{good})$$

It has long been argued whether the above elements can describe what trust is. The above are conditions that exist in all situations requiring trust: the presence of choices, the dependence on the trustee in the trusting choice, the uncertainty and the vulnerability of the trustor at the decision time. However, these conditions are not necessarily complete since they abstract out elements like the common understanding of expectations between the trustor and the trustee. Many sophisticated computational trust models have been developed with the aim of understanding how people trust others or for building agents that can trust another entity.

Trust is affected by the trustor's knowledge of her environment, as perceived by her. The environment provides her with input relevant to trust where the relevance is determined by her goals. Exactly how the environment and the trustor's knowledge impacts trust is typically part of the trust model implicitly. However, we will make this aspect of trust models explicit by our emphasis on the trust context in a wide range of applications. We will not concentrate on a single formalism for modeling trust, but instead we will describe how contextual elements change the trust models.

Next, we describe the crucial elements of trust context.

2.3.2 *Decision Maker: The Trustor*

Trust is studied in social sciences as a human construct. When people make decisions, trust allows them to choose one decision over the others by relying on each other. The trustee can be another person, a team or an organization. However, only a human being can trust a trustee. Even when discussing organizations trusting other organizations, social sciences take the view that decisions in organizations are made by human decision makers. In essence, the trust literature in this area aims to analyze how and why people trust others, how trust impacts their behavior and how systems can be devised to achieve certain social outcomes by influencing trust. Various perspectives shed light on this relationship: their social network, cognitive heuristics and biases and specific goals. Some researchers, for example concentrate on how people form opinions of others. There is also research in how different traits of a person affect their trust relationships or how emotions impact trust. Other researchers may look at various social advantages offered by trusting relationships. Some studies concentrate on how people attribute positive and negative features to different social groups that impact their trust. In some approaches, people are modeled as agents acting in self-interest and trying to modify their own utility. In this case, trust is a way of achieving future gains and systems designed to calibrate people's utilities can impact their trust.

These different approaches sometimes conflict with each other. There is continuing debate on whether human beings are rational [9]. If rationality is described as adherence to the assumptions of mathematical utility theory, there are many findings about how people make decisions that are at odds with this notion of rationality. Ultimately, human decision making depends on the trustor's view of the world; her value system, what she knows and remembers, and the cognitive process she uses to make a decision all play a role. However, these conditions are not fully known externally. All trust models with a human trustor operates with an underlying set of assumptions about how people evaluate trust. Sometimes these assumptions are explicit, and sometimes they are embedded in complex algorithms. Regardless of how these assumptions are defined, they are an approximation of human decision making.

In our previous examples, Chip trusted Alan also. Is this a new type of trust? Chip is evaluating trust using an algorithm based on a model of trust. In computing, a large number of trust models are proposed. Some attempt to mimic the cognitive trust by incorporating how trust beliefs form and are used in decision making [1, 10]. Some models do not have a social cognitive component, but attempt to assess properties of trustworthy entities and introduce optimal protocols for trusting. For example, an algorithm for routing in mobile ad-hoc networks measures how well each node is behaving to assess whether they are working properly and are not compromised [6]. The algorithm then concludes that a computational node can trust another. Another notion of trust is introduced by ranking algorithms, which try to assess which sites are likely to have more reliable links and updates its rank computation by valuing the links from these sites more highly [7]. These models of trust are often based on

observations of network behavior that incorporate system based and human input. For example, web sites are created and maintained by people aided by applications that automate many tasks. Trust for both needs to be taken into account. The algorithmic models of trust may differ from cognitive aspects of trust in many different ways. For example, they may incorporate fine-tuned processing of the environment with attention to statistical properties of different variables, criteria rarely considered in cognitive processing. They may also make some simplifying assumptions, disregarding inputs crucial to cognitive models.

In short, the trustor describes the underlying trust model. The trust model describes the trust evaluation system and how trust evaluation changes depending on context. We still distinguish between two broad classes of trust models. The first class, called cognitive trust tries to closely resemble cognitive and social processing of trust signals. The second class, called computational trust, is not strictly based how people trust each other, even though it may borrow some terminology from cognitive trust. A trust model may fall anywhere between these two broad classes.

It is debatable when a trust model can be called purely computational, not based on any cognitive model. This is a topic of study in many agent-based trust models where different cognitive components are introduced into trust models. An example of such a model is one in which agents evaluate trustees' actions based on their current trust beliefs of their trustworthiness. For our purposes, it is not important to draw this line accurately. What is important is to appreciate that many different system level descriptions of trust exist depending on who the trustor is.

2.3.3 The Trustor's Goals and Trust Constructs

In the previous sections, we gave many examples of cases in which the trustor had a choice in which she depends on a trustee to accomplish a certain goal. We have also shown cases in which the trustor's goal is complex and depends on multiple trustees. In this section, we will further discuss how the trust goal can be complex along two dimensions: the subgoals that make up a goal and the constructs used to evaluate trust for each construct. In particular, we make a distinction between information trust and action trust. We first start by discussing these two.

2.3.3.1 Evaluating Information Trust

We have seen that the trustor considers two separate subgoals when deciding whether or not to trust certain information.

- First, does she trust the information source to provide information that can be trusted?
- Second, does she trust herself to evaluate the credibility of the information?

The trustor is dependent both on herself and the source of information for this trust evaluation. How much weight each subgoal carries in the final evaluation of trust depends on many factors, one of which is the familiarity of the trustor with the information topic. If the trustor knows very little about a specific topic, her trust for the information may rely almost completely on her trust in the information source.

The information trust evaluation may involve other subgoals. Similar to the telephone game, online information travels through many media and can be altered along the way by people as well as programs. The trustor is also dependent on the entities that were involved in the transmission of a message. She may consider how much she trusts the transmission media as part of her information trust evaluation.

Suppose Alice sees facebook a post by Bob that says: "I lost 10 lbs on this amazing new diet, take a look at the following web site!" Alice may find it quite unlikely that Bob would post this type of information, and considers the possibility that this is spam. In this case, she does not trust this information. However, this is not due to her trust for Bob. She doubts the credibility of the message and considers the possibility that Bob's account used in this post was compromised. As a result, she attributes the message to someone other than Bob. This subgoal is sometimes referred to as the identity trust, which is the trust that the poster is in fact who he says he is. Note that identity trust is not necessarily an independent trust goal; it is likely evaluated conditionally.

In another example, suppose Alice is watching a video in which Bob is supporting a certain point of view. Again, Alice finds it unlikely that this video reflects Bob's true opinion. She may then consider the possibility that video may have been edited to alter its meaning. This may be due to the fact that the trust for the source presenting the video is not trustworthy. In these examples, the entities involved in the transmission and editing of information are conditionally considered in the trust assessment and are subgoals of information trust.

As a different example, suppose Alice and Bob are walking to a restaurant. Bob tells Alice that the restaurant is at 45th St, so Alice trusts him. It is possible that Bob got his information from Google Maps, but he has not told this to Alice. If he did, then Alice would base her trust evaluation on Google Maps' reliability, not Bob's. The trustee is in fact determined by the perceived source of the information. Alice could have also considered Bob's ability to look up Google Maps, but probably that would be a bit insulting. Here, Bob could be considered the transmission medium for this information.

Let us now consider a crowdsourcing system that uses a large amount of input, from many sources. The main assumption behind such a system is that if sufficient number of independent sources of information are consulted, then the final answer will be trustworthy regardless of how trustworthy each individual is. The design of this system in fact significantly reduces the dependence on any individual information source. The trust for information in such a system depends predominantly on the design of the system: how it collects and aggregates information. The properties of the underlying population of information providers are also a crucial part of assessing trust. Such a system provides an institution that

supports trust, and depends on the trustworthiness of its inputs. We will examine many examples of these types of institutions in the next chapters.

In short, trust for information may depend on a number of subgoals such as how much the perceived source of information is trusted to provide trustworthy information, how much the various entities along the information transmission path can be trusted to deliver it correctly, and how credible the information appears to the trustor.

2.3.3.2 Evaluating Trust for Actions

When trusting another entity to accomplish a specific task, the trustor's goal is determined by the given task. Suppose Alice is buying a book from an online seller, Bob. She pays Bob for the book. At this point, Alice is fully depending on Bob to send the book, and Bob controls the outcome of this transaction completely. Alice's goal is to get the book from Bob. What guarantees that Bob will not take the money and run? This is a topic of very old debate. Is it the fear of sanctions from the community? Is it the promise of future rewards for Bob if he cooperates? Is it morality and expected behavior patterns? Are these fundamentally different things?

Regardless of the underlying assumptions of why people honor contracts, we note that there are many institutions that constrain Bob's actions. These could be social institutions based on expected behavior, or institutions providing legal protections. Many online services provide computational reputation mechanisms that serve as a soft form of legal protection. Bob's reputation as a seller allows him to continue selling to other customers. The reputation management mechanism guarantees that if he misbehaves, this reputation will be damaged. As a result, he has reason to act trustworthy. In this instance, Alice is dependent on the effectiveness of the reputation system for her purchase in addition to her dependence on Bob.

Reputation management is also referred to as trust management in some computing literature. Some argue that this is not an appropriate term. In fact, trust management schemes of this form reduce the need for trust instead of enhancing it [4]. Instead, we define this as a shift of dependence from one trustee (Bob) to another (the reputation system), either partially or completely. A reputation system cannot completely eliminate Alice's dependence on Bob for arbitrary online markets, especially if Bob can easily create another online identity as soon as his current reputation is ruined. To this day, identity management is one of the most important problems in online transactions [2, 15]. To address this, the reputation system may require a history of transactions. Such a requirement may end up increasing the cost of entry into the market, which may not be desirable.

Alice's goals may be more complex than the successful completion of a task. She also cares about the likelihood that the product will be shipped quickly, will be packed well, and will arrive undamaged. Reputation alone may not be sufficient to support all these goals and additional institutional mechanisms may be needed, each resulting in a different dependency.

Similar considerations exist when the trustee is a computational agent or a robot. Alice needs to trust the robot, the specific piece of hardware that she is interacting with, to be in good operating condition. Furthermore, she needs to trust the underlying programming of the robot to be capable and trustworthy.

In short, similar to the case of information trust, for a specific trusting choice, there might be multiple goals and trustees. In such a case, the trustor depends on each trustee to some degree. Each trustee has some level of control over the outcome of the goal through a specific subgoal.

2.3.3.3 Trust Constructs: Trustworthiness and Competence

As we have seen in the discussion of information and action based trust, the trustor typically has a complex goal involving subgoals. Each subgoal defines the evaluation of specific aspects of the trustee. Trust research has generated a long list of keywords used in this effort, such as goodness, morality, benevolence, expertness, credibility, predictability, dependability, etc. There have been many efforts to categorize these considerations into distinct types [11, 17]. We will review these in the next chapter.

In the case of trusting actions, an often-used categorization separates trustworthiness from ability.

Trustworthiness refers to our expectation that a person will do as they say. This term incorporates dimensions like the trustee's integrity and good intentions.

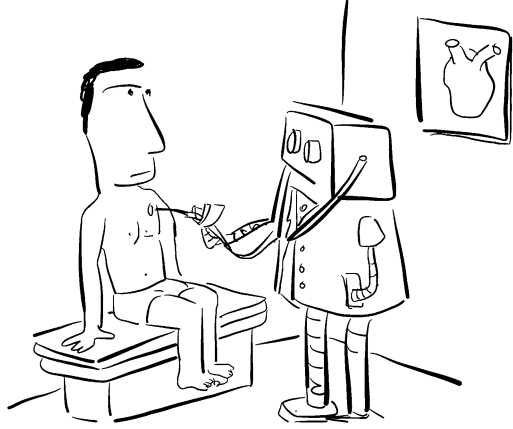
Ability is the trustee's competence in accomplishing a specific task (Fig. 2.6).

A trustworthy person may be dependable for completing a task, but their performance may not be satisfactory if they are not capable.

In reputation management, a system designed to force people to act in a trustworthy manner may sanction bad behavior [2]. However, sanctioning alone is not sufficient to separate competent but untrustworthy people from incompetent but trustworthy people. To assess the ability of people, online reviews have been found to be more useful [3]. These are called signaling mechanisms. For trustworthiness, tracking bad behavior is considered more diagnostic, while for ability, tracking and accumulation of good behavior is more meaningful. This agrees with research that explains how people form opinions of others [5]. Trustworthiness is a general measure of the trustee's reliability and is not dependent on a specific topic. If someone is our friend, we expect that he will do what he says and will help us when we ask for it. However, ability is dependent on a topic: a doctor's competence is in medicine and a mechanic's is in car repair.

Some research also concentrates on a third type of trust construct. In this case, the intentions of the trustees are considered [14]. Even if Bob does not act in a way that Alice prefers, he may still have Alice's best interests at heart. Sometimes these are called **good intentions**. An adversary with bad intentions may act in a trustworthy way to achieve advantages that he may later use against the trustor. In cognitive psychology, the evaluations of another's intent are studied as part of the theory of mind. According to this theory, people learn about other's beliefs and

Fig. 2.6 Competence of the trustee



thoughts by comparing them to themselves [12, 13]. Our understanding of ourselves and others evolves over time through shared social activities and cultural learning. These activities create a common set of expected behavioral patterns and social norms [18]. In short, we tend to think that people who act and talk like us probably think like us and have similar intentions. We project ourselves onto others. In-group and out-group dynamics and friend or foe categories typically result in judgments of intent. Those who are in the in-group are associated with positive intent, while those in the out-group are perceived as having negative intent. Due to the close links of good intentions to the trustworthiness dimension, most categorizations consider only the first two constructs for trust. However, most point out that an untrustworthy person is treated very differently than a person with bad intent, i.e., a distrusted person for whom trustworthiness is considered negative. When an untrustworthy person starts to act trustworthy, trust may eventually develop. However, a distrusted person's trustworthy actions may be met with suspicion and disregarded. As a result, trust may never develop. A simple example of this type distrust can be seen between individuals from social groups that have historically been in conflict. The burden of proof is much higher in these cases for trust relationships to develop.

Research in information trust [8] shows that people use various different constructs in this case as well such as truthfulness, believability, accuracy, objectivity, timeliness or reliability. It is also possible to classify these into two main categories. Truthfulness is more of an affective factor that represents a dependence on the trustworthiness of the source. Accuracy and timeliness on the other hand, are more closely linked to the competence of the source. Similar to ability, they are specific to a topic. In essence, parallel notions of trustworthiness and competence of the sources exist when considering information.

Believability, on the other hand, can be tied more closely to the information itself and the trustor's evaluation of it. As we discussed earlier, evaluation of information trust also involves the trustor's evaluations of the information. As this subgoal involves only the trustor, we can also consider this as a significant component of the trust evaluation context.

Trust goals generally map to constructs that fall in these very general classes, but they can be much more specific. Alan trusts his doctor to diagnose him correctly (competence) and tell him the truth (trustworthiness). He trusts Chip to hold the ladder so he does not fall (competence). Alice trusts Google Maps to have the correct answer (competence) and have the best coverage of location information (competence). Alice trusts Wikipedia to have the most objective (trustworthiness of the source) and the most comprehensive (competence of the source) information on a topic. Alice trusts Bob to send her purchase quickly as he promises (trustworthiness), and the electronic marketplace to provide accurate information about Bob (trustworthiness) and reduce his reputation score if she gives a bad comment (trustworthiness).

All these subgoals describe a trust construct and a trustee that the trustor is dependent on for that construct. Is there a distinction between how competence and trustworthiness are evaluated when the trustee is a person vs. when it is an automated system? This is a topic of research. We will review some of the literature on this issue in the next chapter.

2.3.4 Trust Evaluation: Cues, Network and Time Effects

The final critical component of trust context is the environment in which trust is evaluated. There are some significant effects that must be considered when human cognition is involved. We review some of these issues in this section. One can also find parallel issues of concern in the design of algorithms that compute trust. We will review those in Chap. 4.

The first effect we have discussed at length is the credibility of the information that is being considered. We have seen the complex decision making process involved in deciding whether to trust information or not. The source's trustworthiness may trump the credibility of information in some cases, and in other cases the reverse may happen.

Let's go back to Alice and Bob walking to the restaurant that Bob thinks is at 45th St. Even though Alice remembers quite well that it is around 34th, she decides to believe Bob. After a long walk, they reach 45th Street and find out that the restaurant is not there. A quick search reveals that in fact Bob's online source was wrong; Alice was right all along. She knew this, too! Why did she believe Bob? A well-known fact is that we are willing to override our own evaluations of information, even when we are quite sure, by those of a so-called expert. Often we say that we knew it all along, but still somehow ended up believing someone else. The reason is that evaluating information and reconciling conflicts take cognitive effort. Often, we are not willing to spend this effort unless we do not trust the source completely or some other outside conditions prime us to do so. Hence, the dependence for the information could change drastically depending on the situation.

Utilities are discussed in almost all trust research to define what the trustor expects to gain or lose in various situations. For cognitive trust, how the choices

are framed makes a big difference. For example, people value gains and losses differently. Framing the same situation as a potential gain or a potential loss that is averted may impact the decision greatly. The other well-known factor is the endowment effect: we will value something that we own much more than an equivalent item that is not ours. When multiple options are presented, the desirability of an option may also change when an irrelevant option is added. All of this points to the inevitable conclusion that utility is not a simple value evaluation: the framing effects must be considered in trust evaluation [9].

Almost all trust computation is based on a set of cues that are used to assess the trust for the trustees with respect to the specific goals. People use these cues daily, and computational models aim to mimic the human cognitive process. The social cognition of others is based on many different cues, including their faces, their social position, the stories we hear about them (social reputation) and our own experience with them. A similar set of cues exists for text, ranging from the appearance of the content to the authority of the source. We also form opinions of other entities, such as computer systems, information systems, or intelligent agents helping us using many similar cues. We will examine these in detail in the next chapter. We note that at the cognitive level, some of these cues are very easy to execute while others require effort. Kahneman [9] argues that cues that are less costly to process are more frequently used than those that are more costly. Furthermore, the order in which we evaluate cues may impact the trust evaluation; for example, when interacting with a person with an untrustworthy face, it may take us longer to trust them since trustworthiness of faces is processed much more quickly, and the resulting first impressions may impact our subsequent evaluations of trust.

Where the trustor's past experience of the trustee is concerned, some models consider all past experience, while some models consider that the trustor can forget or forgive various past events. The other important factor to consider is priming. For example, people who have been reading about social justice may view an article about wealth distribution differently than those who have been reading about problems with the welfare system. Priming is a well-known effect and used frequently by those involved in advertising or political propaganda. Different external cues may alter the evaluation of information credibility for the same article.

The other thing to remember is that the processing capability of the trustor plays a role in how much of the relevant input will be considered when making a decision. When we are busy or tired, we have fewer cognitive resources and may rely on simpler cues to evaluate trust. For example, relying on Bob's information about the restaurant's location was quite simple. It did not require Alice to think at all about whether the location is indeed incorrect by retrieving information about past visits from her own memory. It was much easier to acquiesce to Bob's opinion, especially at the end of the day when Alice was tired and hungry.

When considering networking effects, the nature of social relations can also be relevant. For example, an often debated issue is whether or not trust is transitive or not. Ultimately, it depends on the underlying social and trust context. For example, transitive closure is often found in social relations involving close friends. This is because two of Alice's friends end up hanging out together and become friends.

Friendship in this sense is a symmetric relation, and transitive closure implies that trust between friends is transitive. This is also reflected in the way we evaluate the intentions of others. Our friends have positive intentions towards us. However, not all friends are considered competent for a specific goal. There is no reason to expect that ability is symmetric or transitive. We will investigate this issue in more detail in the later chapters. Clearly, transitivity is an assumption that may apply to some trust constructs but not all.

Similar considerations exist when computational trust is concerned. The cues used to assess trust must be useful, and free of bias as much as possible. System design must take into account how to correctly optimize for trust. For example, a crowdsourcing system that makes previous votes available to voters introduces a bias towards the votes of the earlier voters. There is reason not to trust such a system. In our opinion, these issues impact trust computation greatly but have not yet been discussed in great detail.

In summary, we must consider which cues are available to the trustor to judge trust as well as the trustor's goals. Factors in the environment may impact how the trustor's choices are framed and whether the trustor is primed to evaluate them in a specific way. The mental load and alertness of the trustor may impact which factors will be evaluated. The order in which factors become available and are evaluated may change the final trust evaluation. The relevant environmental factors depend on the trust constructs and trustees. Modeling these factors is an important part of any trust model.

2.3.5 *tl;dr*

A common norm in online sites is to complete a long chunk of text with a section called *tl;dr* (Too Long Didn't Read). This section summarizes the trust context in a few sentences to accommodate readers with limited cognitive resources.

The trust context defines who the trustor is and who the trustees are. The dependence for a specific trusting choice specifies in detail which trustee to depend on for which goal and how much. A utility is defined for each specific choice, which in turn is used in defining trust.

However, to evaluate trust, a set of external cues regarding the trustees must be used based on the environmental factors. Their evaluation is greatly impacted by a set of external factors that need to be considered when modeling and measuring trust.

References

1. C. Castelfranchi, R. Falcone, *Trust Theory: A Socio-Cognitive and Computational Model* (Wiley, 2010)
2. C. Dellarocas, Reputation mechanism design in online trading environments with pure moral hazard. *Inf. Syst. Res.* **16**(2), 209–230 (2005)

3. C. Dellarocas, X.M. Zhang, N.F. Awad, Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interact. Mark.* **21**(4), 23–45 (2007)
4. D. Elgesem, Normative structures in trust management, in *Proceedings of the 4th International Conference on Trust Management*, Pittsburgh, PA, pp. 48–61 (2006)
5. S.T. Fiske, A.J. Cuddy, P. Glick, Universal dimensions of social cognition: warmth and competence. *Trends. Cognit. Sci.* **11**(2), 77–83 (2007)
6. K. Govindan, P. Mohapatra, Trust computations and trust dynamics in mobile adhoc networks: a survey. *IEEE Commun. Surv. Tutor.* **14**(2), 279–298 (2011)
7. Z. Gyongyi, H. Garcia-Molina, J. Pedersen, Combating web spam with trustrank, in *Proceedings of the 30th International Conference on Very Large Data Bases*, Toronto, 2004
8. B. Hilligoss, S.Y. Rieh, Developing a unifying framework of credibility assessment: construct, heuristics and interaction in context. *Inf. Process. Manag.* **44**, 1467–1484 (2008)
9. D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York, 2011)
10. S.P. Marsh, *Formalising trust as a computational concept*. PhD thesis, University of Stirling, 1994
11. D.H. McKnight, N.L. Chervany, What trust means in e-commerce customer relationships: an interdisciplinary conceptual typology. *Int. J. Electron. Commer.* **6**, 35–59 (2001)
12. A.N. Meltzoff, ‘Like me’: a foundation for social cognition. *Dev. Sci.* **10**(1), 126–134 (2007)
13. A.N. Meltzoff, R. Brooks, Self-experience as a mechanism for learning about others: a training study in social cognition. *Dev. Psychol.* **44**(5), 1257–1265 (2008)
14. M.C. Moldoveanu, J.A.C. Baum, “I think you think i think you’re lying”: the interactive epistemology of trust in social networks. *Manag. Sci.* **57**(2), 393–412 (2011)
15. H. Nissenbaum, Will security enhance trust online, or supplant it? in *Trust and Distrust in Organizations* ed. by R.M. Kramer, K.S. Cook. Russell Sage Foundation Series on Trust (Russell Sage Foundation, New York, 2004) pp. 155–188
16. D.M. Rousseau, S.B. Sitkin, R.S. Burt, C. Camerer, Not so different after all: a cross-discipline view of trust. *Acad. Manag. Rev.* **23**, 393–404 (1998)
17. F.D. Schoorman, R.C. Mayer, J.H. Davis, An integrative model of organizational trust: past, present and future. *Acad. Manag. Rev.* **32**(2), 344–354 (2007)
18. M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**(5), 675–91; discussion 691–735 (2005)
19. A.R. Wagner, R.C. Arkin, Recognizing situations that demand trust, in *International Symposium on Robots and Human Interactive Communications*, IEEE RO-MAN, Atlanta, 2011



<http://www.springer.com/978-1-4614-7030-4>

Modeling Trust Context in Networks

Adali, S.

2013, VI, 83 p. 13 illus., Softcover

ISBN: 978-1-4614-7030-4