universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Interpersonal Trust: The Pivotal Role of the Relationship between Trustor and Trustee“

verfasst von / submitted by

Johanna Rosenberger, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 941
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Philosophie
Betreut von / Supervisor:	Univ.-Prof. Max Kölbel, MPhil MA PhD

Plagiatsverzicht

Hiermit erkläre ich, die vorgelegte Arbeit selbständig verfasst und ausschließlich die angegebenen Quellen und Hilfsmittel benutzt zu haben. Alle wörtlich oder dem Sinn nach aus anderen Werken entnommenen Textpassagen und Gedankengänge sind durch genaue Angabe der Quelle in Form von Anmerkungen bzw. In-Text-Zitationen ausgewiesen. Dies gilt auch für Quellen aus dem Internet, bei denen zusätzlich URL und Zugriffsdatum angeführt sind. Mir ist bekannt, dass jeder Fall von Plagiat zur Nicht-Bewertung der gesamten Lehrveranstaltung führt und der Studienprogrammleitung gemeldet werden muss. Ferner versichere ich, diese Arbeit nicht bereits andernorts zur Beurteilung vorgelegt zu haben.

Wien, 10.01.2022

Danksagung

Weder kann, noch möchte ich diese Arbeit einreichen und meine Studienjahre abschließen, ohne einigen Personen meinen Dank auszusprechen. Dieser Dank gebührt nicht nur denjenigen Personen, welche mich durch mein Studium begleitet haben, sondern auch denjenigen, welche es mir ermöglicht und erleichtert haben.

Zuallererst möchte ich mich bei meinem Diskussionspartner, Korrekturleser, Mentor, Kollegen und besten Freund bedanken. Danke, Clemens, für all die unzähligen Stunden voller Hilfe, Kritik und Ermutigung über die letzten acht Jahre hinweg – von der ersten bis zur letzten Prüfung. Ohne dich wäre ich weder akademisch noch persönlich da, wo ich nun bin.

Weiter gilt mein Dank meinen Eltern, Gerlinde und Gerhard, meiner Schwester Theresa und meiner weiteren Familie für die Begleitung und Unterstützung während meines Studiums. Auch ohne eure Hilfe wäre mir mein Studienabschluss nicht möglich gewesen.

Ebenso möchte ich meinem Partner Christoph für die ständigen lieben Worte und Kraft gebenden Umarmungen danken. Sie haben mir meine letzten Studienjahre – und besonders das Verfassen dieser Arbeit während einer globalen Pandemie – sehr erleichtert.

Natürlich gebührt dieser Dank auch meinen Freundinnen und Freunden. Allen voran Victoria, mit der ich seit einer Dekade nicht nur mein Privat- sondern auch Berufsleben teile. Danke für all dein Verständnis, deine Ratschläge und deine Rücksicht. Ebenso möchte ich Emanuela, meiner Mitbewohnerin, für die Begleitung durch diese Etappe meines Lebens und die vielen ausgleichenden Crazy Dance Minutes danken. Ein großes Danke auch an alle weiteren geschätzten Mitglieder der 1. EG, ihr seid die Besten. Ebenso will ich Carina, Barbara und Eva erwähnen: Auch ihr habt mit all eurer Hilfe und euren Gesprächen einen großen Beitrag geleistet.

Zuletzt möchte ich meinem Betreuer Max Kölbel den größten Dank aussprechen. Ohne seine Geduld, Zeit und Expertise wäre diese Arbeit nicht zu dem geworden, was sie ist.

Table of Contents

I. Introduction

Trust is everywhere – or at least it seems to be. Trust is not only believing your partner to be faithful but walking in the city streets without fearing being attacked. Trust is lending your colleague your toolbox, accepting the help of a stranger or going to the doctor thinking that you will be taken care of. Trust may be an attitude that is implicit in our actions or explicitly expressed. I can show you that I trust you when I hand you the keys to my flat and I can also tell you that I trust you with words. This makes it apparent that our lives are interwoven with trust. Once you start thinking about the importance of trust in social situations of any type, you cannot unsee that trusting is an essential component of our lives. Still, the amount of philosophical interaction with the phenomenon of trust does not measure up to the role it plays in our world. Although trust is pivotal to not only relationships but also whole societies and despite its ubiquity in our everyday lives, the philosophical research and especially the conceptual analysis of trust is only getting started. Current theories of trust do not yet offer convincing definitions of trust which adequately represent the vast variety of trusting situations. Therefore, I want to add a contribution to the discussion about what trust actually *is*. Not only do I want to suggest a novel understanding of the concept, but also show why many of the existing theories fail to offer a satisfying definition of trust. I aim to help close the gap between the practical and the theoretical side of trust. By that, I mean offering an approach that does justice to the broad phenomenon of trust in our everyday lives, but doing so in a comprehensible, argumentative way which will lead to definitions of trust that accurately depict all instances of trust.

Put broadly, this work is thus concerned with an answer to the question of what trust is. Narrowed down, the research question, that this thesis follows, is: *What are the limitations of current theories of interpersonal trust and how can its definition be improved?*

This is worthwhile to the philosophical research because the existing theories about interpersonal trust are flawed. There are certain shortfalls and unclarities in the literature on what trust is. One main problem is that there is no consensus about what the conditions for an attitude are to be identified as a trusting attitude. This becomes apparent when looking at the theories' susceptibility to counterexamples. Many accounts offer definitions that are constructed in a restrictive manner and do not acknowledge the variety of the phenomenon of trust. Therefore, these definitions are flawed because they exclude some instances of trust. Because of that, critics can easily create counterexamples. To solve this issue, I want to offer an understanding of interpersonal trust, which appreciates the broad

spectrum of trusting situations without committing the error of being too permissive and thus failing to sufficiently distinguish the concept of trust from other, closely related phenomena like reliance.

Furthermore, what many existing theories of trust do not acknowledge either is the pivotal role of the relationship between trustor and trustee. While some of the literature highlights the emotional component of trust, it does neither conclude that some trusting attitudes are more affective than others, nor that it is the relationship between trustor and trustee that affects the way they trust each other. My approach shows that this is the case and subsequently implements these findings.

Next, I want to give the reader a walkthrough of how the question that leads this project, will be tackled. To give a precise answer to this thesis' research question, I start with a response to the first part of the question about the limitations of current theories. Then, I present how the definition of trust can be improved. Nevertheless, before diving into this endeavor, I offer a short formal introduction to the topic of trust employing an analysis of paradigmatic examples.

In Chapter 1 I outline the theoretical foundations. Since I am planning to explore trust on an intricate level, I need the reader to be familiar with the basics of the topic. Thus, I start with six paradigmatic examples of trust to highlight both how many-faceted instances of interpersonal trust are and what the similarities are. Furthermore, by providing an analysis of the commonalities of those examples, I can offer the reader a first, uncontroversial approximation to a definition of trust. This comprises conditions about who the subjects of a trusting relationship are, what the object of trust is, which role risk plays and what happens if trust is broken. Furthermore, this chapter introduces the terms mental attitude of trust, reasons for trusting and rationality of trust, as well as discusses what the value of trust is.

Then, in Chapter 2, I examine the current research and highlight certain limitations, thereby giving reasons for why it is important to further investigate the concept of trust. I introduce the important distinction between cognitive and non-cognitive theories of trust and offer a short summary as well as a critique of three theories from both these camps.

Next, Chapter 3 offers a list of desiderata for a theory of trust and shows how the three theories of concern cannot meet them. Subsequently, my own account is outlined, and the second part of the research question is answered: If the current theories are flawed, how can the definition be improved? I advocate for an understanding, that acknowledges the variety of trust and does not conceive of trust as a uniform phenomenon. In my view, trust should be considered a polysemous expression which refers to three different forms of trust.

Finally, in Chapter 4, these three forms of trust are explained in detail. The main takeaway here is, that the specific relationship between trustor and trustee determines which kind of trust is at play. I claim that there is trust in strangers, trust in acquaintances and trust in friends and family – and that all of them are different attitudes and should thus, be defined differently. Therefore, further in this chapter, all three forms of trust will be distinguished and defined. Also, I show how my account can meet the demands of the desiderata, which the conventional theories from Chapter 2 could not fulfil. Lastly, I highlight the advantages that my work can offer for the current state of philosophical research before offering a final precise conclusion of the thesis.

What is important to me to note at this point is that this work should not be considered a complete theory of trust. First and foremost, this thesis is only concerned with interpersonal trust. Thus, the scope of my work is narrowed: It does not include claims about other types of trust such as self-trust, collective trust or trust in institutions. Furthermore, the three forms of trust I describe are preliminary. I do not want to preclude the possibility of further forms of interpersonal trust, that might be beneficial to consider. I am happy to accept another form of trust if there are good reasons that trustee and trustor may be in a kind of relationship that is different from those that I identified. Lastly, the definitions I give for the three forms of trust might be worthy of refinement. Despite these limitations of my thesis, I stand by my main claims: a. that the definitions provided by current theories of trust are flawed because they are susceptible to counterexamples and do not highlight the importance of the relationship between trustor and trustee, b. that trust should not be considered a uniform phenomenon and c. that it is the specific relationship between trustor and trustee that affects how the attitude of trust should be defined.

The argumentation of these claims is enough for my work to be a valuable contribution the philosophical research on the concept of trust. It is a novel understanding of the phenomenon that highlights the plurality of trust and puts emphasis on the affective component of trust: the personal relationship between two persons. Thereby it connects two ideas, that have been addressed but not necessarily developed within the philosophical literature.

II. Main Body

1. What Is Trust?

1.a. Paradigmatic Examples

This thesis is concerned with the phenomenon of trust – more precisely with what is known as 'interpersonal trust'. Interpersonal trust is the most commonly discussed variant of trust and depicts the trusting relationship between two persons: the trustor and the trustee. However, interpersonal trust does not yet exhaust the concept of trust - there are other variants of trust which are highly interesting and plentily debated in philosophical literature as well: examples being trust in institutions, collective trust, self-trust and distrust. Although a discussion of all these variants of trusting has its merits, this thesis will concentrate exclusively on interpersonal trust, including one person, the trustor, trusting another person, the trustee[1]. Thereby I will follow the vast majority of the current philosophical research on trust, but especially those publications, whose contents will be central to my claims: Hieronymi (2008), Holton (1994), Lahno (2001) as well as Hawley (2014), Frost-Arnold (2014) and Jones (1996). To illuminate characteristic cases of this kind of trust, I will introduce the research field by presenting six paradigmatic examples:

Library: Anna is situated in a library with her books, phone, and personal computer to study. Next to her is a stranger. When she gets up to go to the toilet, she asks the stranger to watch her belongings for a moment. There are no signs that this person is trustworthy, and she does not know anything about them. What she does know is, that she is putting herself at risk by doing so. However, she still puts her trust in this person and leaves.

Shoplifting: Resa and Sarah are in a store for electronic devices. Resa leaves Sarah in the section for mobile phones and continues looking at dishwashers. When they eventually leave the store, an alarm goes off. Sarah is accused of stealing a phone by a security guard. The evidence speaks against Sarah: There is a phone in her purse. Luckily, the girls are dismissed with a warning. On the way home, Sarah claims that she did not try to steal nor put the phone in her purse – she must have been framed. Since they have been friends for years and Resa knows Sarah to be an honest person, Resa trusts her to be telling the truth.

[1] From now on, 'trust' and 'interpersonal trust' will thus be used interchangeably.

Letter: Paul has been working for a company for a couple of months. He has to go on a work trip spontaneously and asks his colleague Laila to drop off a very important, personal letter at the post office for him because he has to leave in a hurry. He has known Laila for a while and knows that she is a precise and diligent worker and person. Since he knows that she has never missed an appointment, always calls back and returns every penny she borrows, he trusts her to bring the letter to the post office and leaves for his trip.

Climbing: Felix decides to try rock climbing. There are only so many routes he can climb without somebody securing him with a rope. The guy next to him asks him whether he wants to try a higher and more difficult route, which would require him to secure Felix. Felix has not seen this person securing anybody else since both of them are at the walls alone. Despite not knowing whether the stranger is able to handle a security rope, he decides to trust the stranger. He allows him to secure him and climbs the wall – up to a height, from which a fall could end his life if he was not secured properly.

Key: Hadis has a friend staying at her apartment for some nights. Unfortunately, she cannot be there herself to hand over the key to her friend when they arrive because she will be at a gala. Her friend says that he could pick up the key from Hadis' university if she left it there with someone. Because of that, she asks Eva, an acquaintance from her research group who always stays late at their office, whether she could take the key and give it to her friend when he arrives. Hadis has relied on Eva's help with research projects a couple of times already and has never been disappointed. Hence, she trusts Eva to not lose the key, forget about the friend or even worse, rob her apartment.

Assault: Kian is in a pickle. His twin brother Armin is accused of being part of a group that assaulted an old man. Armin has been distant over the last weeks and often returned home later than regularly. However, Kian assumed that his brother just had a new girlfriend and that would be the reason for his recent absence. Kian knows that Armin was definitely not at home when the assault took place. Furthermore, the police's evidence includes blurry footage of a person who could be his brother. However, Armin claims that it was not him and that the police are mistaken. Kian knows his twin inside out and trusts him more than anyone else. Although the person in the video could easily be his brother, he believes Armin's words over the police's footage.

1.b. Commonalities

These examples do not only give the reader an initial idea of various trust situations, but I will also refer back to them in the course of this thesis. This is, because I am of the opinion, that any complete theory of trust should account for all these paradigmatic trusting situations and must not exclude any of them. However, in this first chapter, I will use them to introduce what seem to be essential components of each situation that involves the attitude of interpersonal trust. Mere examples are not enough for a theoretical understanding of a certain concept, hence the need for a more sophisticated approach in which I want to make sense of the nature of trust. Still, this analysis of trust is no particular theory that I am introducing. It is a first approximation to the phenomenon and definition. By listing the commonalities of the paradigmatic examples, I am introducing the reader to those aspects of interpersonal trust that are uncontroversial. Or in other words: those characteristics that even competing theories agree on.

Readers might ask themselves whether discussions like this are needed. I admit that naturally, one is familiar with the phenomenon that is addressed when reading the words 'trust' or 'trusting'. First of all, this is because trust is ubiquitous – or, as Acton puts it: Trust is like cement that holds our society together (Acton 1974). We all trust, all the time, in a way. One person somewhat more and another somewhat less. Still, we usually dare to walk in the streets, because we trust other people not to draw a gun and shoot us. We trust the supermarket staff not to hand us back fake bills. We trust the tea company not to put psychedelics into their teabags. Secondly, we know what trust is because we know what it means to do so: In some way knowing that another person will do what I expect them to do and thereby not betray me. Yet, it is important to gain a deeper, more theoretical understanding of the concept in order to engage in a philosophical discussion and also to follow the line of my argumentation. Therefore, let me start with these first formalities of trust.

First of all, one circumstance that is observable in each of the paradigmatic examples as well as in a vast part of the literature about trust, is that the attitude of trust is usually considered a *three-place relation*. Since this consideration allows for the most fruitful discussion, I am going to follow e.g. Annette Baier (1986), Karen Jones (1996) or Pamela Hieronymi (2008) in treating trust as a three-place relation. Before I will explore what it means for trust to be a three-place-relation, I will quickly introduce but also preclude trust as one- and two-place relation.

Trust might be generally considered a one-place relation: A (the trustor) trusts. However, the attitude of trust described here is not directed towards another subject and does not involve a scope or object to say what it is, that the trustor puts their trust in. Thus, an analysis of trust as one-place relation is not only less in-depth than an analysis of trust as two- or three-place relation, but also not suited for researching interpersonal trust, since a particular trustee is omitted. Hence, a discussion of trust as a one-place relation is merely about what it means to have a trusting nature.

Considering trust as two-place relation on the other hand does involve a trustor as well as a trustee: A (the trustor) trusts B (the trustee). In this scenario, trust describes a trusting attitude by A towards B, without any particular action X to execute. In ordinary life, we use this a lot. You might ask your best friend about their partner: 'Do you trust him?' instead of 'Do you trust him to do this or that?'. This gives rise to say that trust as two-place relation is more of an elliptical phrase, within which we simply do not vocalize the action X.

I, however, will treat interpersonal trust as three-place relation: A (the trustor) trusts B (the trustee) to do X (the object of the trusting relation). Usually, X is considered to be an action. A, the trustor, trusts B, the trustee, to do something: giving them a lift, bringing them groceries, telling them the truth. Referring back to the paradigmatic examples above, one can easily see that each one involves a person A who trusts another person B to execute some particular action X. All apparent two-place trust claims will be treated as elliptical descriptions of three-place cases.

For now, it has been made clear that treating interpersonal trust as three-place relation always involves a trustor A and a trustee B (with the trustor trusting the trustee to do action X). The next issue I will tackle here is, what the conditions are for somebody to be a trustor or a trustee. Keep in mind, that the aim here is not to suggest a new theory of trust but to offer the reader certain know-how to understand the theoretical landscape of trust.

In each of the paradigmatic examples, there is a *trustor*. Being a trustor is characterized by certain conditions: First of all, trusting features a certain expectation towards the trustee. The trustor does not only hope for the trustee to do the desired action X but expects them to do so. Furthermore, trusting involves a certain degree of *risk* for the trustor: While they expect the trustee to act, there is no guarantee that they come through. This risk is often said to be connected to an initial dependency on the trustor's side (e.g. cf. Faulkner 2007). Let us take another look at *Climbing* to make this claim more transparent: Felix depends on a stranger to do something for him, namely securing him when he climbs up the wall. Since Felix cannot know whether the stranger will secure him, he takes a risk when

accepting the offer. Furthermore, if the trustee was not to carry out the desired action and thereby break the trustor's trust, the trustor is entitled to a certain *reactive attitude*, which is often associated with the feeling of betrayal or resentment. It goes without saying that Felix would be justified in feeling betrayed by the stranger if he let him fall to the ground.
To sum up this first quick introduction to a theoretical approach to explaining the characteristics of a trustor, I will highlight them once more: On the trustor's side, the attitude of interpersonal trust involves expectation, dependency, and risk as well as the entitlement to a certain reactive attitude if the trust were to be broken.
Moving on, there is the counterpart to the trustor, which is part of each two- or three-place trusting relation: the *trustee*. Being a trustee is characterized by being trustworthy. To be considered trustworthy, the trustee must be motivated to fulfil the desired action X. Different theories of trust have different ways of explaining where this motivation stems from – maybe because the trustee is a good person, maybe because they know about the trustor's helplessness, or because they have the trustor's interest in mind – yet they do agree that motivation is there. (Were the trustee to lack motivation we would not talk about an instance of a genuine trust situation.) However, being motivated is not enough. Even if the stranger at the climbing wall wanted to help Felix, his motivation would be void if he did not know how to handle a security rope. Thus, the conditions on trustworthiness on part of the trustee are both *motivation* and *competence*. "Although both the competence and motivational elements of trustworthiness are crucial, the exact nature of the latter is unclear" (McLeod 2020: 7). Some authors will use the terms skill or knowledge, some will use the term commitment, some will include a notion of honesty – but for my goal here, which is covering trust, trustworthiness, trustor and trustee to show the commonalities of the introductory examples, this broad understanding is sufficient. Each of the examples revolves around a situation in which a person A, the trustor, puts their trust into another person B, the trustee, who is regarded as trustworthy. In each scenario, the trustor trusts the trustee by expecting them to be trustworthy – which means being motivated and competent enough to fulfil the desired action. In each case, there is no guarantee or promise, that the trustee will come through. Thus, there is risk involved on the trustor's side, which is peculiar to the attitude of trust (cf. Lahno 2001: 171). However, were the trustee to break the trust, the trustor would be entitled to feel a certain kind of reactive attitude, like disappointment, a feeling of betrayal or resentment.
I have now clarified two parts of the three-part relation that is trust, namely person A, the trustor, who trusts person B, the trustee. What is missing is the *object of trust*: what A trusts

B about. In the paradigm examples, these 'objects' differ: Anna trusts the person in the library to watch her belongings, Resa trusts Sarah to be truthful about not stealing the phone, Paul trusts Laila to drop off his letter etc. These objects are exactly those actions that the trustor expects the trustee to perform and the ones that the trustee needs to be motivated and skilled enough to execute.

1.c. Rational Trust

Since this thesis is concerned not only with trust simpliciter, but also with reasons for trust and rational trust, it is relevant to introduce on the one hand, what rationality of trust is as well as on the other hand, why the attitude of trust should be rational. Although my work will not include strong claims about the rationality of trust, it is of significance to address it for two reasons: firstly, one can observe certain conflicts in the realm of rationality of trust. One of these problems, a tension between rationality and trust as cognitive attitude, as presented by Leo Townsend and Jeremy Wanderer (2013)[2], is highly relevant to the remainder of this thesis. Since the origin of this tension is the way the attitude of trust is defined, it is important to my endeavor. Secondly, I will offer a novel approach to understanding trust in the latter part of this thesis. Naturally, this comes with implications for how rational trust is understood. Although I will only sketch these implications, it is beneficial to the reader to have some knowledge about the rationality of trust in order to follow them.

Therefore, I want to take the time in this section to offer an introduction to the rationality of trust. The strategy for this subchapter is the following: firstly, there will be a section about what rational trust is. Secondly, I will discuss why trust is valuable and how rationality of trust is connected to this and lastly, the focus will be on certain doubts about whether rational trust can ever be achieved.

Thus, let me start with an examination of rational trust. This will include a distinction between theoretical and practical rationality as well as between truth- and end-directed rationality of trust. Furthermore, I want to state how this difference is connected to the mental attitude of trust. However, I will not introduce distinct theories of trust yet, since this is the program of the next chapter. For now, I want the reader to understand how

[2] This problem will be examined thoroughly later on.

rationality of trust can be approached, as well as why this is both valuable and desirable but also difficult and dubious.

1.c.i. What Is Rational Trust?

First and foremost, I am following the idea that rational trust is trust that is supported by good reasons. To answer the question of whether one instance of trust is rational, it is crucial to assess whether this instance is supported by good reasons.

Before trying to explore the rationality of trust more thoroughly, it is reasonable to explore the concept of rationality on its own first. Ordinarily, rationality is divided into two realms: The practical and the theoretical. Practical rationality concerns the rationality of an action, whereas theoretical rationality is concerned with what it means for a belief to be rational.

Within the domain of practical rationality, an agent's action is considered rational if it is the right means to an end the agent wants to achieve. Acting is thus rational if it involves reasonable decision-making when trying to reach a specific goal. This decision is based on the evidence available to the agent. Within the domain of theoretical rationality, a belief is considered rational if there are reasons that make its truth sufficiently likely. Put simply, this means that we only want to accept those beliefs which are (likely to be) true. To gather information about the validity of our beliefs, we resort to what Robert Audi (2004) calls the essential or classical basic sources of theoretically rational cognition. That is: perception, memory, consciousness, reason and testimony. These are the supporting items that give evidence for the truth of our beliefs (Audi 2004, Baker 1987).

Now let us quickly apply these ideas to rationality of trust. The findings about rationality suggest that the phenomenon of trust may be considered rational in two different ways: practically rational or theoretically rational. Philosophers who identify trust as an action consider practical rationality as the desired object of investigation (e.g., cf. Dormandy 2020, Castelfranchi/Falcone 2020). So, if you consider trust to be an action then trust is rational if trust as action is rational. Note that this implies shifting to an evaluation by the standards of practical rationality. Therefore, what renders trust rational is not true belief, but the action being "best means available to the agent, at least as far as he can tell on the basis of his evidence, for achieving an end that he values" (Dormandy 2020: 10).

Philosophers who consider trust to be a cognitive attitude identify trust as belief – often as belief about someone's trustworthiness. Therefore, they usually engage in an examination of theoretical rationality. However, and this is difficult to argue for, trust may be identified

as belief although the evaluation of its rationality is not a theoretical one. Trust is then still considered a belief, but there are other reasons which justify this belief as true. Rationality is in this case not directed at truth, but something else. What I am referring to is end-directed rationality, as McLeod (2020) calls it, or strategic rationality, according to Baker (1987). If the cognitive attitude of trust was rationally evaluated in an end-directed manner, then reasons which show the attitude useful for example are accepted as well. On the other hand, to reiterate, if the attitude of trust was rationally evaluated in a truth-directed manner, then only reasons which support the truth of the belief are accepted. Thus, the reader should keep in mind that different accounts of trust come with different accounts of what rational trust is.

Let me quickly take stock: What I have stated so far is that trust may be rational in different ways: theoretically or practically rational, as well as in a truth-directed or end-directed manner. An important question is: Which standard of rationality is the most useful one? To figure out the answer to this question, we need a clear-cut definition of trust. Is trust an action? Is it a belief? Is it something else? It is crucial to know what trust is in order to know which evaluation of rationality is applicable. Thus, an investigation of the rationality of trust always depends on how trust is defined. Since this thesis' research revolves around the definition of trust, it is important to understand how it is connected to the analysis of the rationality of trust.

The important connection is between the mental attitude of trust and rationality. Without knowing what the exact (mental) attitude of trust is, it is impossible to come up with conditions to judge whether an instance of trust is rational. Trust is then considered rational if the mental attitude of trust is rational (cf. Nolfi 2015). So, if trust – as commonly suggested – is a belief about someone's trustworthiness, then it is rational if the belief is rational. You could analogously say that trust is rational, if the emotion, stance or decision to trust is rational.

What is crucial at this point is to understand the following: how trust's mental state is defined influences how its rationality is understood. However, since trust is such a complex phenomenon, it has been a very difficult endeavor to argue for trust as one particular mental attitude. To show this I will shortly introduce a prominent conflict about whether or not trust's mental state may be identified as belief.

The mental state of belief is involuntary. One cannot voluntarily choose beliefs one might want to have. So, if trust was a belief, that would entail that one also cannot choose to adopt the attitude of trust. This seems reasonable: if I do not trust you I cannot make myself do

so. I might say – to borrow an example by Annette Baier – that 'I wish, I could trust you', but I cannot just decide to adopt a trusting attitude (Baier 1986). Hence, if trust was identified with the mental state of belief, it had to be considered strictly involuntarily.

Yet – and this seems to be intuitive as well for some – there seem to be instances in which I can decide to trust. Imagine the girl next to you in a lecture asks you whether she could borrow your handwritten notes until next week. You do not know her; you do not have much that could help you make an educated judgement about whether she is trustworthy or not. There are too many people in the lecture to know whether she is attending regularly and is probably coming again next week. There is only a very slim basis to form a belief about her trustworthiness. So the only thing is to take a shot in the dark. When you hand your notes to her, you are deciding that she is trustworthy (enough). Recall, that I also presented examples like this in the opening examples of this chapter: *Library* and *Climbing*. In both examples, the trustee is a stranger to the trustor and there is little to no evidence about their competence or motivation to fulfil the trustor's expectation. Thus, the trustor does not have the means to form a belief about their trustworthiness, but only a presumption or guess. It seems more fitting to say that in cases like these the trustor decides to trust the trustee based on the presumption that she is trustworthy instead of saying that the trustor trusts the trustee because she has the belief that the trustee is trustworthy.

What we are left with here is a conflict about whether trust should be considered a belief or not[3]. Furthermore, belief is not the only mental state that is an option – as already stated, it may also be considered an emotion, a stance, a disposition or even an action (cf. Lahno 2001, Holton 1994, Kappel 2014, Castelfranchi/Falcone 2020). This shows, that defining trust as one particular mental state proves to be difficult – and without consensus about what this phenomenon is, we are not able to analyze its rationality. In the beginning, I stated that trust is rational if it is supported by good reasons. To properly examine whether this is the case, we need a definition of trust. Exploring the options more closely and finding a definition is a pressing endeavor since an analysis of trust's rationality depends on it. That is why in Chapter 2 I will introduce the reader to different theories of trust to show exactly what difference the definition of trust's mental state entails and which difficulties arise from there. This investigation will lead us to another understanding of the phenomenon of trust, which may result in a novel apprehension of the rationality of trust. A complete

[3] The following chapters will go into more depth about specific positions in the conflict about trust being a cognitive, involuntary or non-cognitive, voluntary attitude.

analysis of these implications for an analysis of trust's rationality would go beyond the scope of this thesis, but it will show that there *are* implications and explain where they stem from. However, for the remainder of this chapter, I want to touch upon the value of rational trust as well as some doubts about whether an evaluation of trust even makes any sense.

1.c.ii. Why Trust?

In this subchapter, I would like to give some reasons for why an examination of trust and the rationality of trust is not only interesting but important.

As one might have gathered already, trusting someone is quite a risky endeavor. Not only does trust involve the risk of feeling let down by someone with who one thought one had a trusting relationship with, but often also putting an object (in a broad sense), which is cared for, at risk. May it be your car that you trusted your daughter with, may it be your dog which you trusted the neighbor with or may it be a secret, which you told a friend. If there was a guarantee, that the trustee would always come through and do whatever they are trusted to do, then there would be no need for trusting in the first place. The possibility of trust being broken and getting betrayed is necessary for trust – and this is for sure, since being entitled to a certain reactive attitude is a necessary condition for being a trustor. In the foreword to the *Routledge Handbook of Trust and Philosophy* (2020), Maria Baghramian writes "[…] to trust, therefore, is to take a risk and to be susceptible to disappointment, if not outright betrayal" (Baghramian 2020: xix). Pamela Hieronymi states this claim similarly: "I will take vulnerability to betrayal as a kind of touchstone for trust" (Hieronymi 2008: 215).

Yet, we trust. This is because trust is valuable. Its benefits tend to come with trust that is justified, which is closely related to trust being rational. Trust is vastly important to a cooperative society. Were it not for trust, we could not live life as simply as we do: Asking for directions, letting our neighbors accept our packages, ordering food without doubting the cooks' good intentions. If both parties of any relationship are cooperative and can rely on each other, there is vast improvement for a society. One can observe this by looking at individuals with trusting issues: Their lives will not be as easygoing as others. They may be running less risk of getting betrayed; however, they are missing out on the advantages of being able to hand off some of the burdens of their lives to others (cf. Hawley 2012: 18ff.).

Another good associated with trust is the possibility of not only cooperative but meaningful, deep relationships with others. A more detailed section about the importance of trusting personal relationships for one's mental wellbeing would exceed the scope of this thesis. However, it seems intuitive to accept the claim that trust is a vital factor and valuable good in child-parent relationships or romantic long-term relationships for instance.

Lastly, I want to touch upon the topic of knowledge as something valuable that trusting can bring about. As already mentioned in the beginning, a big part of a person's knowledge is somehow dependent on the testimony of others, or more explicitly: it is dependent on trust in the testimony of others. If you did or could not trust others' testimony, you could never acquire the amount of knowledge a trusting person would have. Think about the time-consuming task of checking whether the proposition that 'Benjamin's grandmother lives in Amsterdam' is true instead of trusting him for the truth. Or even worse, imagine having to verify for yourself that the circumference of the earth is 40 075 kilometers instead of believing the scientists' testimonies. Therefore, knowledge-creation is both easier and more time-efficient for someone with a trusting attitude (cf. Faulkner 2011, McLeod 2020).

Hence, trust has its benefits. However, and this is important, these can only fully unfold whenever trust is rational – not every kind of trust will do. If you were trusting for the wrong reasons, your trust will not necessarily be beneficial in the ways just described – or as McLeod puts it: "[The goods of trust] tend to accompany justified trust, rather than any old trust" (McLeod 2020: 36).

1.c.iii. Can Trust Be Rational?

So, one should trust in order to achieve the goods which come with trust. But also, one should be careful with one's trust to not get betrayed. Only rational trust – trust that is supported by good reasons – is valuable and also minimizes the risk of trust being broken. However, there are certain reasons for doubting whether trust even is an attitude whose rationality should be evaluated. This brings me to the last question I want to discuss in this chapter: Can trust ever really be rational?

The doubts about trust's rationality originate from the fact, that risk and therefore vulnerability to betrayal is a necessary condition for trust. For a trustor to be at risk, in this context, means to not be sure whether the trustee will prove trustworthy and action X will be carried out. So, trusting involves risk; it involves putting at least some of your eggs into the trustee's basket. This risk is greater, the higher the chance of being betrayed is.

Therefore, if a trustor does not know whether the trustee is trustworthy (skilled and motivated), then there is greater risk involved.

The question I would like to pose at this point is the following: Can doing something risky even be rational? Can jumping from a rooftop only to get a cool photo be rational? Can betting money on the worst football team be rational?

If risk is considered irrational, would trust be the most rational then, if there was the least risk – or, in other words, if you could be 100% sure that the trustee will come through and there is no chance of getting betrayed? However, if you had that degree of certainty, would that even be trusting? It does not seem like it. In that case, you would not expect the trustee to do the desired action X but you would just know that it would happen.

However, knowing that X is going to happen and simultaneously risking that X might not happen, is contradictory. If you know, then you do not risk – but risk is so central to trusting, that e.g. Victoria McGeer even says, that trust is purer if the available evidence does not support it (cf. McGeer 2008). Consider the following example: In situation A, you are as sure as you can be, that a friend will do desired action X and thus you trust her to do it. In situation B, you are pretty sure, that a friend will not pull through, but still, you put your trust in her. In which situation do you trust more? (cf. Hawley 2012).

If you pick situation A, you are denying the importance of vulnerability to betrayal as part of the definiens of trust. If you pick situation B, you are saying that trust is not as rational as one would desire it to be because you are doing something risky. As we can see, there is some kind of conflict between trust and rationality.

Another, probably even more striking reason for the difficulty of defining rationality of trust is that trust is prone to running counter the evidence. This phenomenon is at the center of the discussions of many philosophers working on the rationality of trust (Baker 1987, Hawley 2014, Frost-Arnold 2014). Consider the paradigmatic example *Shoplifting*: When Sarah tells Resa that she did not steal the phone, Resa believes her – she trusts her for the truth, although all other evidence clearly speaks against her. If she were to be rational, she would have to consider the evidence: that Sarah did spend time alone in the phone section and that there was a phone in her handbag. Eventually, Resa would conclude that Sarah probably did steal. However, when Sarah tells Resa otherwise, Resa trusts Sarah's words more than the other evidence available – although this other evidence was stronger, in the sense that perception might be considered more persuasive than someone else's word. In cases like this, a friendship appears to bring with it a form of evidence (like personal history or former successful trusting situations) that outweighs other evidence, as convincing as it

may be. Thus, trust within personal relationships seems to compete with evidential reasons – which makes it difficult to find a way to declare trust as a rational attitude.

You can also construe examples like this without the protagonists being best friends: Imagine the political party someone has been voting for years getting into a huge scandal for stealing tax money. Counter the evidence the newspapers print, the voter might still trust the party and believe that there is for sure a better explanation, like them being framed for fraud by another party in order to lose voters (cf. Baker 1987).

Moreover, sometimes there are no evidential reasons to assume someone's trustworthiness but still, you trust. That means, it might occur without any rational consideration – for instance, when trusting the nurses in the emergency room to keep you safe. You do not know if they are trustworthy people, yet you trust them.

Again, the attitude of trust and the concept of rationality do not seem to be in accordance with each other. Yet, I do not want to get discouraged by these doubts about the rationality of trust and find a way to explain where these problems stem from and suggest an idea, that might solve them.

For now, I have laid the groundwork: I introduced the topics trust and rational trust, as well as its values and doubts. I explained why rationality of trust is desirable and important on the one hand, but also what difficulties there are on the other. The final aim of this thesis will be to provide a suggestion of how to think about trust which escapes these problems.

To pursue this objective, I will stick to the following agenda: First and foremost, to move forward, a theory of trust is needed. Right now, within philosophical research, there is no consensus on how to define trust. Most striking is the disagreement about whether trust should be considered a cognitive or non-cognitive attitude. To introduce the reader to this problem, I will analyze three main strands of thinking about trust in Chapter 2. I will show that there is no consensus about how trust should be defined within those theories as well as that they cannot account for all of the paradigmatic examples. In Chapter 3, I will first develop certain desiderata for an ideal theory of trust, then state why those theories cannot fulfil these desiderata and lastly, suggest a way of thinking about trust, that can do so. In Chapter 4, this account will be properly explained and refined. In the end, I want to have shown that my own understanding of trust has advantages over the other, 'conventional' theories. One of these advantages, which will be addressed in the final parts of this thesis, is that some of the tension between trust and rationality can be resolved.

2. Theories of Trust

In this chapter, I will take a closer look at the question of what it means for trust to be identified as cognitive or non-cognitive attitude and why these two camps, the cognitive and non-cognitive, are in conflict. This discussion is fruitful for understanding my claim that there is no consensus within the conceptual analyses of trust (and therefore neither about how trust's rationality is understood). To tackle this endeavor, I will give a short overview of what it means to recognize trust as either cognitive or non-cognitive attitude and introduce the implications this has for a theory of trust. I will work with three established theories, as found in the literature, that are useful for highlighting these problematic aspects of trust, namely those by Pamela Hieronymi (2008), Richard Holton (1994) and Bernd Lahno (2001, 2020). In addition to introducing the conflict, I want to add support to my claims by drawing on works by Townsend and Wanderer (2013), Katherine Hawley (2014) and Karen Frost-Arnold (2014), who agree that this is a field of research with certain problems to solve.

My argument in this chapter shall be as follows: first, I will differentiate between trust as cognitive and non-cognitive attitude. Next, I will give examples for theories of trust of both variants and evaluate their strengths and weaknesses. Then, I will show that these example theories do not capture each of the paradigmatic examples, as introduced in the first chapter and, thus, that they fail in accounting for all of them. From these findings, I will subsequently conclude that none of these theories can cover all situations that should be considered instances of trust. Therefore, the main aim of this chapter is to establish the claim that the prevailing theories of trust are incomplete and that they should be expanded and/or improved. Thereby I am laying the groundwork for presenting my own account about how trust should be understood instead in the following chapter.

2.a. Trust as Cognitive Attitude

To precisely state what trust as cognitive attitude implies, I will follow Townsend and Wanderer's (2013) definition. According to them, being an advocate of the cognitive camp involves commitment to two main conditions about trust:

i. “It must involve taking a stance on the truth content that can be captured in a that-clause[4] […]. Standardly, the cognitive attitude is that of belief or expectation, and the content of a ‘trusting belief’ (or expectation) is ordinarily thought to involve the trustworthiness of the person trusted.” (ibid.: 1)

ii. “[T]hat it is a cognitive attitude […] is somehow central to an adequate understanding of trust. So, not only must cases of trust involve a cognitive element, but this element must not be seen as merely the consequence of some other, more basic, element in an analysis of trust.” (ibid.: 1)

Townsend and Wanderer state, that if trust is considered a cognitive attitude, then the trustor believes, that the trustee is trustworthy (at least in regard to an action X), to be true. Furthermore, trust’s central element is the belief, that the trustee is trustworthy. This coincides with the definition of cognitivist Russell Hardin (2002), who claims that trusting simply equals believing that somebody is trustworthy. However, the content that the trustor believes, might not be ‘that the trustee is trustworthy’ but ‘that the trustee can be trusted to do action X’ (or simply ‘that the trustee will do action X’, which I understand as including that the trustor thinks of the trustor as trustworthy in that case). Similarly, Keren (2014) states that cognitivists agree that trust is a belief about the trustor’s trustworthiness, at least in the sense that the trustee will carry out action X. I suggest the following differentiation: a strong cognitivist position identifies trust with the belief, that the trustee is trustworthy. A weak cognitivist position identifies trust with the belief, that the trustee is trustworthy only in regard to one action X. In cases, in which the differentiation is of relevance, I will differentiate between those positions – if not, then ‘cognitivist’ henceforth refers to both of them.

Moving on, there is one certain merit to the definition of trust as cognitive attitude, which is especially interesting to my endeavor – namely, that it is not very demanding to define what rational trust is. If the mental attitude of trust is considered a belief, then trust is rational if this belief is rational, and it is straightforward to say for beliefs under what conditions they are rational (i.e. if they are supported by evidence etc.). As suggested in Chapter 1, with trust’s mental attitude as belief, we are typically dealing with theoretical, truth-directed rationality – which means, that a belief is rational, if it is true; or rather, if

[4] (such as, to come back to the paradigmatic examples for Chapter 1, ‘that Laila will drop off the letter’, ‘that the girl from the library watches my belongings’).

there are reasons (most commonly, evidential reasons) that make its truth sufficiently likely.

To conclude, according to the cognitivist camp, trust is identified as a belief about the trustee's trustworthiness (in a strong or weak sense). Subsequently, trust is rational if there are reasons that support this belief's truth. However, there are also profound difficulties when trust is considered a cognitive attitude. To demonstrate the merits and, more importantly, the problems of trust as cognitive attitude, I will now evaluate one particular cognitive account established by Pamela Hieronymi (2008).

2.a.i. Hieronymi: A Cognitive Theory of Trust

For Hiernonymi, genuine trusting necessarily involves the belief that action X will be carried out. Her theory of trust clearly states that the reasons of trust only concern the *belief* about the trustworthiness of the person in question (cf. Hieronymi 2008: 213). Her account is thus concerned with theoretical rationality, instead of practical rationality, since real, genuine trust is considered a cognitive attitude (namely, a belief) and not an action. Moreover, it is strictly truth-directed. If the reasons for trust are to achieve an end, for instance, considering usefulness (e.g. to remain calm in the emergency room) or 'goodness' (e.g. to maintain or build a relationship), they do not support your trusting belief. This trusting belief, solely linked to the trustee's trustworthiness, can only be supported by reasons which support its truth. Instrumental reasons, such as reasons that show trust useful, important, required etc. are not those for which one trusts – they rather show that one lacks confidence in someone's trustworthiness (cf. ibid.: 214). To exemplify what Hieronymi has in mind here, imagine this: your phone died and you need to ask for directions. The only person close is a man on the streets who does not appear to be trustworthy. However, because he is your only option and you need help, you ask him and follow his directions. If you trusted him only because it is useful for finding your way, did you really engage in genuine trust? Hieronymi says no – moreover, she states that you do not trust "to the extent that you *must* rely" (ibid.: 214). Thus, the more you are 'in need' of other reasons than those that support your trusting belief[5], the less you trust. However, these other, 'wrong' reasons for trust, might be reasons for acting. Apparent 'trust' which is motivated only by end-directed reasons, is not full-fledged trust, but only a kind of entrusting (cf. ibid.).

[5] I will use ‚trusting belief' and ‚the belief that the trustee is trustworthy' synonymously.

The main takeaway from Hieronymi's account is that real, full-fledged trust involves a belief about the trustee performing some action X: „[t]rust is full-fledged only if one believes that the person in question will do the thing in question" (ibid.: 227). This belief "can be supported only by reasons that bear on the truth of that content" (ibid.: 232). Every other kind of trust, which we engage in for reasons of end-directedness, is not.

This theory of trust certainly has its merits: It gives a clear-cut definition of what trust is and also, what it is *not*. There is a sharp distinction between trust and 'entrusting', and thereby also a differentiation from mere reliance. Thus, her definition successfully picks out full-fledged trust situations that do not represent entrusting or mere reliance. I want to reinforce here that I am not claiming, that Hieronymi's theory is picking out the right trust situations in general – only that, within her account, she successfully offers conditions that distinguish her account of trusting from non-trusting.

That said, I claim, that there are two major problems with Hieronymi's cognitive account, which will be explained in the following paragraphs. Firstly, there is reason to doubt whether it is too restrictive. This entails that this theory fails to account for some instances of trust. It wrongly excludes situations from counting as trust. Namely, those situations, in which we trust for end-directed reasons. Secondly, Hieronymi's account, as well as other cognitive theories have to account for a certain difficulty linked to the evaluation of rationality. This problem is concerned with the tension between trust and evidentialism.

I will now look for convincing reasons for both these claims: namely, that Hieronymi's account is too narrow to account for situations that should be considered instances of real trust. Then, I will also check whether the account falsely renders trust situations irrational that should be considered rational. If one of these faults is the case, then there are counterexamples to the theory, and it thus should be either revised or rejected.

I will begin by tackling the problem of the theory being too restrictive by presenting objections from the literature. The main criticism is that a theory of trust like Hieronymi's excludes instances of trust which are directed by reasons other than supporting the trusting belief as true. In other words: Cases of trusting for end-directed reasons are dismissed. This is more of a hindrance than it is an advantage to a theory of trust since people trust although the reasons for their trusting belief are end-directed or even without a trusting belief at all. To make this clearer I will borrow an example from Paul Faulkner (2007). This is a counterexample to Hieronymi's account: A reformer employs someone, who is convicted of theft. Not only is there no evidence that the convict is trustworthy, there even is evidence that he is *not* trustworthy. The reason for which the reformer employs the convict is to give

them a chance in society – a reason which does not support the truth of the reformer's trusting belief. Maybe the reformer does not even have the belief, that the convict is trustworthy – and *yet* they trust. On Hieronymi's account, the reformer does not genuinely trust the convict since either a) there is no trusting belief, or b) the reasons for trust are not supporting the truth of the trusting belief.

This is exactly my claim here: there are situations in which there is trust, although there is no reason that supports the truth of the trusting belief and trusting situations in which there is no trusting belief at all.

Support for my claim also comes from Judith Baker's (1987) work on trust, which emphasizes the importance of allowing for end-directed reasons as well. Baker states that belief – including a trusting belief – "aims at the truth" and that we "should only accept those beliefs which are likely to be true" (Baker 1987: 807). Thus, reasons for beliefs should be those that give evidence for their contents. However, and this is crucial: "[T]rust seems to involve beliefs which are not accepted on the basis of evidence" but are formed for end-directed reasons (ibid: 807). Moreover, Baker states that in some instances of trust we think it rational to stick to a belief although there is counterevidence (as portrayed in *Assault*). Therefore, trust should not be treated in the same way as beliefs. She suggests thinking of trust as a kind of commitment instead of solely a belief, in order to escape these problems.

In addition, also Karen Frost-Arnold (2014) argues that while trust is a cognitive attitude, it should be considered something broader than a belief alone. The reasoning for this claim is that there are more reasons to trust than epistemic ones, namely coping, therapeutical or corrective reasons. Since these non-epistemic reasons for trusting will play a grander role in the last chapter of this thesis, I will refrain from extending the examination of Frost-Arnold's work at this point.

However, it is important to remind the reader that both of these critiques do not object to trust being a cognitive attitude. At the least, there is mutual agreement that the 'main kind of trust' or 'conventional' trust should be considered a cognitive attitude. Therefore, I want to draw on more findings from the literature which *do* object to trust's attitude being of a cognitive kind. This is a more drastic disagreement since it does not only pertain to nuances of how the cognitive attitude should be conceived but to the attitude being cognitive in the first place. According to these non-cognitive theories, the concept of a trusting belief about the trustee's trustworthiness should be dismissed entirely. These critiques will be used later to back up my argument that, in the light of Hieronymi's account, some of the paradigmatic trust examples cannot be designated trust situations.

Although non-cognitive theories of trust will only be discussed more closely in the second half of this chapter, I will quickly sketch the idea now, since they represent an objection to Hieronymi as well. Non-cognitive accounts deny the premise that trust necessarily involves a belief of some sort, that the trustee is or will be trustworthy. If there is a cognitive component to the account, then it is not playing a central role or is itself grounded in some other non-cognitive element. The claim is that trust is more of an affective or emotional attitude or even a certain 'way of seeing the other', as opposed to a belief about their trustworthiness (cf. Jones 1996; Lahno 2001, 2020). Moreover, as introduced in Chapter 1, some arguments concern the voluntariness of trust: since one can construct arguments that trust can be willed and belief cannot, trust should not be identified with the mental state of belief (cf. Holton 1994).

This concludes my presentation of literature that supports my objection that Hieronymi's theory is too restrictive. Next, I want to tackle the second major difficulty with cognitive accounts such as Hieronymi's. This issue concerns the topic of rationality of trust and revolves around the problems between cognitive accounts and evidentialism[6].

This tension may be set out as resulting from the following contradiction, as introduced by Townsend and Wanderer (2013): cognitive theories of trust view trust as an essentially cognitive attitude. Furthermore, that means that the standard of rationality is theoretical rationality. Central to theoretical rationality is (some form of) evidentialism, which means that "a core part of what it means to be theoretically rational is that one's cognitive attitudes stand in the right sort of relationship to evidence" (Townsend/Wanderer 2013: 1). Thus, evidence or evidential reasons are central to trust, if considered a cognitive attitude. However, and this is the crux of the matter to Townsend and Wanderer, trust is in conflict with the norms governing evidentialism – and this conflict is not merely superficial.

> "Rather, it is claimed that the tension between trust and Evidentialism is constitutive: part of what it is to trust is that one's attitude of trust bears a relationship to the relevant evidential base that is explicitly prohibited by Evidentialism." (ibid.: 2)

Put into other words, this means that it is necessary for the attitude of trust to not depend on evidence since it must, by definition, involve some degree of risk. A trustor, who collects

[6] This conflict has already been addressed in Chapter 1.c., as the general tension between trust and rationality. Townsend and Wanderer's work, however, applies to cognitive accounts of trust and the role evidence plans in particular.

evidence to be sure that action X will occur, is what Townsend and Wanderer call "a lousy trustor" (ibid.: 1). Moreover, it should be noted that this problem is not the only difficulty an account of trust faces concerning evidentialism. Not only is there a need for this essential lack of evidence that a theory of trust must account for, but it also has to deal with the phenomenon that trust occasionally runs *counter* the evidence. This will be addressed in more detail throughout this thesis, but for now, it is enough to state that this is in obvious conflict with a cognitive theory of trust as well.

To sum up, the authors conclude that trust, if defined as cognitive attitude, cannot be considered a rationally appropriate attitude – which is highly problematic for advocates of the cognitive camp. This conflict between cognitivism and evidentialism, as presented by Townsend and Wanderer, will be highly relevant for my further analysis and is crucial to why I shall eventually suggest another way of understanding trust.

2.a.ii. Application to the Paradigmatic Examples

Now that I have made the problems that cognitive theories like Hieronymi's face clear, I will apply these insights to the paradigmatic examples of trust, as introduced at the beginning of Chapter 1. My demand for a theory of trust is that it can make sense of each of these paradigmatic examples. I want to show that Hieronymi's theory cannot meet this demand since it either excludes certain examples from being a proper trusting situation or has to render them irrational.

Let me start with both examples *Library* and *Climbing*. In these situations, the trustor has little to no evidence about whether the person they are about to trust, is trustworthy. Neither Anna, who leaves her personal belongings in the care of a stranger, nor Felix, who agrees to trust the guy at the climbing wall to secure him with a rope, knows whether the respective trustee deserves to be trusted. It could be that the girl at the library does not really care and leaves Anna's computer unattended as soon as she enters the bathroom. There is also a chance that the stranger at the climbing wall is not skilled at securing and is putting Felix in a very dangerous situation. Since there is no evidence to justify a trusting belief, the trustor cannot know whether trusting the trustee is reasonable. It is not possible to believe (for truth-directed reasons) that the trustees are trustworthy in regard to guarding Anna's belongings or securing Felix' rope. Remember that 'being trustworthy' entails a certain motivation and skill/knowledge. The trustors might assume that the trustees are motivated

and skilled to guard or secure and will come through – but they do not have evidence that supports this assumption's truth value.

Thus, on Hieronymi's account, Anna and Felix do not trust the trustees. Anna and Felix might rely or entrust but they do not trust. Were Hieronymi to allow for other than truth-directed reasons, then we could identify these two situations as trusting situations. Examples for end-directed reasons could be given the following way: Anna could have various end-directed reasons. The most intuitive one would be a reason of usefulness. If she trusts the girl, she does not have to pack all her things and bring them to the bathroom and she does not run risk that someone else takes her seat while she is gone. She may also trust because she wants to challenge herself to be more trusting altogether or maybe she wants to spark a positive feeling inside the girl. Felix, who trusts the stranger to help him, does so because he could achieve the end of trying a more difficult route. Alternatively, he might not want to come across as unfriendly or maybe he does not want to make the helping stranger feel as if he is not strong or skilled enough to secure him. To conclude, in both of these examples, the conditions for them to count as a trusting situation are not fulfilled according to Hieronymi. Anna and Felix are not trusting the respective trustees because of epistemic reasons. They neither have reasons to believe that the trustees are trustworthy, nor that the trustees will successfully perform action X. On the contrary, they trust for considerations that "show trust useful, valuable, important, or required" and these are not the reasons for which one genuinely trusts according to Hieronymi (Hieronymi 2008: 213). Thus, both examples *Library* and *Climbing* are excluded from the domain of real trust on this account. Next, let us take a look at the examples *Shoplifting* and *Assault*. In both situations, there is an underlying personal relationship between trustor and trustee. In *Shoplifting*, the protagonists Resa and Sarah are best friends; in *Assault* they are brothers. Thus, they share history, they know each other very well.

In *Shoplifting*, Resa trusts Sarah to tell the truth about not stealing the phone. Resa has a trusting belief, so she believes that Sarah can be trusted regarding the truth about the incident. Or, in other words, that she is motivated and skilled enough to do so. According to Hieronymi, a trusting belief can only be justified by reasons which support its truth. The most valuable reasons to support the truth of a belief are evidential reasons. However, it seems that Resa does not trust Sarah for (only) evidential reasons, but for reasons that Katherine Hawley (2014) calls reasons of friendship. These kinds of reasons are non-epistemic reasons and are therefore not meeting the requirements for Hieronymi's genuine trust. The evidence – that Sarah spent time alone in the phone section and that there was a

phone in her purse – speaks against Sarah and against trusting her to tell the truth. Resa trusts Sarah although there is obvious evidence that contradicts her trusting belief. To summarize, Resa does not only trust for reasons which support the truth of her trusting belief but for other reasons that show trust good or appropriate (f.e. continuation of their friendship or not upsetting Sarah). She prioritizes their friendship over the evidence at hand, that does not support the truth of her trusting belief.

The conditions in *Assault* are very similar, hence the analogous analysis: Kian trusts Armin's words over the police's evidence. Thus, he holds on to a belief although there is striking evidence that the belief might be false. This is a certain peculiarity of trust: it gives us "blinkered vision", as Jones says (Jones 1996: 12). If the trustor trusts somebody, and this is mostly the case in trusting situations between close friends or family, then one "will resist the truth of any evidence to contrary" (McLeod 2020: 15). Although Armin has no alibi and there is security footage, Kian holds on to his trusting belief. It is observable that trust is an attitude that is more emotional and less cognitive than Hieronymi wants to admit. By means of these examples, the tension between trust as cognitive attitude and evidence, as described by Townsend and Wanderer, becomes apparent. According to Hieronymi, both examples should be dismissed as rational trusting situations.

Lastly, I want to see how the remaining examples hold up on Hieronymi's cognitive account of trust. First, I want to take a look at *Letter*: in this situation, Paul trusts Laila to drop off his letter because he believes, that she is trustworthy. His trust is based on reasons, that support the truth of this belief (namely that she has proven trustworthy in different situations in the past) and also, there is no evidence, which contradicts the trusting belief. Paul does not trust Laila because there is no other option but because he is convinced that she is the best for the job. Hieronymi argues that one trusts the most if one relies the least (cf. Hieronymi 2008: 213). While in *Library* and *Climbing*, the trustors have to rely 'a lot' because they do not know the trustees, Paul does not rely: he is convinced that Laila is trustworthy. Thus, on Hieronymi's account, *Letter* seems to be a perfect example of genuine, rational trust.

In *Key*, Hadis trusts Eva to hand her key over to her friend. The example is similar to *Letter*: the trustor does not trust for any other reason of usefulness or necessity but simply because she believes that Eva is trustworthy and there are various evidential reasons which support the truth of this belief. Like *Letter*, *Key* is another example of what Hieronymi has in mind when it comes to rational trust.

Before moving on to the second part of this chapter, I would like to quickly sum up the findings about Pamela Hieronymi's theory of trust as a representative for cognitive theories of trust and what there is to take away from applying it to the paradigmatic examples of trust. According to Hieronymi, genuine trust is equivalent to the belief, that someone is trustworthy. The reasons for which a trustor engages in real trust are those, which support the truth of the trusting belief. Since in *Library* and *Climbing* the reasons for trusting are other reasons, such as reasons of usefulness or necessity, both situations should not be identified with a genuine trust. While *Shoplifting* and *Assault* could be identified with genuine trust, because the trustors do have a trusting belief, there is another problem: there are evidential reasons that do not support the truth, but falsehood of their trusting belief and yet they trust – although their trust runs counter the evidence. Therefore, these situations are no instances of genuine trust either, according to Hieronymi. Thus, all four examples should either be identified as irrational, or as no trusting situation at all.
Lastly, *Letter* and *Key* have proven to be instances of trust because they match Hieronymi's conditions. Thus, this cognitive theory of trust can only account for two out of six situations, which should be considered instances of rational trust. I will now move on to non-cognitive theories of trust and discuss what their application to the paradigmatic examples implies.

2.b. Trust as Non-Cognitive Attitude

If trust is defined as a non-cognitive attitude, then this means that either condition i. or condition ii. of the definition of trust as cognitive attitude is denied. Spelled out, this means that either trust is not considered a belief (or expectation) whose content concerns the trustworthiness of the trustee *or* said belief (or expectation) is not an essential or central element to trust. Examples of non-cognitive attitudes are an emotional or optimistic attitude, a way of seeing or a kind of stance the trustor takes towards the trustee. Yet, just like trust as cognitive attitude, trust as non-cognitive attitude also has some problematic implications. While the tension between trust and evidentialism, as explicated by Townsend and Wanderer, is dissolved if trust is not identified as belief, there are other difficulties a non-cognitive account has to account for. I will discuss both advantages and disadvantages of non-cognitive attitude more closely on the basis of two accounts, them being the participant-stance theory as suggested by Richard Holton (1994) and a theory about trust as emotional attitude by Bernd Lahno (2001, 2020).

The plan for this subchapter is to introduce both of these accounts, discuss their benefits and weaknesses and subsequently apply them to the paradigmatic examples as well, in order to see whether one of them meets my demand of providing a definition that is both necessary and sufficient for all of them.

2.b.i. Holton: Trust as Participant Stance

At the heart of Holton's theory of trust is the conclusion, that it is, in some cases, possible to decide to trust. Before explaining this conclusion, I want to give a general overview of the account. According to Holton, cognitive accounts of trust are misconstrued in their assumption that trust is a belief about something (e.g. a trusting belief like Hieronymi's or a belief about the trustee's goodwill, as Baier (1986) suggests). Rather, trust involves a "more general attitude", which entails seeing the trustor as a person (Holton 1994: 66). This attitude is a certain stance towards the trustee, which Holton adopts from Strawson (1974) and calls the participant stance. By means of this claim, it is easy for Holton to differentiate between trust and reliance: while a trusting attitude involves a participant stance, mere reliance does not. Thus, reliance is not necessarily directed at a person. This entails that one cannot trust, but merely rely on objects. Furthermore, it is also justified to react a certain way when trust is broken. A central point about Strawson's participant attitude is that it allows for reactive attitudes. Holton argues that the relevant reactive attitude of trust as participant stance is that of betrayal. If you do not adopt the stance, you are not entitled to the feeling of betrayal if something goes wrong. This explains why misplaced reliance is only followed by disappointment, not betrayal.
So, Holton turns against the tenets of cognitive theories of trust:

> "Trusting someone does not involve relying on them and having some belief about them: a belief, perhaps, that they are trustworthy. What it involves is relying on them to do something, and investing that reliance with a certain attitude. This is to take a practical stance." (Holton 1994: 67)

Moreover, he states, that there are no beliefs about the trustee's trustworthiness which will influence whether we adopt the participant stance. Rather, we can only reach these beliefs

when taking the stance – it leads us to form certain beliefs (cf. ibid.: 68)[7]. Since trusting does not equal believing but taking a stance, Holton argues that we can decide to do so without committing to doxastic voluntarism (cf. ibid.: 69ff.). What is meant by this term is the (controversial) view that it is possible to choose one's own beliefs.

Holton makes another relevant point regarding trusting at will: we need not forget that we are concerned with trust as three-place relation. While he agrees that one cannot choose two-place trusting, which is a trusting relationship simpliciter, he says that it is possible to decide to trust someone to carry out one particular action X (trust as three-place-relation). A trustor might decide to decide to trust a trustee even if "they have not yet proven their trustworthiness" (Holton 1994: 70). Thus, you can choose to trust someone to do an action X without evidential basis that they can be trusted in this regard. However, Holton also states that trust cannot be willed if it is not appropriate or justified. Thus, you cannot decide to trust without good reasons at all – but you certainly do not need evidence.

Before moving on to the critique of Holton's account, I will quickly list the benefits of Holton's non-cognitive theory. Firstly, it can account for a difference between trusting and mere reliance by introducing trust as participant stance. This is not to be regarded as an advantage over cognitive theories but as general benefit to a theory of trust. Secondly, and most importantly, it can explain trusting at will without committing to doxastic voluntarism, d.i. the possibility to choose beliefs voluntarily. Thirdly, it escapes the aforementioned tension between trust and rationality by moving away from a strictly cognitive theory that identifies trust with belief. Lastly, it is able to dismiss the dated idea that trusting is mere risk calculation and adds a personal component to the issue[8] (cf. Holton 1994).

The most prominent critique of Holton's account concerns the claim, that trust is a voluntary attitude. It is commonly said that trust cannot be invited nor willed (e.g. cf. Baier 1986, Hiernonymi 2008, Keren 2014). If there is doubt about the trustworthiness of a trustee that forces a trustor to actively decide to trust instead of them just trusting, then this should not be considered genuine trust but mere reliance. Baier for example says, that an invitation like "Trust me!" cannot be accepted at will, since there are no reasons that would justify

[7] Since beliefs only follow from trust but are not a necessary condition for trust, they are not essential to the attitude. Therefore, Holton's account is not a cognitive account in disguise. He clearly denies ii. of the definition of cognitive accounts.

[8] Keep in mind here that this is neither a critique of all cognitive theories nor of Hieronymi's theory, since not every cognitive theory is a mere risk assessment theory. However, it is said that cognitive theories are grounded in theories like these (cf. Lahno 2001). Holton, on the other hand, lets go of these cognitive roots.

the trusting belief (cf. Baier 1986: 244). Hieronymi would object that trust, which the trustor decides on for other reasons than those which show the trusting belief true, can never be full-fledged trust.

Before moving on and applying Holton's theory to the paradigmatic examples of trust, I will first introduce another appealing theory of trust as non-cognitive attitude. Then, I will investigate whether one of them can account for all paradigmatic trust situations from Chapter 1.

2.b.ii. Lahno: Trust as Emotional Attitude

Lahno's (2001) theory of trust originates in a critique of merely cognitive accounts, according to which "trust is nothing but a rational belief about the character or the probable behavior of another person" (Lahno 2001: 172). Lahno states that trust should not be approached with a game-theoretic mindset in which the essence of trust is nothing more than calculated risk[9]. Risk is a central component of trust and thus should not be minimized or disposed. Moreover, since calculating risk contradicts real trust, its definition should be approached in a completely different way. Thus, Lahno states that trust is "not subject to direct rational control" and is thereby similar to an emotion (ibid.: 172). He admits that emotions are not independent of beliefs and reasons and that there is a certain connection between trust and belief – however, belief is not essential to the attitude. In this respect, Holton and Lahno agree with each other. Lahno is drawing on Karen Jones' (1996) theory about trust as an affective attitude and Ronald de Sousa's (1987) influential work about the nature and rationality of emotions, in order to develop his account. So, for Lahno, trust is essentially non-cognitive and similar to emotion, but still, there is a cognitive aspect to trust. This is, because emotions can determine how the world is perceived, what we think, and which judgements we make.

> "Trust is necessarily tied to a particular perception of the world or some part of the world. It may be characterized by certain patterns in the way the world is represented in thought and the way certain contents of thought are associated with each other. There is a causal relationship between trust and belief. Yet, because trust somehow determines how we think, it cannot be understood as the immediate result of rational consideration." (ibid.: 177)

[9] To get to know game-theoretic approaches to trust, I refer the reader to Williams (1988), or Gambetta (1988) in general.

By identifying trust with an emotional, not a cognitive attitude, Lahno's theory has major advantages. Since trusting is not rational calculation of risk, he can carve out a difference between trust and reliance. This difference concerns the relationship between trustor and trustee, which is sufficient reason for the trustee to be motivated to perform X. The relationship is not only reason to expect the trustee to behave a certain way, but also to feel like one is obliged to do so. This is also due to a certain feeling of 'connectedness', that is best described as the trustor's perception of common interests and shared norms, "which the trustor himself takes to be authoritative" (Lahno 2020: 152). Thus, trust as emotional attitude involves a certain stance or way of seeing each other, which a merely cognitive attitude certainly neglects. While reliance is then a simple expectation without a normative dimension, trusting as an emotional attitude goes beyond mere reliance (Lahno 2001, 2020). Lahno (2001) gives the example of a burglar who expects someone not to be in their house at night because he studied the owner's routine. The burglar only relies, he only calculates the risk; but does not trust the house owner not to be there. In conclusion, what is central to trusting on Lahno's account is the focus on the personal relationship, the emotional engagement and also the obligations that derive from it.

> „Any comprehensive and general attempt to define genuine trust as a special sort of reliance must at least in part refer to the way people treat each other when trusting, how they relate to each other, how they perceive each other when choosing how to act." (Lahno 2020: 151)

Now that Lahno's account of trust as emotional attitude has been introduced, I shall discuss its weaknesses. Specifically, there seem to be two problematic implications of his theory. Firstly, since trust as an emotional attitude is grounded in a relationship and a feeling of connectedness, it does not seem to fit cases in which there is no such underlying tie between trustor and trustee. If there is no basis of shared values and norms, then there also is no basis on which the trustor can develop a justified expectation about the trustor's actions. In other words: if the trustor cannot perceive the trustee as somebody "whose actions […] are guided by common interests", then they cannot genuinely trust the trustee, but merely rely (ibid.: 152). These considerations give rise to the assumption that trust, as defined by Lahno, is an attitude that is limited to trustors and trustees who already have a kind of relationship and therefore also share a way of seeing and a way of feeling connected to each other.

Secondly, Lahno claims that genuine trust has emotional aspects and is therefore "essentially identified by features that we generally find characteristic of emotions" (ibid.: 153). This entails that trust "guides our thought just like emotions characteristically do" (ibid.: 154). If these statements are taken at face value, then this leads to severe difficult implications. If we do not trust for reasons of rational consideration but emotional reasons, we are prone to prejudice. A racist white cancer sufferer could rather trust his white friend instead of a black doctor and therefore be ill-advised. A troubled teen could trust her abusive boyfriend over a worried teacher. While this is certainly often the case, it does not necessarily have to be so. One can easily construct counterexamples of trusting in which the trustor actively resists being led by emotions. Imagine an HR boss putting his trust in a woman and giving her the job instead of the male candidate, *although* his conservative upbringing makes him feel like the latter would do a better job. The kind of trust I am addressing here is what Karen Frost-Arnold (2014) calls 'corrective trust'. Frost-Arnold argues that this corrective trust is necessary to fight instances of testimonial or epistemic injustice. We are dealing with a kind of testimonial injustice if a hearer's or trustor's prejudice about the speaker or trustee does not allow them to believe them, or in other words to grant them the credibility they deserve. Thus, Frost-Arnold states that it is possible to choose to trust someone in order to fight this problem and compensate the injustice[10].
This is a highly important point that Frost-Arnold makes, and it is especially pivotal for a trusting societal climate and a stable and fair society. However, Lahno's account does not include cases like that. The problem is that Lahno ascribes too much of an affective component to trust. Identifying trust as guided by emotions neglects other reasons for trusting (epistemic, practical, normative etc.). To conclude, a theory of trust cannot (and also should not) consider trust analogously to only emotions since it does not cover the variety of trusting situations then.

2.b.iii. Application to the Paradigmatic Examples

Now that two influential non-cognitive accounts have been introduced, I will examine whether they can account for all of the paradigmatic examples. The question that will be answered here is whether each of the examples counts as an instance of trust on these

[10] I will come back to the phenomenon in the last chapter when I outline how my own understanding of trust can account for cases of trust at will – which is important to undermine testimonial injustice.

accounts. If this is the case, I want to say why they succeed. If not, I want to carve out the reasons for it. These findings will be used to create a list of desiderata for a theory of trust in the following chapter.

I will start investigation with those two examples, in which trustor and trustee are sharing a deep, long-lasting relationship: *Shoplifting* and *Assault*. Firstly, I want to discuss whether Holton's account is fitting to describe those situations. At first glance, in both examples, the trustor seems to rely from a participant stance – which is how trusting is defined on Holton's account. Thus, this justifies them to be considered instances of trust. However, the implication, that trust at will is possible, is a problematic one. Holton's theory is more useful for cases in which trustor and trustee do not share much history and there is little evidence that the trustee is trustworthy. This is visible in the main opening example by Holton about someone deciding to trust a group of strangers at a drama class (cf. Holton 1994: 63). Holton does not elevate his account to cases, in which a certain trusting relationship is already established, like in *Shoplifting* and *Assault*. As mentioned, his theory defines trust simply as relying from the participant stance, which is a way of seeing the other as a person. However, it seems as if it is missing a nuance: in those cases, in which trustor and trustee have an underlying relationship, they do not see each other merely as persons, but as friends, partners or family. While Holton does state that it is not possible to decide to have a trusting relationship, he does not investigate situations in which trust is in tension with evidence. There is no definite answer to the question of whether trust like this is simply relying from a participant stance, or whether it should be treated differently. There are reasons to doubt that Holton's definiens covers situations like *Assault* and *Shoplifting*. Resa's and Kian's attitudes towards Sarah and Armin seem to be more committed than a simple 'way of seeing', which explains why they are so determined to trust them for the truth. A personal relationship is more than a stance. However, if Holton defined trust as a closer bond than a simple way of seeing, his argument about trust at will would not be convincing – since it is not intuitive that one can decide to have a tight bond with someone. Thus, it remains questionable whether Holton's theory can account for the peculiarities of trust as described in the two examples discussed.

Lahno's theory of trust as an emotional attitude, on the other hand, is more promising when it comes to describing *Assault* and *Shoplifting*. Both are trusting situations on Lahno's account since we are not dealing with a kind of rational risk calculation but with an emotional attitude on the trustor's side. There is a deep relationship between trustor and trustee, which justifies the trust's strength and resistance to evidence. The underlying

trusting relationship, which is described as emotional, determines how the trustors think; how they perceive the world. That is the reason why the trustors in the examples acknowledge the relationship more than the evidence presented – although this evidence should make them doubt. Lahno's account is very fitting to explain those trust situations in which trusting goes counter the evidence since its focus is on the affective, not cognitive component of the attitude. Trust as emotional attitude guides our perception and reasoning, explaining why trustors trust, even if it is in tension with evidence. It is, to a certain degree, "independent of objective information", such as evidence (Lahno 2001: 171). This is most visible when it comes to personal relationships:

> "[T]he trust of a friend encompasses patterns of normative evaluation that will motivate the trusting friend in a certain way […]. The trusting person perceives the friend and that part of the world that is related to the friend and his actions in a certain light. This is the core of his trust. It is an emotional attitude as defined above." (Lahno 2001: 178)

Next up is the examination of the examples *Library* and *Climbing*. As mentioned earlier, these situations are interesting because trustors and trustees are strangers to each other. For Holton, this poses no problem: since the only condition for trust is relying from a participant stance, these situations are trust situations on his account. The participant stance theory is particularly useful in this matter. While it is difficult for cognitive theories to account for trust in strangers, Holton's theory can deal with it well. Since he argues for the possibility of trusting at will, Anna's and Felix' trust does not need to be grounded in evidential reasons – it may have been willed for other reasons, such as reasons of usefulness or sympathy.

However, Lahno's theory falls short of properly accounting for these two examples as trust situations. The reason for this is that trust in strangers is not properly examined, while trust in personal relationships is highlighted. This comes as no surprise, since an account whose building blocks are emotions is better equipped to describe trust between friends than trust between strangers. It is not easily applicable to strangers since his paper suggests that the stronger trust is, the more emotional the underlying relationship between trustor and trustee is. If trust is stronger, the stronger the relationship is and weaker, the weaker the relationship is – what if there is no relationship at all? Does this entail that trust is not possible between strangers?

Lahno writes, for example, that a special relationship might induce a trustor to see situations "in a certain favorable light" – which means trusting friends over objective information or trying to reevaluate given information so that it suits the (perfect) picture of the friend (ibid.: 178). However, all of these are peculiarities of personal relationships do not apply to strangers. Therefore, it seems as if trust in strangers is more of a calculated risk evaluation or mere reliance, not an emotional attitude. This forbids for situations like *Library* and *Climbing* to count as instances of trust. Since there is no emotional attitude towards the trustees, there are no explicit shared values or norms between trustor and trustee, and the attitude is also not independent of objective information.

So far, I have shown why Lahno's is not the right account to make sense of trust as described in *Library* and *Climbing*, as well as that Holton's account is more useful for trust situations like these than those in which there is an underlying personal relationship. This is sufficient for justifying my claim that neither of them is extensive enough to cover all six examples. Some examples are trusting situations that are not met by the definiens that the authors suggest, hence they are counterexamples to those theories. However, for the sake of completeness, I will lastly touch upon how Holton and Lahno deal with the paradigmatic examples *Key* and *Letter*. Both of these trusting situations are very straightforward – or 'standard', if you like. Trustor and trustee are in no kind of special relationship with each other; that means that they are neither strangers nor do they stand in a close personal relationship. Since it is only those kinds of relationships that demand a special investigation when it comes to trust, examples like *Key* and *Letter* are mostly unproblematic for both accounts.

To conclude this second chapter, I will summarize the main findings. Intuitively, I indicated that it is easy for a theory to account for trusting situations in which trustor and trustee already know each other. However, in cases in which trustee and trustor are strangers – like in *Library* and *Climbing* – the element of trust at will seems necessary. Lahno and Hieronymi cannot account for these situations since deciding to adopt a belief or an emotion is dismissed. While Holton's account can capture that, it does not grasp the essence of a deep, trusting relationship between close friends, partners, or family. Thus, it does not accurately describe trusting situations like *Assault* or *Shoplifting*.

The findings lead to the main conclusion of this chapter, which is that there is no consensus on how to define trust among these major theories of trust. Furthermore, none of them manages to account for all paradigmatic examples as legitimate, genuine trusting situations. For that reason, I want to suggest an alternative view. I will do so by first carving out what

an ‘ideal’ theory of trust should account for by establishing a list of desiderata it has to meet. An ideal theory of trust is a theory that meets the demands of said desiderata and can account for all the paradigmatic examples. The desiderata will be constructed in a way so that if they are fulfilled, then the theory escapes those problems I just ascribed to Hieronymi’s, Holton’s and Lahno’s theories. Thus, an ideal theory of trust should not face any of the issues that these three theories face.

Once the desiderata are established, I will show that one theory of trust – especially none that is conventional – cannot meet all these desiderata, without running into incoherence. Thus, I will look at trust in another way. This way will include moving away from the conventional thoughts that trust is a uniform phenomenon and that there is only one kind of trust. I am convinced that this new perspective can move the discussion forward and resolve the current theories’ problems.

3. Trust Is Not a Uniform Phenomenon

Before diving into this chapter's main claim and its defense, I will first take stock of what led me to this point and then introduce the structure of the argument. This is to illustrate how the contents of the previous chapters provide the basis for the following endeavor as well as to make it more easy for the reader to follow the upcoming steps and to connect them to the points that I already have made.

This thesis is looking for an understanding of trust, which is neither too narrow and therefore prone to counterexamples, nor too wide and therefore not capturing trust in a way that properly demarcates it from other similar phenomena like reliance. Chapter 1 focused on an introduction to trust. I presented six paradigmatic examples of trust that form the basis for many investigations throughout this thesis. I demanded, that a proper theory of trust can account for each of these cases. Then, the reader was further provided with a more formal analysis of the concept of trust. Moreover, some implications of my work are about the rationality of trust. Therefore, to complete this introductory chapter, I covered the formalities, values and problems of rational trust. Next, I aimed to introduce the reader to the issues with the current state of research on trust. To do so in a proper manner, I resorted to existing theories of trust from the philosophical literature. Thus, in Chapter 2, I highlighted three different, influential accounts to find out in which regard they differ and what their advantages and weaknesses are. The main finding here is that, within the current philosophical literature, there is no consensus about what interpersonal trust entails. Moreover, each of the highlighted theories had difficulties in accounting for some of the paradigmatic cases of trust. Since the analysis of prominent theories of trust in Chapter 2 did not provide us with an answer to the question about what conditions need to be satisfied for a mental attitude to be a trusting attitude, I will examine the problem further.

Thus, in the ongoing Chapter 3, I will carve out the demands that a theory of trust has to meet to account for all cases which I consider to be trusting situations – in particular the introductory paradigmatic examples. Then, I will show that those theories, which had been the focus in Chapter 2, do not fulfil these desiderata and subsequently establish the claim that no conventional theory of trust can meet them all. Thus, I am facing the issue that no theory can satisfy my conditions, d.i. neither meet the demands of the desiderata nor account for all paradigmatic examples. Following that, a solution will be offered: the problem dissolves by acknowledging a novel understanding of trust, which does not consider trust to be a uniform phenomenon.

3.a. The Difficulties

3.a.i. The Desiderata

What I have established so far is that instances of trust are more varied than one would expect. While the six opening examples from Chapter 1 appear as if their common denominator is 'being a trust situation', they now seem to involve different but related phenomena. Since the attitude of trust varies substantially from instance to instance, it proves difficult to pinpoint its properties. Several philosophers have tried to define what trust is. In Chapter 2, I have shown that none of the theories that I dealt with, is able to satisfyingly interpret all six examples. Rather, each theory only fits some of the paradigmatic examples. In this chapter, I will present this difficulty in a more formal way by establishing a list of desiderata a theory of trust should fulfil. Besides that, I will also show that neither Hieronymi's, Holton's and Lahno's theory, nor any other theory can meet them all. Finally, I will suggest another understanding of trust, one that is able to fulfil all desiderata and does justice to all paradigmatic examples.

Let me start with my demands that a theory of trust should satisfy. The first requirement is for trust to be sufficiently distinguished from reliance[11]. This means that trust is not only characterized by depending on someone to perform an action X and hoping that the other will come through, but also by some element that makes the trustor think that the trustee is trustworthy and thus makes the trustor expect that the trustee carries out X (cf. Faulkner 2007). This gives rise to the idea of including a certain cognitive element into the list of desiderata of trust. To explain this, I will (like many) follow Baier (1986) in saying that the most salient difference between trust and reliance is that broken trust gives rise to feelings of betrayal, while misplaced reliance is only followed by disappointment. To feel betrayed, the trustor must have been convinced that the trustee will carry out action X. Thus, it makes sense to say that the trustor is only justified in feeling betrayed if they believed or expected the trustor to prove trustworthy and perform X. The attitude of reliance is then similar to hope, while the attitude of trust is more similar to belief. To conclude, this is the first desideratum for a theory of trust: a theory of trust has to sufficiently distinguish between

[11] This also applies to other similar attitudes, for example pretend-trusting. However, the differentiation between trust and reliance seems to be primary within the literature, thus I will keep with e.g. Baier (1986), Holton (1994) and Faulkner (2007).

trust and mere reliance. Incorporating a cognitive element is an excellent starting point since it makes the distinction between trust and reliance clear-cut and comprehensible.

Secondly, a theory of trust should allow for trust in strangers. This is crucial for trusting situations like *Library* and *Climbing*, in which trustor and trustee do not know each other and have no specific evidential basis concerning the potential trustee's trustworthiness. For a belief to be properly justified, many philosophers agree that it "has to stand in the right sort of relationship to evidence" (Townsend/Wanderer 2013: 1). However, in cases in which trustor and trustee are strangers to each other, the evidence about the trustworthiness of the trustee is vanishingly small. Keep in mind that 'being trustworthy in regard to an action X' is understood as a certain combination of motivation and skill/knowledge. When trusting strangers to do X, the trustor does not have evidence about whether the trustee is motivated or skilled to perform X. Thus, in the case of trust in strangers, trustors do not have (sufficient) evidential backup to form a belief about the trustworthiness of the trustee and/or about whether they will actually carry out X.

Therefore, one is entitled to two inferences, namely a. that people do not trust strangers for evidential reasons (but other ones) and b. that the attitude of trust towards strangers is not an entirely cognitive one. An understanding like mine, which wants to integrate how to apprehend trust in strangers, also needs to attend to trusting at will. In other words, since trust in strangers does not emerge because of evidential reasons, there have to be other reasons for this kind of trust. As exemplified in *Climbing*, one reason to trust a stranger could be usefulness or making the stranger feel good. These are non-evidential reasons, which lead to the *decision* to trust. I do agree with the literature (f.e. Baier 1986) that one ordinarily cannot choose to trust. Especially not, if one already believes someone to be distrustful. However, in the case of trust in strangers, the trustor does neither believe the trustee to be trustworthy nor untrustworthy. Because of that, it is sensible to include an element of voluntariness into a theory of trust: if there is no underlying relationship (and no evidence about whether or not the trustee is trustworthy), it is possible to decide to trust somebody (cf. Hinchman 2020). This demand on a theory of trust contradicts the assumption that the attitude of trust is solely cognitive, since cognitive attitudes, like beliefs, cannot be chosen at will.

Thus, this is the second desideratum for a theory of trust: to properly account for trust in strangers and, in achieving this, to also account for cases of trust at will. This may be

achieved by allowing for a wider range of reasons for trust (like practical reasons) than merely evidential reasons.

Thirdly, a theory of trust should be able to account for cases in which trustors trust although it goes against the evidence. That this is an element of trust, which cannot be neglected, becomes evidential in situations exemplified in *Assault* and *Shoplifting*. In the last chapter, Townsend and Wanderer's argument showed that a trusting attitude is not always governed by the norms of theoretical rationality and thereby, by evidentialism (cf. Townsend/Wanderer 2013). Also, philosophers like Lahno (2001), as described in Chapter 2, and also Jones (1996) and Faulkner (2007) give convincing arguments for why a non-cognitive, affective part of trust should be considered by a theory of trust. If a theory of trust only considers trust as cognitive, then it cannot make sense of cases in which trust goes against the evidence. Normally, evidence is considered as a reason to trust or not trust. However, in some situations, trustors neglect the evidence and believe the trustees instead. Thus, evidential reasons sometimes compete with non-evidential reasons – and a theory of trust has to make sense of that.
To conclude, I argue that a theory of trust needs to account for the impression that in some cases trust seems to justifiably go against the trustor's evidence. Trust is then not a cognitive attitude like belief, but an emotional one, since there are other (affective) reasons for trust, instead of evidential ones. This is the third desideratum for a theory of trust.

3.a.ii. How the Desiderata Are Not Fulfilled

Now that I have stated the three desiderata for a theory of trust, I want to first show how none of the theories of concern by Hieronymi, Holton and Lahno is capable of meeting their demands and subsequently, that no other conventional theory can meet all of them. I will deal with the theories in the same order they are presented in Chapter 2. Thus, first up shall be Hieronymi's cognitive theory of trust. This account highlights that trust should be considered a belief about the trustor's trustworthiness. The only reasons for trust that Hieronymi allows are evidential reasons which show the trusting belief to be true. Thus, trust is not justified by other reasons, such as those of usefulness or goodness. This theory has a very strong cognitive element, which demarcates genuine trust from other attitudes like mere entrusting and reliance. Therefore, it does fulfil the first desideratum. However,

I argue that Hieronymi's theory falls short of meeting the expectations of the latter two demands.

It is obvious why Hieronymi's conditions on trust and my second desideratum are not consistent. Hieronymi explicitly forbids for practical reasons like reasons of usefulness or reasons of sympathy to count as reasons for trusting. Trust is identified as belief and beliefs are governed by epistemic reasons, first and foremost evidential ones. Thus, if Hieronymi wants to allow for trust at will, she has to commit to doxastic voluntarism – that is, claiming that it is possible to choose one's beliefs. With Hieronymi's account being strictly cognitive, she cannot fulfil the second desideratum: making sense of genuine trust in strangers, since trust in strangers needs a voluntary element.

Furthermore, the theory is also not consistent with the third desideratum, which states that a complete theory of trust should also account for trust, that goes against the trustor's evidence. Trust on Hieronymi's account is considered a belief and beliefs are justified by evidence. Thus, it is contradictory to allow for the attitude of trust while certain evidence is dismissed. Fulfilling the desideratum would mean acknowledging the non-cognitive, affective component of trust that Hieronymi's strictly cognitive theory does not.

Let me now turn to Holton's participant stance theory and show that it also fails to satisfy the demands that have been established in this chapter. While his theory is non-cognitive and defines trust as a stance rather than a belief, it does not commit the errors of Hieronymi's theory. Since a stance is epistemically less demanding than a belief, Holton can easily account for trust at will and therefore trust in strangers – which is not surprising, since arguing that trust can be willed at times is his whole program (cf. Holton 1994). However, it does not meet the demands of the first desideratum to sufficiently distinguish between trust and reliance. While it is fairly easy for cognitive theories to demarcate trust from reliance[12], it is more demanding for non-cognitive theories. Holton states that trust is different from relying insofar as it is reliance *from the participant stance* and that, therefore, the different reactive attitude is different when trust is broken than when reliance is disappointed. Nevertheless, the definition of Holton's participant stance is sparse. He describes it as something that involves seeing the other as a person and having a readiness towards the other to be betrayed (cf. Holton 1994: 67). However, seeing the other as a person is not a sufficient condition for trust since this is also the case with relying on

[12] A popular distinction is that trust is considered a belief, while reliance is more similar to a non-cognitive attitude like hope (cf. Baier 1986).

someone. Furthermore, Holton does not state how the apparent attitude of readiness to being betrayed differs from readiness to being disappointed. Therefore, it seems to me that the missing cognitive element of Holton's theory leads to an imprecise distinction between trust and reliance. Since this is what the first desideratum should ensure, Holton's theory does not meet its demands. Concerning the third and last of the desiderata: Holton does not engage with the emotional aspect of trust and how it may compete with evidential reason. While he does highlight that trust is not completely cognitive, he does not make claims about trust going against the trustor's evidence. Hence, I will omit the evaluation of whether Holton fulfils the third desideratum. His account not meeting the demands of the first one is sufficient to build my case that none of the theories from Chapter 2 can meet them all.

Lastly, I will argue that Lahno's theory also fails to meet the demands of some of the established desiderata. The weak point of Lahno's theory is how it does not fulfil the second desideratum about allowing for trust at will. He highlights the emotional component of trust and critiques an overly cognitive, risk-assessing definition of trust. While this is important in order to account for the demands of the third desideratum, it is inconsistent with those of the second one. Since one cannot choose emotional attitudes, trusting at will is denied. The second desideratum ensures that trust in strangers (like in two of the opening examples) is properly accounted for. However, since Lahno's emotional attitude is grounded in a relationship, it cannot meet this demand. Still, this theory is the best one to explain how trust might compete with evidence, which is the content of the third desideratum. Lastly, for the sake of completeness, Lahno does demarcate trust from mere reliance. He says that it is exactly the emotional component of trust which gives the trustee reason to perform X, and thereby gives the trustor reason to trust. Without an underlying personal relationship, trust cannot be established. While this does distinguish reliance from trust within relationships, it does not distinguish reliance from trust between people without one. Because of that, trust in strangers seems to be dismissed as mere reliance.

It is possible to draw from these investigations that each of the theories of concern highlights another element that is peculiar to trust[13]. Hieronymi's account focuses on the cognitive aspect: that trust should be identified with the belief, that the trustee is trustworthy and not simply with hope or habit. This is highly important to the attitude of trust since a

[13] Simpson (2012) claims something very similar: There is no consensus between current theories of trust, because the word 'trust' is used to refer to cognitive, conative and also affective mental states. Each of these theories highlight another peculiarity of trust and exclude those accounts, who focus on another one. Therefore, different strands of thinking about trust focus on different dispositions or motivations (cf. Simpson 2012: 564f.). I will come back to Simpson's findings in the last chapter of this thesis.

trustor does not doubt that the trustee will perform action X – a trustor *believes* that the trustee will do so. If not, the attitude would be mere reliance. Holton's account focuses on cases in which we have to decide to trust because we do not have the belief (yet) that the trustee is worth our trust. While this is contradictory to Hieronymi, it is equally important – if we did not consider cases like that, we could not make sense of trusting strangers. Since trust in strangers is possible and relevant, it has to be taken into consideration. Lahno's theory highlights the emotional aspect of trust and can thus explain another component of trust, that Hieronymi neglects. Thus, Lahno's theory is fitting to describe the phenomenon of how trust in personal relationships competes with evidence – a phenomenon that surely is part of trusting relationships.

Nevertheless, none of these theories can account for all of the desiderata. To me, this is not surprising, since the demands contradict each other. While the first desideratum asks for trust to involve a cognitive attitude, which is tied to evidential reasons, the second and third desideratum ask for trust which is a non-cognitive attitude and for non-evidential reasons of trust.

I conclude at this point that no conventional theory, which identifies trust as *either* cognitive *or* non-cognitive attitude, can meet all of the desiderata, and additionally, that no theory can account for all of the paradigmatic examples from Chapter 1. No conventional – or standard – theory of trust can view trust as both a cognitive and a non-cognitive attitude at the same time and consider both evidential and non-evidential reasons for trusting as rightful without being either inconsistent or too vague. Thus, I would like to introduce the reader to another way of thinking about trust in the next subchapter. This new approach to trust should account for all of the peculiarities of trust and neither runs the risk of neglecting important aspects nor of being inconsistent.

3.b. A Solution

3.b.i. Trust as Manifold, Not Uniform

Instances of trust are manifold – and any theory of trust has to make sense of this variety. In some cases, one is dealing with trust between new acquaintances or even strangers; with relationships in which trust has not yet been established before. In cases like that, a trustor might decide to give someone a chance and to put trust into a trustee, although there is no way to know whether they will prove trustworthy – like when meeting a stranger from the internet. On the other side of the spectrum, trust between family, friends or partners is

already there. Oftentimes, it has been there for so long and has been established so firmly, that it, as Karen Jones (1996) puts it, gives the trustor blinkered vision. In cases of trust that strong, trust might compete with evidence – like believing your boyfriend's word over the pictures of him cheating that your colleague sent you.

So, trust might be fresh and not established yet or, on the opposite, has grown strong and stronger over a long period of time. However, the point is that both kinds - and also everything in between - should be considered trust and should be covered by trust's definition. The question is, how is that possible? So far, I have established that trust is manifold in its occurrences. Moreover, I have stated that the mental attitudes of these trusting situations and the reasons for trust can not only be different but also contradictory (recall, e.g., that trust sometimes occurs as an attitude that *can* be willed and sometimes it *cannot* be willed). Now, how can we account for a phenomenon with these characteristics in a meaningful way?

The solution I would like to propose in this thesis is to recede from the thought that trust is a uniform phenomenon and alternatively consider the term 'trust' as a polysemous expression for different forms of trust. I claim that the tension between rival theories of trust originates from the authors unknowingly having different forms of trust in mind. Besides that, not only do the authors of concern miss that there is more than one kind of trust, they also do not acknowledge that these forms of trust are connected to the sort of relationship that the trustor and the trustee share. While some philosophical work on trust does grant that trust is not uniform and that the broad term of 'trust' actually refers to different types, there has neither explicitly been drawn the connection that different kinds of relationships call for different forms of trust, nor has the term 'trust' been identified as polysemous. What is new as well is that these forms of trust are defined differently, not only when it comes to their mental attitude, but also when it comes to voluntariness, justification and rationality. Since the definitions of the types of trust contradict each other to a certain degree (e.g. in being cognitive or non-cognitive, voluntary or involuntary), but we still use the same word to denote these related phenomena, trust should be considered polysemous. Explaining these claims shall be the main goal of the rest of this thesis. In the end, I also want to offer some thoughts on how this new understanding of trust is beneficial to the philosophical debate about trust, especially to those about the rationality of trust. For now, I will start by stating how a theory of trust that acknowledges different forms of trust can fulfil the demands of the desiderata and can also account for each of the paradigmatic

examples. At last, the forms of trust I have in mind will be introduced in detail in the next chapter and the advantages of trust as polysemous expression will be listed.

Attentive readers might have noticed that two of the paradigmatic examples exemplify one kind of relationship, respectively. In *Library* and *Climbing*, trustor and trustee do not have an underlying relationship; they are strangers to each other. In *Key* and *Letter*, trustor and trustee do share a relationship, however simply as colleagues or acquaintances. Lastly, in *Shoplifting* and *Assault*, trustor and trustee have a deep, personal relationship. The examples teach us that the kind of relationship between a trustor and trustee is pivotal for how trust between them should be perceived. That is why I am certain that different relationships between trustor and trustee call for a different analysis of the trust between them. To me, being strangers, acquaintances or friends/family, are the three main relationships, which occur between trustor and trustee. That is why I suggest three forms of trust. Let me elaborate on this theory.
Each instance of trust falls into one of those three categories and is therefore considered one particular form of trust. One kind of trust cannot fulfil all the desiderata but together, they can. Thus, there is not one kind of 'trust' that simultaneously is a cognitive and non-cognitive attitude, but trust can be either cognitive or non-cognitive, depending on which form of trust is discussed. In other words: one definition of one single form of trust cannot meet the desiderata. However, by distinguishing different types of trust those desiderata can be met by their separate definitions. Also, if we acknowledge that we are dealing with three different forms of trust instead of only one, the problems which have been discussed in the first part of this chapter, are resolved.

3.b.ii. The Impact of the Relationship

So far, I have claimed that there are different forms of trust, which depend on the relationship between trustor and trustee. I will here give an argument about why the underlying relationship impacts the way one trusts. I want to convince the reader that there is a clear variance in the attitude of trust, and that this variance depends on how the trustor relates to the trustee. In other words: the attitude of trust, attended by the rationality of this attitude of trust is conditional on whether the trustee is a stranger, an acquaintance or a beloved friend, family member or partner to the trustor.

Trust differs, depending on the relationship between trustor and trustee. To me, this is the reason why it is so hard to determine what trust is – and, also, why it is so simple to find counterexamples for possible definitions. Several accounts of trust focus on only one type of trust and are therefore less inclusive. For example, definitions like Hieronymi's are designed to fit only full-fledged trust between people, who already know and expect from each other that they are trustworthy. That is trust between on the one side, a trustor with evidence about a trustee's trustworthiness and hence, on the other side, a trustee about whose trustworthiness it is easy to form a belief. However, accounts like this are susceptible to counterexamples – like *Library* or *Climbing*. Recall: these instances of trusting, on Hieronymi's account would not be considered genuine trust but mere entrusting since the reasons for which the trustors trust, do not support the truth of the trusting belief. Instead, those reasons are practical reasons for acting and not believing, such as reasons of usefulness.

On the other hand, some accounts allow for (too) many situations to count as trust and thus are more inclusive. For example, it is often said that one cannot adopt the attitude of trust at will. Still, Holton argues that this is possible despite other philosophers' objections that these situations should not be regarded as real trust. Therefore, Holton can include instances of trust at will, while others' definitions are narrower. There are more theories than Holton's that support a non-cognitive and therefore broader understanding of trust. Many claim that it is possible to trust without evidence for someone's trustworthiness. One of these accounts, that I already mentioned in Chapter 1, is by Karen Frost-Arnold (2014). She argues that the attitude of trust should not be considered a belief only and therefore, that also non-evidential reasons should be accepted to justify the attitude. While some instances of trust indeed should be considered beliefs about the trustee's trustworthiness, trustors also engage in certain forms of trust for exactly those non-evidential reasons, that do not support belief (cf. Frost-Arnold 2014: 1960ff.). I fully agree with Frost-Arnold on this issue: trust, simply defined as belief, is too narrow to wholly represent the variety of different forms of trust.

Therefore, I suggest that cognitive accounts are usually too narrow to meet all the desiderata of trust and non-cognitive accounts are usually too broad to precisely state what trust is and is not and thus, put insufficient constraints on the concept of trust. Again, we are confronted with no consensus on what trust is.

It is important to put this into order. Conventionally, 'trust' is considered a concept, which refers to exactly one phenomenon (hence my use of the term 'conventional theory').

Philosophers have tried to analyze the concept and give sufficient and necessary conditions to give a definition that does justice to the phenomenon. However, as shown, there is no agreement on how to do so, since counterexamples to the definitions offered can be construed. Since I am convinced that this conventional way of analyzing trust will always be susceptible to counterexamples, I suggest a different approach to trust. This new understanding comprises understanding 'trust' as a case of polysemy. I consider 'trust' to be a term, which denotes related, but distinct phenomena – namely, different forms of trust. The goal for this following part is to show that there *are* various forms of trust that – and this is crucial to my approach – depend on the kind of relationship trustor and trustee have. In order to convince the reader of this proposition, I shall do the following: in the remaining part of this chapter, I want to describe three nearly identical trusting situations, in which the only difference is the relationship between trustor and trustee. By keeping the conditions fixed and only changing the parameters of how trustor and trustee relate to each other, I want to pinpoint the differences in the trusting attitude. Subsequently, in the fourth and last chapter, I will take a closer look at those different kinds of relationships separately, introduce the respective forms of trust and their characteristics as well as give a formal definition for each of these forms of trust.

3.b.iii. There Is a Difference

Let us consider the fact that philosophers do not agree on what the mental attitude of trust is. Depending on their own view of trust, they state different things. Trust is a stance, which one can adopt at will. It is a belief, whose truth should be supported by evidence. It is an emotional attitude, which can run counter said evidence. Not only are these views different, but they are also competing. However, if we agree that the mental attitude can differ depending on the particular trusting situation, then this tension dissolves. This is exactly my approach: there is more than one 'trust'. It is not a uniform phenomenon – and the differences can be located within the initial relationship between trustor and trustee. Depending on their personal relationship, the attitude of trust shifts.

Another insightful notion to discuss is the reactive attitude of trust. The 'breach of trust' and the feeling of betrayal that comes with it, is a unique feature of trusting. Only if I really trusted, then I can feel this certain reaction (if my trust is broken). Consequently, if I do not feel betrayed, I did not initially trust. However, I claim that the reactive attitude of trusting is not uniform as well. That could mean either that sometimes there was trust and

sometimes there was none – *or*, it means that there are different kinds of trust, which come with different kinds of reactive attitudes. I shall claim the latter: different forms of trust manifest in these different kinds of reactive attitudes.
Now, let us have a look at three examples, which should make this claim clearer.

i. Imagine the following scenario: it is summer. You are at a lake, alone. You brought along your backpack, with your belongings, such as your phone, wallet etc. Then, you decide to go for a swim. Politely, you ask the person who sits next to you, to watch your backpack while you are in the water. You trust them not to steal from you and watch it, so nobody else steals from you. Everything goes smoothly. However, on the way home, on the train, you notice that €50,- are missing from your wallet. Your trust has been broken. How do you react? Of course, you will be angry with the person you trusted. They looked nice, they were there alone as well and thus knew that leaving your belongings unattended constitutes an unsettling feeling. By asking them to watch your backpack, you showed them that you trust them. Yet, they stole from you. Hence, you are angry. But more importantly, the person you will be most annoyed with, will be yourself. How could you be so stupid and naïve?

ii. Now imagine the same scenario, but this time you are at the lake with someone you have been dating for a while. Maybe it is your third date. You do not know them very well yet, but they are no stranger to you. You decide to go swimming and trust them with your backpack. On your way home, you find out that €50,- are missing. How do you react here? You will feel betrayed by this person. They simply broke your trust and you resent them for doing so.

iii. For the last example, imagine the same scenario but you are there with your best friend of ten years, Gustaf. Once again, you go for a swim. You leave your backpack with them and on your way home, €50,- are missing from your wallet. What would your reaction be? You would be confused – how could this happen? There has got to be an explanation for this. First of all, Gustaf would not steal from you. If he needed money, he would ask. Secondly, he would not leave your backpack without a very good reason. He is a reasonable, trustworthy person who would not do something like this. You are starting to wonder: maybe the money got stolen while you guys were talking? Maybe it was stolen on the train ride? And, are you really sure that you had those €50,- in your wallet from the

beginning? You are doubting yourself. You will not accept the fact that Gustaf just broke your trust. Your reaction will not be resentment – but puzzlement.

These three examples are highly interesting concerning what they tell us about trust in regard to the reactive attitudes. The difference of these attitudes points to differences in the initial trusting. It indicates that there had to be a slight disparity in the mental attitude of trust in the beginning. In example i., I described the reactive attitude as anger, at both the stranger and oneself. Why would I be mad at myself in that situation? I think this is because it was *my decision* to trust the stranger. The stranger did not invite me to trust them. I misjudged their trustworthiness. I could not know whether they will actually guard my belongings, but I trusted them because of practical reasons: because it was useful to have somebody watch them while I went swimming. Thus, I infer, that the attitude of trust was non-cognitive and voluntary. In example ii., I feel betrayed, angry at my company. Because of our ongoing relationship, I felt like I could trust my date. I knew that person and had gathered evidence, that they can be trusted – especially regarding guarding my things and not stealing from me. I believed them to be trustworthy – but they failed me. I claim that the initial trusting attitude was cognitive; a trusting belief about the date watching my belongings. Lastly, in example iii., the attitude of trust was an emotional attitude, which is so strong, that it made me doubt the evidence – that is, that Gustaf very likely stole from me. Thus, the reactive attitude is different from examples i. and ii. – before I feel betrayed by my best friend, I will look for another explanation. One, that does not involve Gustaf stealing from me.

The underlying relationship between trustor and trustee is very clearly the reason for the shift in the reactive attitudes. This supports my claim that the mental attitude of trust is impacted by this relationship. Thus, a substantial examination and justification of this hypothesis is necessary. This shall be the subject of the next chapter. As of now, keep the following preliminary claim in mind: there are differences in the mental attitude of trust, depending on how strong the relationship between trustor and trustee is.

4. Three Forms of Trust

In the previous chapter, I have sketched a new approach to understanding trust which embraces the variety of the phenomenon – that is, apprehending the concept of trust as a polysemous expression. The initial groundwork for this enterprise has illustrated that there is a missing consensus about the attitude of trust within the current theories. Support for my claims about trust not being a uniform phenomenon comes from the conclusion, that any theory, that identifies trust with only one specific mental attitude, cannot meet the demands of the desiderata I introduced in the last chapter. Furthermore, I explored how the differences between distinct forms of trust are connected to the relationship that trustor and trustee share. Specifically highlighted were the differences in the reactive attitudes of trust. These are of relevance since they are partly constitutive of the initial attitude of trust. Thus, the main claim of the last chapter was that trust differs depending on the relationship between trustor and trustee. Specific relationships demand separate analyses of the trusting behavior exhibited within them. Moving on, I want to provide a non-formal description of each of the forms of trust T1, T2 and T3. The reason for this is to make the reader understand to which subcategories the concept 'trust' may refer to and how they differ from each other. Moreover, I want to motivate why it is beneficial to do so. To support my claims here, I will implement considerations from the philosophical literature. What follows will be a formal definition of each form of trust as well as an analysis of how they fulfil the desiderata I establish for theories of trust as well as which paradigmatic examples they account for. In the end, I will have shown that my approach fulfils all the desiderata and accounts for each of the opening examples.

4.a. Trusting Strangers (T1)

Trust in strangers is a mysterious phenomenon. That is because it involves genuinely trusting someone you do not know. What makes it hard to trust a stranger is that there is no (or only very little and general) evidence about a stranger's trustworthiness accessible to the trustor[14]. I postulate that a trustor can trust somebody even if they cannot be certain

[14] What I mean by that is the following: The fact, that the stranger you met at the bus station does not carry a weapon and bought a ticket, can be considered evidence that he can be trusted to be a civilized member of society and to not hurt you during the bus ride – however, this is not the kind of evidence you share with colleagues, friends or family. The kind of evidence I have in mind is evidence of their trustworthiness towards *you*. Since you never trusted a stranger before, you have no evidence whether they will prove trustworthy to *you*.

whether the trustee will prove trustworthy, as is exemplified in *Library* and *Climbing*. However, this kind of trust is different from trusting someone you do know.

I admit, that in most cases, a relationship is built before trust is established. Nevertheless, as in *Library* and *Climbing*, it is also often the case that there is no time to bond or test someone's trustworthiness before having to – or deciding to – trust. While trusting complete strangers might be the exception rather than the rule, it is possible and has to be acknowledged on its own – since it has obvious features like its voluntariness, that e.g. trust in friends is missing.

Let us have a look at the literature on trust that supports my claim that there is trust in strangers and see, how it is motivated. As stated in the last chapter, Richard Holton's theory of trust is useful for understanding trust in strangers. To eliminate any doubts and to convince the reader that trusting strangers exists and is fundamentally different to trusting someone you know, I will now discuss some further authors' accounts.

Of relevance are the works by Edward Hinchman (2020), Karen Jones (1996) as well as Karen Frost-Arnold (2014). Hinchman elaborates on the role evidence plays in trust and its rationality. He claims, and this is helpful to understand trust in strangers, that one does not require evidence to trust. According to his account, one can trust someone even if one is missing the evidential basis to evaluate their trustworthiness.

> "You may lack sufficient evidence for the judgment that your trustee is worthy of your trust, but that evidential deficit need not constrain your ability to trust. Even if you have such evidence, that evidence should not form the basis of your trust." (Hinchman 2020: 134)

Moreover, Jones' theory of trust as an affective attitude is non-cognitive and does without the mental attitude of 'belief'. This is promising for the stranger case: Jones simply considers trust to be an optimistic attitude about the trustee being moved by the fact, that the trustor is counting on them. That means, that the trustor is optimistic that the trustee will carry out X because they can sense that their help is needed. Furthermore, Jones highlights how an instance of trust is comprised to one domain, namely the action X. This is beneficial when thinking about trust in strangers because it does not assume an underlying trusting relationship but simply an attitude towards one person, regarding one action. Although Jones does not discuss the voluntariness of trust, this remark makes it easier to accept trusting at will: it is more intuitive to accept that one wills oneself into trusting someone to do one thing, than that one wills themselves into a trusting relationship.

Secondly, I will examine Frost-Arnold's work more closely, as mentioned in the first half of this thesis. Most importantly, Frost-Arnold supports the claim that there are non-cognitive types of trust. This is because we also trust for non-epistemic reasons, which are just as legitimate as epistemic ones are, even if they do not support the truth of a trusting belief. Thus, instances of trust that are based on non-epistemic reasons are rational instances of trust as well as those, that do involve belief. Frost-Arnold's suggested forms of trust – which we engage in not for epistemic, but normative reasons – are therapeutical, corrective and coping trust. Therapeutical trust "can inspire positive change in the trusted part" (Frost-Arnold 2014: 1960). Imagine Mrs. Rogers, who puts her trust into her neighbor, who just got out of prison and struggles to find his way back into society, to mow her lawn for an allowance. She knows that he is not a diligent garden worker, but she trusts him anyway with her garden and equipment, with the goal of sparking something positive inside of him. That Mrs. Rogers is counting on him should motivate him to do his work properly.
Secondly, corrective trust is a form of trust, that has the goal of correcting the phenomenon of testimonial injustice, as already mentioned in Chapter 2. It is a non-cognitive form of trust, which can be adopted at will when a trustor finds that their judgement of the trustee's trustworthiness is impaired.

> "When a hearer suspects that a prejudice is undermining her ability to judge the trustworthiness of someone's testimony accurately, she ought […] to choose to trust the speaker by revising the credibility upwards to compensate for the prejudice." (ibid.: 1961)

Lastly, coping trust is a form of trust which we engage in because it makes it easier to cope with the complexity of life. This involves only trusting experts about, for example, certain scientific findings, simply because one cannot check these themselves. An example would be trusting your doctor and the scientific community about the usefulness and low risk of vaccines because you could not research and develop them yourself. None of these forms of trust is constricted to attitudes that one only holds towards non-strangers, thus they support my claim that trust in strangers exists.
Frost-Arnold's claim involves that all of these forms of trust may be chosen (cf. ibid: 1960ff.). Thereby, the author agrees with Holtons participant-stance theory, which allows for some instances of trust to involve a decision about the trustee's trustworthiness. Note here, that neither of these authors claims that no instance of trust ever involves the cognitive attitude of belief or that each one must involve a decision. Both merely state that there exist

types of trust which allow for another mental attitude than belief. This is in line with my claim that there are differences in trust – however, neither of these accounts highlight the connection to the ongoing relationship between trustor and trustee or the fact that these thoughts are especially useful when thinking about whether to trust a stranger.

What Holton's, Jones' and Frost-Arnold's theories have in common is, that they make a claim about the trustee's trustworthiness without a necessary cognitive element like a belief. Since trustors often do not have a belief of this sort about strangers, it is of importance to have an alternative non-cognitive theory for cases of stranger trust.

To summarize, it is possible to trust strangers despite being confronted with the problem that we do not know for sure whether they are trustworthy. For some accounts, this problem excludes trust in strangers from the definition, as trusting is assumed equal to believing that the trustee is trustworthy (f.e. Hardin 2002). This means, that you can only rely on, or pretend-trust strangers, but never fully trust them. However, there are more convincing arguments that trust in strangers *is* possible: firstly, evidence about someone's trustworthiness is not essential to the attitude of trust. Secondly, there are special forms of trust like Frost-Arnold suggests, which occur between strangers. Thirdly, there are various accounts on which trust in strangers does not contradict the cognitive-based definition: e.g. Holton's participant-stance theory or Jones' theory about affective trust. Thus, I argue that it is possible to trust strangers, but in another way than you trust your friends for example. However: what *is* trust in strangers? I want to offer a definition to show how trust in strangers accounts for two of the paradigmatic examples, how it partly fulfils the desiderata and how it differs from other forms of trust.

Definition T1: T1, or trust in strangers, is one form of trust, that is identified with the following conditions:

a. It is an attitude of trust[15], which is expressed towards strangers (people whom you do not know and have never met before).

b. It is a non-cognitive attitude (a stance).

c. It can be willed for certain non-evidential or non-epistemic reasons[16].

[15] Nota bene, that I do not try to define trust, but trust *in strangers*. Trust here is considered the uncontroversial way I described it in Chapter 1: An interpersonal, three-place relation between a trustor and trustee regarding an action X, with the trustor depending on the trustee to do X and expecting that they will do X; while being entitled to a certain reactive attitude (such as feelings of betrayal, resentment or puzzlement) in case said trust is broken. This also applies to the definitions of T2 and T3.

[16] Like reasons of usefulness or sympathy, or for normative reasons; also in order to spark a certain feeling of proudness or friendship in the trustee.

Now that a formal definition has been established, it is time to see which of the desiderata T1 fulfils. The first desideratum demands that a theory of trust should sufficiently distinguish between trust and reliance. Trust, or forms of trust, which are considered a cognitive attitude are beneficial in having a concrete, quasi built-in, difference to reliance: trust is identified with belief and reliance is not. Since T1 is considered a non-cognitive attitude, the lines are blurred. Even with condition a. stating that T1 meets the demands of the uncontroversial, universal kind of trust, there are no further conditions that make sure that reliance and trust are sufficiently distinguished. Neither condition b., nor condition c. are exclusive to T1, they could also be applied to reliance.

The second desideratum demands a theory of trust to account for trust at will and thereby trust in strangers. This is ensured by condition c. of T1's definition.

Lastly, the third desideratum that a theory of trust has to account for is the impression that in some cases trust seems to go against the trustor's evidence justifiably. As mentioned above, this is a peculiarity to emotional attitudes. Put shortly, since T1 is not considered an emotional attitude, but a stance, it does not meet this third demand. However, this argumentation is the lazy route. That is because one might object that in cases of Frost-Arnold's therapeutical trust (which I identified as trust that can occur between strangers), the attitude of trust *does* go against evidence – as seen in the example of Mrs. Rogers and her neighbor. Because this is an interesting objection I want to take the time to address it: firstly, keep in mind, that Mrs. Rogers and her neighbor are not complete strangers. Maybe they are also not very familiar with each other, but they are certainly not strangers. But what if they were? Then Mrs. Rogers would decide to trust this person to care for her garden, although there is evidence that he should not be trusted because he is a criminal. It seems as if it is possible to construct examples in which T1 fulfils the demands of the third desideratum. However, these cases are exceptions. That is because usually, one does not have any information about strangers (as exemplified in *Library* or *Climbing*). Furthermore, as mentioned at the beginning of this paragraph, the third desideratum is meant to account for cases in which the personal relationship is so strong that it undermines evidence. In the example of Mrs. Rogers and her neighbor it is not their relationship that is very strong but the 'therapeutical' element of Mrs. Rogers' trusting attitude, that makes her go against the evidence. Thus, I conclude, that T1 ordinarily does not fulfil the third desideratum.

To conclude the analysis of T1, I want to touch on the paradigmatic examples. T1 is fitting for assessing the trusting relationship as exemplified in *Library* and *Climbing*. In both

situations, conditions a., b. and c. are fulfilled: trustor and trustee are strangers, the trustors' attitudes are non-cognitive, and they decide to trust the respective trustees. The other paradigm cases contradict at least condition a. Next, the focus will be on another form of relationship, namely on being acquaintances.

4.b. Trusting Acquaintances (T2)

What is considered an acquaintance here, is someone who is not entirely a stranger but also not e.g. your family, best friend or life partner – thus, someone in between. For example, a work colleague, your new neighbors, a friends' friend or the owner of your favorite coffeeshop.

The literature suggests that in most cases, some kind of underlying relationship between trustor and trustee is assumed when looking for a definition of trust. More specifically, trusting situations are often investigated as if this relationship is an acquainted one. This means, that neither strangers, nor family or best friends are considered by default – they usually are the exception to the rule, the counterexample, or the reason why the proposed definition is not sufficient. Trust between acquaintances seems to be favored as the standard form of trust, since it is less mysterious to access than other forms of trust.

What follows is an analysis of trust in acquaintances (T2). Then, I will look at some accounts from the literature to show which are fitting to describe the type of trust between acquaintances. Lastly, some differences to T1 will be highlighted.

When a trustor is putting trust into an acquaintance, they already know something about them. There is some kind of relationship between them. Maybe there is not yet the thickest bond; however, they share enough experiences in order to form a belief about the trustee's trustworthiness. Therefore, if I am in an acquainted relationship with you, it is possible for me to infer whether I can trust you with something, based on my experiences and evidence. I can – at least – predictively trust you (cf. Dormandy 2020).

However, this does not mean that it is impossible to trust acquaintances in a deeper sense. Paul Faulkner (2007), for example, introduces the concept of affective trusting. Faulkner highlights that the underlying relationship between trustor and trustee is the reason for a form of trust, which is affective and not merely predictive. Although he does not define the relationship he has in mind, he states that it is the reason why, within an affective trust relationship, the trustee is responsive to the trustor's dependence. An account like this has its weaknesses when trying to make sense of full, affective trust between a trustor and a

stranger. However, it seems fitting for acquaintances and other relationships (Faulkner 2007). Moreover, it is noticeable that Faulkner's theory also demonstrates how differences in the descriptions of trust have something to do with how trustor and trustee relate to each other. Thus, it supports my claim that trusting differs, depending on the relationship between trustor and trustee.

But now let me come back to the analyses of the three forms of trust. What exactly are the differences between T1 and T2? Most importantly, the differences concern the mental attitude. T1's mental attitude is non-cognitive, T2's is a cognitive one. What follows from this discrepancy, are differences in their voluntariness and the degree of risk involved. Lastly, there they differ in regard to their reactive attitudes in case of the trust being broken.

To be more precise, I – and several philosophers as well (Baier 1986, Hieronymi 2008, Hardin 2002) – claim that the mental attitude behind trusting an acquaintance is a cognitive one: a belief. Beliefs are based on evidence such as experience, perception, memory. However, we neither have any experience with a stranger, nor a lot of perception or any memory of them. That is the reason why T1 was defined as a non-cognitive attitude: a stance. As we do have (at least some) data about acquaintances, it is possible to believe in the trustee's trustworthiness. While e.g. Hardin's, Hieronymi's and Baier's accounts were not fitting to describe trust in strangers, they are to describe trust of this sort.

Furthermore, this entails that one cannot decide or choose to trust an acquaintance[17]. This is because one also cannot choose to believe something. In the stranger case, however, it might be possible, as we have nothing else to base trust on when we do not know this person. As Hawley for example points out: „[n]on-epistemic reasons do not support a belief, but they allow you to decide to trust someone" (Hawley 2014: 2035). With acquaintances, we usually already know, due to their past behavior, whether we can tell them a secret or not – and we cannot just choose to believe that this time they will not spill the beans.

Another, quite obvious, difference concerns the involved risk. T1 involves more risk than T2, for it is harder to assess a stranger's trustworthiness. The trustor knows little to nothing about the trustee in the case of T1 and has to rely on their skill, motivation, kindness or

[17] It is important to mark here that Frost-Arnolds argument concerning therapeutical, corrective and coping trust contradicts my claim. She does not say that it is only strangers whom we choose to trust in these ways (as shown in the example about Mrs. Rogers and her neighbor). In order to accommodate Frost-Arnolds and my account I would suggest considering an acquaintance, about whom there is no evidence, and whom one trusts for corrective, therapeutical or coping reasons, as a kind of stranger. Keep in mind that T1, T2 and T3 are ad hoc classifications and that there might be exceptions to my suggestions. However, I stand by my claim that typically, one cannot choose to trust an acquaintance.

goodwill instead of having any kind of knowledge about them being worthy of trust. However, this is not the case with T2: since trustor and trustee do have some kind of relationship with each other, the trustor has more evidence and safety for assuming the trustee's trustworthiness. Also, only the fact that trustor and trustee have a relationship, is a reason for the trustee to perform X – since they would probably like to continue this relationship. That there is a reason for the trustee to act, is a good reason for the trustor to initially trust them, and also a good reason to claim that trusting someone you know is less risky than trusting strangers.
Lastly, T1 and T2 differ in regard to the distinct reactive attitude, as indicated by the introductory lake examples found in Chapter 3.
Now that I have focused on explaining T2 and stating how it differs from T1, it is time to give a more formal definition, to see whether it can meet the demands of my desiderata for a theory of trust and to assess whether it accounts for the paradigmatic opening examples.

Definition T2: T2, or trust in acquaintances, is one form of trust, which is identified with the following conditions:
a. It is an attitude of trust, which is expressed towards people, with whom one shares an acquainted relation- or friendship, but no deep personal relationship.
b. It is a cognitive attitude (a trusting belief about the trustee's trustworthiness).
c. The reasons for this form of trust are epistemic and concern the truth of the trusting belief.
d. It cannot be willed/decided on voluntarily.

With an established definition of T2, it is possible to assess which of the desiderata are fulfilled. As a cognitive attitude, which is identified with belief, T2 easily meets the demands of the first desideratum, which asks for a sufficient distinction between trust and reliance. While trusting is considered believing that the other one is trustworthy and will therefore carry out X, relying is not considered a belief. When someone A relies on an acquaintance B, then A depends on B but does not think or expect B to do X. Insofar, the attitude of reliance is more similar to the non-cognitive attitude of hope. Nevertheless, T2 does neither fulfil the second nor the third desideratum. Let me first tackle the reasoning why T2 does not meet the demands of the second desideratum. This desideratum calls for adequately accounting for trust in strangers and thereby cases of trust at will. Since T2 is identified as belief, it cannot be adopted at will without committing to doxastic voluntarism. So, trust as cognitive attitude cannot meet the demands of the second desideratum unless

the questionable implication is also accepted, namely that beliefs can be decided on at will. Since I do not accept the implication about beliefs-at-will, T2 does not fulfil the second desideratum.

Secondly, I will analyze whether T2 is able to meet the demands of the third desideratum, which asks for a theory of trust to account for the impression that in some cases, trust, justifiably, goes against the trustor's evidence. Since T2 is defined as a cognitive attitude, this further means – and I will stick to Townsend/Wanderer (2013) here again – that it is tightly tied to the norms of evidentialism. Evidence, or evidential reasons, such as perception or memory, are reasons for trust if considered belief. Therefore, it cannot make sense of cases in which trust goes against the trustor's evidence. This is a peculiarity of emotions, not beliefs. For it is found between acquaintances and not deep, personal relationships, T2 is not considered an emotional, but a cognitive attitude (this will be clearer once T3 is considered). Thus, neither the second nor the third desideratum are fulfilled – T2 only meets the demands of the first one.

Lastly, it is possible to assess that T2 is the most useful form of trust to make sense of two of the paradigmatic examples, *Key* and *Letter*. In both situations, trustor and trustee know each other – they are not strangers but also do not share a very personal relationship. The respective trustors build their trusting belief on evidential reasons and does not go against them. That is the kind of trust, which T2 covers.

4.c. Trusting Friends and Family (T3)

After taking a look at what is particular about trusting strangers and acquaintances, the following section will focus on the case of trust between persons who share personal relationships. These might be people like long-term friends, romantic partners or family.

Because trustors and trustees, who are friends or family, share a lot of history and therefore knowledge about each other's trustworthiness, it seems uncontroversial to suggest that the mental attitude of T3 is belief. However, friends and family do not only have epistemic reasons for trusting each other – but they also have non-epistemic reasons. Imagine, you would have to answer the following question: 'why do you trust your mother with your car?'. The reasons you provide can both be epistemic and non-epistemic. For example, 'Because I know that she will take care', 'Because I did numerous times already', but also 'Because I love her' or 'Because she would be disappointed if I did not'.

That there are epistemic and non-epistemic reasons of trust, is the reason why something other than belief is also plausible for describing a trusting attitude: an emotional attitude. It might not seem tenable to the reader to view the mental attitude of trust as an emotion at first, but there are convincing arguments about the striking similarities between trust and emotion (cf. Lahno 2001). Trust, and especially trust in people whom you are close with, and emotion share the following property: favorably considering the evidence that supports the emotion/trusting attitude (cf. McLeod 2020, Jones 1996, Lahno 2001). I will come back to this similarity below.

Before offering a formal definition of T3, I would like to refer to some works that support the idea, that trusting within a personal relationship differs from other instances of trust. Some of these findings will be reiterations from statements, that I already mentioned in previous chapters. However, now these works will be used to support my claim that trust in friends should be considered an individual form of trust that differs from other kinds of trust.

Many authors, such as Jones (1996, 2019), Hawley (2014) and Lahno (2001, 2020), describe a certain peculiarity to T3: it neglects (some) evidence. Jones, for example, describes the phenomenon of 'affective looping': a trustor will only regard the kind of evidence, which affirms the trustee's trustworthiness, and will dismiss evidence that might prove otherwise. Because of this, she claims that trust can only be defined as an affective attitude and not a belief (cf. Jones 2019). Building on this, Lahno develops his theory of trust as an emotional attitude, which I have already discussed at length.

Katherine Hawley (2014) claims something similar: because of the evidence available within an established relationship, we have epistemic reasons to trust. However, the deeper this relationship is, the more probable it is that there are conflicting non-epistemic reasons which are in contradiction with the trusting belief. It is important to highlight that this is unique to the case of trusting friends and does not concern other instances of trust: "these are reasons which do not apply to beliefs about the trustworthiness of nonfriends" (Hawley 2014: 2034). While knowledge or evidence is considered an epistemic reason, an established relationship like friendship is considered a non-epistemic reason. Often, we trust friends or family so much, that our judgement runs counter the evidence. We take their word over what might be presented as evidence. Both Hawley and Baker argue, contrary to many, that these instances of trust are still reasonable (Hawley 2014, Baker 1987).

This phenomenon of trust being in tension with evidence may also be observed in the reactive attitude of betrayal by good friends, as introduced in Chapter 3. If your belief in

someone's trustworthiness is so strong that it conflicts with epistemic reasons, you will also not believe that your trust was broken: you think that there has to be a better explanation. Below, I will offer a definition for T3, or trust between friends, before I move to an analysis of T3 in regard to the desiderata and paradigmatic examples.

Definition T3: T3, or trust in friends, is one form of trust, which is identified with the following conditions:
a. It is an attitude of trust, which is expressed towards people, with whom one shares a deep, personal relationship.
b. It is a non-cognitive attitude (an emotional attitude).
c. The non-epistemic reasons for this form of trust may compete with evidence.
d. It cannot be willed/decided on voluntarily.

Next, the examination of how T3 compares to the three desiderata will be outlined. Firstly, I will take a look at the first desideratum, which asks for a sufficient distinction between trust and reliance. Since T3 is not considered a cognitive attitude, the distinction is not as clear-cut as in the case of T2. Since both trust and reliance are non-cognitive on this account, we need to dig deeper. With reliance, the relied-on person does not have a reason to carry out the particular action X. Analogously, the relying person has no reason to expect them to do so. As already mentioned, reliance is similar to hope in this regard. However, T3 as an emotional attitude between friends or family always involves some kind of tie between trustor and trustee. This bond is what Lahno for example describes as a feeling of connectedness, which always gives the trustee a reason to carry out X. So, it seems that it is the connection, which follows from a kind of deep and meaningful relationship, that demarcates trust from reliance. However, there is room for doubt: how does trusting friends then differ from relying on friends? The connection does not disappear. This gives reason to conclude that T3 is only demarcated from reliance in cases in which the trustee is not in a personal relationship with the trustor. T3 is then not sufficiently distinguished from reliance on friends and family, only from reliance on strangers and acquaintances. Consequently, the first desideratum is not fully fulfilled.
Furthermore, T3 does not meet the demands of the second desideratum either. The reason for this is that a trustor cannot choose to have this tie or connection with the trustee at will. Therefore, trust in strangers is not properly accounted for.

Lastly, T3 will be examined concerning the third desideratum, which asks for a theory of trust to account for cases in which trust justifiably goes against the evidence. This is fulfilled *per definition*, to be exact, because of condition c. of T3's definition. Condition b. and c. ensure that T3 is understood as very similar to emotion. What they have in common, is that they narrow our perception of said evidence to what we find to be relevant. An emotion controls how we see facts: we tend to perceive those that affirm what we think to be true and negate those that deny what we think to be true. One can witness the same phenomenon with trust: the trustor will only regard the kind of evidence that affirms someone's trustworthiness and will dismiss evidence that might prove otherwise (as e.g. introduced in *Shoplifting*).

Lastly, T3 will be analyzed in regard to the paradigmatic examples: it is suitable for describing both paradigmatic examples *Shoplifting* and *Assault*. In both trusting situations, trustor and trustee share a very close, personal relationship with an already established underlying relationship of trust. This kind of trust cannot be willed but has to be built over time. This trusting attitude is non-cognitive as it is more similar to emotion than to belief, for it is not tied to epistemic reason and may run counter the evidence.

To summarize, what I have argued for in these last two chapters is that trust is not a uniform phenomenon. On the contrary: 'trust' is a polysemous term, which refers to either T1, T2 or T3[18]. These three forms of trust, which I also called 'trust in strangers', 'trust in acquaintances' and 'trust in friends and family', differ from each other – in some regards only minorly, in others substantially. The most significant difference between T1, T2 and T3 is – obviously – the relationship trustor and trustee share. Subsequently, I claimed that each form of trust is identified with a different mental attitude. Another difference concerns the reasons for which trustors engage in trust as well as the question of whether it is possible to do so voluntarily.

Since those differences are major and the conditions for the forms of trust are not coherent, it is disadvantageous to consider only one kind of trust. By analyzing three current theories of trust, I have shown that the main weakness of each of those accounts is that convincing counterexamples can be constructed. For example, Hieronymi's cognitive theory of trust is not exhaustive, since certain instances of trust in strangers or friends are not accounted for

[18] There may be good reasons to explore whether we should consider more than these three forms of trusts, however I will continue to work with only T1-T3.

and are counterexamples to her definition. Likewise is Holton's participant stance theory at the fault of not properly making sense of trusting situations, in which trustor and trustee share a deep, personal relationship for example. Lastly, Lahno's theory about trust as an emotional attitude is also susceptible to counterexamples, since, e.g., it cannot account for trust at will. A detailed analysis of which paradigmatic examples can be considered counterexamples to those theories can be found in Chapter 2[19]. However, I demand that a complete theory of trust can not only account for each of the paradigmatic examples but more essentially, that it meets the demands of all of the desiderata (naturally, these requirements are connected). Accepting that trust is manifold and that there are variants of the phenomenon, moves us further to a point where we can make sense of a much broader spectrum of trusting situations, which differ first and foremost in the way trustor and trustee relate to each other.

To sum up, the consideration of only one form of trust, as the authors Hieronymi, Holton and Lahno in their accounts do, does neither lead to the goal of being able to account for each of the paradigmatic examples nor to the goal of fulfilling each of the desiderata for a complete theory of trust. Thus, only T1, T2 *or* T3 could not meet these demands either. However, as shown, T1, T2 *and* T3 can achieve this. T1 assures that the second desideratum is fulfilled, T2 meets the demands of the first desideratum and T3 accounts for the third one. Together, they fulfil the established desiderata as well as account for each of the paradigmatic examples.

4.d. Advantages

Before concluding this thesis, I would like to take the time to point out the advantages of this newly proposed, alternate understanding of trust. To add further support to my claims I will first refer to the works of Thomas Simpson (2012), who also favors a pluralistic approach to the concept of trust. However, I mainly want to draw the focus on what is unique to my approach: namely, the crucial relevance of the personal relationship between trustor and trustee. Moreover, I also want to touch on the topic of how the divide between T1, T2 and T3 can be useful in the debate about the rationality of trust. In Chapter 1, I introduced some conflicts regarding rationality and trust. In this last subchapter, I will

[19] For a list of more counterexamples, which are not the paradigmatic examples, see Simpson 2012, 552ff.

round off my work by addressing how my approach might be able to resolve these problems.

Simpson observes the same problem as I have described: counterexamples are easily generated for existing definitions of trust (cf. Simpson 2012: 552). The reason for this is that there is not one *single* definition for trust but that trust is an umbrella term for various forms of trust (ibid.: 554f.). While I agree with Simpson on the problem, his idea for solving it is not the same as mine. While Simpson follows the approach of a genealogical account in which he argues that various forms of trust stem from a certain Ur-notion of trust, my solution for the problem highlights different forms of trust. These forms of trust should be considered separately because of the different relationships that the trustor and trustee share. Nevertheless, Simpson's work supports my claims in many ways. I will refer to those similarities in our findings while going into detail about why the developed account about T1, T2 and T3 is beneficial to the philosophical discussion about trust.

First of all, a notion of trust like mine embraces that trust is ubiquitous in our everyday lives and lives up to the actuality that instances of trust are not always the same in terms of mental attitude, risk or vulnerability. It is capable of making sense of trusting situations, that are excluded by other prominent theories of trust and is not as prone to counterexamples as those. Still, it is not an ad hoc solution – my way of viewing trust does not allow for any stance or attitude to count as trust. There are clear conditions regarding what T1, T2 and T3 are.

This further means that my findings acknowledge that trust is not one special, uniform phenomenon that we only share with particular people, such as a romantic partner, but that there are also ways to trust strangers. Only if trust is understood like this, we can make sense of trust as ubiquitous for our everyday societal life or, according to Baier, as "like the air we breathe" (Baier 1994: 98). Since a society is not only made up out of friends and family, but individuals who are strangers to each other, a notion of trust between strangers seems not only helpful but necessary to accommodate for these instances of trust.

Secondly, my suggestion of different forms of trust first and foremost takes into account the manifold relationships between trustors and trustees. That this is a crucial consideration has been advocated by many other philosophers in this field, above all Faulkner, but of course also Hawley, Lahno and Jones. It does not focus on a common denominator between instances of trust and tries to define it thereby. On the contrary, it highlights the differences and by help of them distinguishes subsets of phenomena, which all should be considered a

type of trusting. As such, it includes aspects of trust as viewed in rational game settings as well as those found within emotional, personal relationships. This way of thinking is supported by Simpson. He states that the prominent accounts of trust share a problem: "[e]ach way of thinking takes some relational situation as paradigmatic, and then builds an account of trust around that" (ibid.: 564). This leads us to the current rival theories, instead of one view of trust which acknowledges that trust is not as uniform as it might initially seem.

Thirdly, my distinction of T1, T2 and T3 allows for various reasons to count as reasons for trust. Some advocates of cognitive theories of trust only allow for epistemic reasons for trusting (cf. Hieronymi 2008). Other authors who favor a non-cognitive approach claim that non-epistemic reasons trump or are in conflict with epistemic reasons when it comes to trusting (or at least, that non-epistemic reasons have to be accepted as reasons for trust as well) (cf. Hawley 2014, Frost-Arnold 2014). Not depicting trust as uniform phenomenon but considering different forms of trust allows for accepting different kinds of reasons for trusting. While T1 calls for non-epistemic reasons (as in Frost-Arnold's corrective trust) instead of epistemic ones, T2 requires the latter. Since T1 and T2 are distinct attitudes, this is not incoherent.

Once again, this does justice to how diverse the phenomenon of trust is. Simpson argues that trust comes in degrees (cf. Simpson 2012: 559). I agree: trust between strangers might be more tentative in the beginning, but grows stronger once you get acquainted with someone and is a very tight bond between close friends and family – and each of these stages of trust needs its own analysis.

Lastly, and maybe even most importantly, it could solve the tension between trust and rationality, as introduced in Chapter 1. Along with Townsend and Wanderer, several authors (e.g. Baker 1987) have tried to analyze or solve the problem of rational trust. However, there is no consensus about what rational trust is and whether trust even can be rational. This does not come as a surprise: since there is no consensus about how trust should be defined, why should there be consensus about how rational trust should be defined?

I want to suggest as follows: first of all, I agree with Pederson et al. (2014) in their assumption that trust is a rationally evaluable mental attitude and thus can be rational or not (cf. Pederson et al. 2014: 1953). Then, following Kate Nolfi (2015), I suggest that trust is rational if its mental attitude is rational – but: not every mental attitude is rationally evaluated the same way. Thus, to analyze trust's rationality, one has to know which kind

of mental attitude trust is identified with. However, conventionally, there is no agreement on one mental attitude that accurately describes trust. The focus of Chapter 2 was introducing the reader to the conflict regarding the mental attitude between the cognitive and non-cognitive camp. This conflict was pivotal for the development of my account, in which the term 'trust' can refer to both a cognitive and a non-cognitive attitude. Due to the manifoldness of trust, I suggested not considering trust as a uniform phenomenon which is identified with only one mental attitude. In other words, I claimed that there is no such thing as an 'one size fits all' account of trust. Likewise, I claim that there cannot be such an account of the rationality of trust.

Therefore, I think that there should be different conditions for rationality, depending on the form of trust. Each form of trust is unique and should be examined in its own way when it comes to rationality. Recall that cognitive and non-cognitive attitudes cannot be analyzed in the same manner. Cognitive attitudes are best analyzed by standards of theoretical rationality, while non-cognitive attitudes should be analyzed by standards of practical rationality. My account can offer both these analyses without being incoherent. Since the mental attitudes of T1, T2 and T3 are not the same either, each form of trust can have its own approach. Furthermore, there are various reasons for which trustors trust, and they differ in regard to the kind of trust, that is at play. For example, while the reasons for T2 are epistemic ones, the reasons for T1 and T3 are not.

Summed up, that means that I believe that considering more than one kind of trust entails that there should also be more than one kind of rational evaluation of trust. How rationality is evaluated is then also connected to the kind of relationship that trustor and trustee share. I assume that subsequently T1, T2 and T3 need unique conditions when it comes to understanding their rationality. Developing those at this point would be out of the scope of this thesis, however, I am sure that this could be highly relevant for the discussion about the rationality of trust and move the philosophical research further.

III. Conclusion

To conclude this thesis, I will firstly restate the research question and its answer in a concise and precise way. Subsequently, I will review the key points of the thesis to show how I reached my final statements. This will include mentions of those works from the philosophical literature, that revealed to be the most relevant for my agenda. Moreover, I will show how my work is relevant to the existing debates within the philosophy of trust and highlight the unique and novel features of my own account. Finally, I want to point out any limitations of my work and introduce a few topic recommendations for future research on the concept of trust.

The question, that is leading my work, is: *What is interpersonal trust?* Narrowed down and specified, the research question was: *What are the limitations of current theories of interpersonal trust and how can its definition be improved?* I chose this very question because the philosophical literature on the concept of trust suggests certain flaws. These mainly consist of those theories putting forward definitions of interpersonal trust, which are either prone to counterexamples, not sufficiently distinguishing trust from related phenomena such as reliance or missing an adequate explanation of what rationality of trust is. Despite these issues, I was eager to try to solve them by contributing arguments about where these problems stem from and how the understanding and definition of interpersonal trust can be refined.

My answer to the specified research question consists of four parts which are respectively presented in the four chapters of the thesis' main body.

In Chapter 1, I introduced the reader to six paradigmatic examples of trust. Initially, these were intended to showcase the variety of trusting situations. However, in the subsequent chapters, they also served the purpose of building a body of instances of trust. A body, that in my opinion, a complete theory of trust must properly account for. Consequently, on the basis of the paradigmatic examples, it was also possible to point out certain flaws of existing theories. If the definition of a theory fails to account for one (or more) of the examples, then it cannot be right. Moreover, the examples were also suitable to point out similarities within various instances of trust. By means of these similarities, as well as findings from the literature, I could give a preliminary, uncontroversial first approximation to a definition of interpersonal trust in Chapter 1. To conclude, I introduced the readers to the topic of rationality of trust as well.

Following from there, in Chapter 2, three very distinct accounts of trust from both the cognitive and non-cognitive camp were examined and their advantages and weaknesses discussed. I observed that, between those accounts, there is no consensus about what trust is, therefore I investigated further to see where this disagreement stems from. Each of the authors seemed to have something different in mind. This manifested in the accounts' major common flaw: their susceptibility to counterexamples. To precisely carve out, where these limitations lie, I tested whether the theories could account for all the paradigmatic examples. Since none of them could, I studied the shortcomings I exposed in this second chapter and, based on them, created certain demands on a theory of trust in the next one.

While I was concerned with answering the first part of my research question about the flaws of the current theories of trust in the first two chapters, I focused on a new, improved understanding of interpersonal trust in the last two chapters. Utilizing the findings from the last chapter, I then established three desiderata in Chapter 3. The desiderata are those demands that a theory of trust must meet to do better than the three accounts from Chapter 2. Then, I showed that neither the theories of concern nor any other conventional theory can meet the desiderata. What I considered a conventional theory is any theory that does not acknowledge the variety of trust and tries to find a definition for one, uniform kind of trust. This led me to an important realization: there is not one single form of trust. Therefore, I suggested that interpersonal trust should not be considered a uniform phenomenon. Instead, I claimed the term 'trust' has more than one meaning and refers to various forms of trust. Thus, it should be considered a polysemous expression. This means, that 'trust' is used to describe related, yet different phenomena.

Moreover, as introduced in Chapter 3 and fleshed out in Chapter 4, I explained that it is the relationship between trustor and trustee that impacts which form of trust the term 'trust' actually refers to. Therefore, I suggested three forms of trust: trust in strangers (T1), trust in acquaintances (T2) and trust in friends or family (T3). I explained my motivation for this new understanding and described how T1, T2 and T3 differ from one another. Subsequently, I gave formal definitions for each of the forms of trust and illustrated that T1, T2 and T3 together can not only meet the demands of all the desiderata but also account for all the paradigmatic examples from Chapter 1. In this way, a comprehensible and more accurate approach to trust, that is acknowledging the variety of the phenomenon, was created. I ended my work with some statements about why my understanding, which denies trust as a uniform phenomenon, is advantageous for the philosophical research on trust. Lastly, I sketched an account about the implications that my suggestions have on the

rationality of trust. I claimed that dismissing trust as uniform and acknowledging T1, T2 and T3 as the three main forms of trust could be beneficial to solving the tension between trust and rationality, as introduced in Chapter 1. However, these thoughts could not be developed into a full argument within the scope of this thesis. Thus, they remained an interesting outline that could turn into a substantial account of rationality of trust one day.

To sum up my answers to both the broad and the specific research question, my findings reveal that there is not one phenomenon that is interpersonal trust, but that there are three forms of interpersonal trust. The term 'trust' may refer to each of the three forms.

More specifically, I have argued throughout this work that the current theories of trust misconceive trust as a uniform phenomenon and are thus susceptible to counterexamples. On the contrary, I suggested a pluralistic approach, because trust should be regarded as a polysemous expression. Thereby, it is appreciated that the term trust may refer to three different forms of trust. Which form of trust is at place in an instance of trust, depends on the kind of relationship between trustor and trustee.

Next, I want to review the key points that were made throughout this thesis. Firstly, I challenged conventional ideas about the phenomenon of trust. These are characterized by the belief that all instances of interpersonal trust fit one single mold and can be defined in the same way. For example, most of the accounts, but especially those by Pamela Hieronymi (2008), Richard Holton (1994) and Bernd Lahno (2001, 2020), which I focused on throughout my work, define trust as either a cognitive or non-cognitive attitude. However, my understanding allows for trust to take on several forms.

This leads me to the second and probably most important key point of the thesis: I observed that trust differs, depending on the kind of relationship trustor and trustee share. For example, trust between strangers differs vastly from trust between friends with regard to the mental attitude, voluntariness or the reasons of trust. By identifying trust as a polysemous expression I am doing justice to the undeniable variety of trust.

Thirdly, an analysis of the current research indicated that present theories of trust offer definitions which can be invalidated by means of counterexamples. My suggestion of trust not being a uniform phenomenon allows for a broader range of trusting situations. Thereby, I avert the risk of the theory being susceptible to counterexamples. Support for the claims regarding this topic is drawn from Simpson (2012), who is one of the few that also favors a more pluralistic approach of trust.

Furthermore, as a fourth key point, I underlined the impossibility to spell out one specific 'reason for trust'. Reasons for trusting differ vastly – especially across different forms of

relationships. An accurate understanding of trust must account for not only a variety of trusting attitudes but also a variety of reasons for trust. This is also connected to the topic of rationality of trust. One observation I was not able to examine properly within the scope of this thesis (and should thus be taken with a grain of salt) is that it seems as if we should consequently consider different understandings of rationality, depending on the form of trust. Problems concerning the reasons for trust and rationality of trust have been taken up by Townsend/Wanderer (2013), Baker (1987) or Hawley (2014).

Lastly, this thesis also aimed at showing how essential it is to spell out what the requirements for a theory of trust are. Without the establishment of the desiderata in Chapter 3, it would neither have been possible to emphasize what a theory has to account for, nor to pinpoint where the flaws of the conventional theories lie.

Next, I would like to take the time to highlight why my work is relevant to the philosophical discussion since naturally, this thesis aimed to drive the research on trust forward. Besides that, I will spell out what is unique and novel about my account.

First of all, I contributed an innovative analysis of the concept of trust. While other concepts, such as knowledge or truth have been at the center of philosophical attention, trust has been overlooked for a long time. This is surprising since it plays such a crucial role in our lives; not only within relationships but also within society. My research thus is not only of relevance for the particular examination of the concept of trust but also for understanding the dynamics of relationships – both on a personal and a more societal level.

However, most significantly, I contributed an approach that unites two things that have been neglected by the current research on interpersonal trust. On the one hand, this is the insight that trust is not a uniform phenomenon but comes in different forms. The current research tends to conceive all trusting situations as instances of the same phenomenon, which can be defined explicitly. On the other hand, this is the pivotal role that the kind of relationship between two people plays when it comes to trusting. Although it seems evident, that the relationship between trustor and trustee is of relevance for an analysis of trust, it has been neglected by many. Only a few highlighted the importance of both these factors – and even more surprising: no one noticed that they are connected. The connection manifests as follows: as stated, trust comes in various forms; and these forms of trust are exemplified by the relationship between trustor and trustee. Thus, there is trust in strangers, trust in acquaintances and trust in friends and family and each of these forms of trust is defined individually.

What is also of relevance to add is that my work also constitutes a useful basis for further research on the rationality of trust. As already established throughout this thesis, we are confronted with a tension between trust and rationality. By acknowledging trust as a polysemous expression that refers to different forms of trust, the clarification of this problem can be driven forward and thereby, this tension between trust and rationality may be resolved.

I want to end this concluding chapter by discussing two limitations of my work and suggesting recommendations for future research on this topic. Firstly, further discussions of the topic need to determine the effects this work has on the discussion of rationality of trust. My own work fell short of giving a convincing argument for the implications of my research for this particular topic. Thus, I am of the opinion, that an ensuing discussion on rationality would be a relevant and fruitful one.

Secondly, while this thesis clearly illustrates that there are three forms instead of one uniform kind of trust, it also raises the question of whether these three forms of trust are sufficient. I did not elaborate on whether the acknowledgement of another form of trust could be advantageous to the discussion. Trust in strangers, acquaintances and friends or family were preliminary divisions that turned out to be sufficient in order to present this new kind of understanding trust that I had in mind. However, it might be the case that the term trust refers to more than these three forms of trust, that I have not been aware of during my research.

IV. Literature

Acton, H. B. (1974) *The Idea of a Spiritual Power*, London: Athlone Press.

Audi, R. (2004) "Theoretical Rationality: Its Sources, Structure, and Scope," in A. Mele and P. Rawling (eds.), *The Oxford Handbook of Rationality*, New York: Oxford University Press.

Baier, A. (1986) "Trust and Antitrust," *Ethics* 96(2): 231-260.

Baier, A. (1994) *Moral Prejudices: Essays on Ethics*, Cambridge, MA: Harvard University Press.

Baker, J. (1987) "Trust and Rationality," *Pacific Philosophical Quarterly* 68: 1-13.

Castelfranchi, C. and Falcone, R. (2020) "Trust: Perspectives in Cognitive Science," in J. Simon (ed.), *The Routledge Handbook to Trust and Philosophy*, New York & London: Routledge.

De Sousa, R. (1987) *The Rationality of Emotion*, Cambridge, MA: MIT Press.

Dormandy, K. (ed.) (2020) *Trust in Epistemology*, New York: Routledge.

Faulkner, P. (2007) "On Telling and Trusting," *Mind* 116(464): 875-902.

Faulkner, P. (2011) *Knowledge on Trust*, Oxford: Oxford University Press.

Frost-Arnold, K. (2014) "The cognitive attitude of rational trust," *Synthese* 191(9): 1957-1974.

Gambetta, D. (ed.) (1988) *Trust: Making and Breaking Cooperative Relations*, Oxford: Basil Blackwell.

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

Hawley, K. (2012) *Trust. A Very Short Introduction*, Oxford: Oxford University Press.

Hawley, K. (2014) "Prejudice and Partiality in trusting," *Synthese* 191(9): 2029-2045.

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

Hinchman, E. (2020) „Trust and Will," in J. Simon (ed.), *The Routledge Handbook to Trust and Philosophy*, New York & London: Routledge.

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(1): 4-25.

Jones, K. (2019) “Trust, Distrust, and Affective Looping,” *Philosophical Studies* 176(4): 955-968.

Kappel, K. (2014) “Believing on Trust,” *Synthese* 191(9): 2009-2028.

Keren, A. (2014) “Trust and Belief: A Preemptive Reasons Account,” *Synthese* 191(12): 2593-2615.

Lahno, B. (2001) “On the Emotional Character of Trust,” *Ethical Theory and Moral Practice* 4(2): 171-189.

Lahno, B. (2020) “Trust and Emotion,” in J. Simon (ed.), *The Routledge Handbook to Trust and Philosophy*, New York & London: Routledge.

McGeer, V. (2008) “Trust, Hope and Empowerment,” *Australasian Journal of Philosophy* 86(2): 237-254.

McLeod, C. (2020) “Trust,” *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/trust/>.

Nolfi, K. (2015) “Which Mental States are Rationally Evaluable and Why?,” *Philosophical Issues* 25(1): 41-63.

Pederson et al. (2014) “Rational Trust,” *Synthese* 191: 1953-1955.

Simon, J. (ed.) (2020) *The Routledge Handbook of Trust and Philosophy*, New York & London: Routledge.

Simpson, T. (2012) “What is trust?,” *Pacific Philosophical Quarterly* 93: 550-569.

Strawson, P.F. (1974) “Freedom and Resentment,” in P.F. Strawson, *Freedom and Resentment and Other Essays*, London: Methuen.

Townsend, L. and Wanderer, J. (2013) “Is it rational to trust?,” *Philosophy Compass* 8(1): 1-14.

Williams, B. (1988) “Formal Structures and Social Reality,” in D. Gambetta (ed.), *Trust: Making and Breaking Cooperative Relations*, Oxford: Basil Blackwell.

V. Appendix

Abstract

It is hard to imagine a world without trust. Our relationships, and even society as whole, is interwoven with the attitude of trust. Yet, we are missing consensus on how to approach this phenomenon. This thesis introduces a novel understanding of trust, that is closely concerned with the particular relationship that a trustor and a trustee share. By means of analyzing the shortcomings of certain other theories of trust, I establish a list of desiderata for a comprehensive theory of trust. By formally showing why these other theories cannot meet them, I can carve out the main problem: trust is conventionally considered a uniform phenomenon. Thus, in contrast, my account identifies 'trust' as a polysemous expression that refers to more than one form of trust. These forms of trust relate to the kind of personal relationship between trustor and trustee. I claim that the mental attitude of trust is different depending on the kind of shared relationship. I distinguish between three main forms of relationships and thereby, three forms of trust, namely trust in strangers, trust in acquaintances as well as trust in friends and family. I contend that each of these forms of trust should be analyzed separately and offer definitions for each of them in the end. Thus, I argue for two main claims: not only that we should dismiss the idea of trust as a uniform phenomenon and acknowledge various forms of trust, but also that the relationship between trustor and trustee determines which form of trust is present.

Zusammenfassung

Eine Welt ohne Vertrauen ist kaum vorstellbar. Vertrauen zieht sich nicht nur durch unsere persönlichen Beziehungen, sondern auch durch unsere Gesellschaft im Ganzen. Obwohl das Phänomen omnipräsent ist, gibt es keine Übereinstimmung darüber, wie es zu verstehen ist. Diese Arbeit stellt ein neuartiges Verständnis von Vertrauen vor, welches besonders die persönliche Beziehung zwischen den Vertrauenden und den Vertrauten miteinbezieht. Es werden einige Vertrauenstheorien vorgestellt sowie kritisiert und, anhand deren Schwächen, eine Liste von Desiderata erstellt, welche eine umfassende Vertrauenstheorie erfüllen muss. Indem ihre Unzulänglichkeiten in Hinsicht auf die Desiderata aufgezeigt werden, kann das Hauptproblem dieser bestehenden Theorien herausgearbeitet werden: Sie fassen das Phänomen Vertrauen konventionell als ein einheitliches auf. Im Gegensatz dazu bestimmt diese Arbeit den Begriff Vertrauen als polysem, welcher mehrere Bedeutungen hat und auf mehr als eine Art von Vertrauen referiert. Unterschiedliche persönliche Beziehungen gehen mit unterschiedlichen geistigen Haltungen von Vertrauen einher. Es werden drei Arten von Beziehungen und deswegen auch drei Arten von Vertrauen unterschieden, nämlich Vertrauen in Fremde, Vertrauen in Bekanntschaften und Vertrauen in Freund*innen und Familienmitglieder. Individuelle Analysen dieser Arten von Vertrauen erlauben, dass schlussendlich auch drei unterschiedliche Definitionen vorgeschlagen werden können. Somit wird für zweierlei argumentiert: Einerseits, dass die Idee von Vertrauen als einheitliches Phänomen aufgegeben werden sollte und unterschiedliche Arten von Vertrauen erkannt werden müssen. Andererseits, dass es die persönlichen Beziehungen zwischen den Vertrauenden und Vertrauten sind, welche ausschlaggebend dafür sind, welche Art von Vertrauen vorliegt.