4-30-2009

The Role of Trust in Judgment

Christophe Sage Hudspeth
University of South Florida

The Role of Trust in Judgment

by

Christopher Sage Hudspeth

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Philosophy
College of Arts and Sciences
University of South Florida

Co-Major Professor: Stephen Turner, Ph.D.
Co-Major Professor: Brook Sadler, Ph.D.
Hugh LaFollette, Ph.D.
Sandra Schneider, Ph.D.
Michael Gibbons, Ph.D.

Date of Approval
April 30, 2009

Keywords: Ethics, Decision-making, Risk, Cooperation, Confidence

To my mother, whose support made this possible.

Table of Contents

The Role of Trust in Judgment

Christopher Hudspeth

ABSTRACT

In this dissertation I defend five claims about trust: 1) trusting and trustworthiness are conceptually but not causally connected; 2) trust is risky; 3) trust requires good will; 4) trust is a two-part relation; and 5) trust is an interpretative framework.

A concern for trust often appears in discussions about testimony and the expectation of truthfulness; Bentley Glass, John Hardwig, and Jonathan Adler each address the role of trust in science while assuming a necessary connection between trusting and trustworthiness. I argue that this conception is untenable because a justification for one fails to suffice as a justification of the other. I show, instead, that it is our assessment that links trustworthiness to trusting.

My second claim, that trust involves risk, is contentious. I argue that the common understanding of risk as harm, which is held by Russell Hardin, contains the more technical understanding of risk as uncertainty, which is suggested by Niklas Luhman. Trust is risky precisely because it is inherently uncertain.

Annette Baier argues that failing to recognize a distinction between trusting and action associated with that trust leads us to accept accounts of trust that are too broad. I argue that trust is to believe that another is concerned with your well-being, has good

will, because she is concerned about you and not that it will necessarily benefit her in some way.

Seeing trust as good will allows me to advance my fourth claim, that trust is a two-part relation, because trust is the assessment of another's good will towards me. Trusting another is believing that the other's good judgment will include concern and consideration for my interests because they are my interests.

My fifth point follows naturally once trust is understood to be a risky, two-part relation requiring good will: it is an interpretive framework. Our trust in another sets the tone for understanding her behavior such that we take her actions as either supporting or blocking our interests. This account takes the capacity to trust to be the foundation for, rather than merely an outgrowth of, our judgments.

Chapter One: The Meaning of Trust

Introduction

What does it mean to trust? This is the question of my dissertation. It sits at the core of my investigation and informs the approach. My concern, thus understood, is descriptive rather than prescriptive even though there can and will be prescriptive claims made based on the description that I present. I want to be clear about this goal in the hopes of avoiding misunderstanding. Because I am attempting to describe rather than define trust it will be possible to present cases that seem to contradict my position. That is, counterexamples can be offered that challenge a prescriptive account but which, rightly understood, are no threat to a descriptive account. For example, were I to define trusting simply as *allowing another to act in your interests* then one could challenge that definition by pointing out cases of trusting where it would be impossible to allow another to act in your interests. On the contrary, by describing trust as *allowing another to act in your interests* then the cases in which you cannot allow another to act in your interests would, quite simply, not be trust at all but rather something else. This might seem to be a kind of intellectual sleight-of-hand, especially since prescribing is generally understood as the act of delineating the scope of the conversation, but I do not intend it that way. Instead, by making this claim I mean only to reinforce that I am chasing after what it means to trust rather than attempting to show when one should or should not be trusting.

A prescriptive account would run afoul of the intricacies of life and would confound my position.

Certain cases, for all that they appear to be trust, are simply not; a criminal mastermind who gives you the keys to Fort Knox after he has irradiated the gold inside might appear to be trusting you not to steal the gold but he is not, in fact, trusting you. Likewise, the Nigerian who needs only your bank account number in order to secretly move $60,000,000 out of the country seems to be trusting you with a fortune but is, in fact, not trusting you at all. The parents that leave you to babysit their child appear to be trusting you to watch their child but the video cameras hidden throughout the house tell a different story. The goal, then, of my dissertation is to consider those cases when one is actually trusting (or being trustworthy) and not simply the cases in which one appears to be trusting (or being trustworthy). I assume that a person is trusting and then investigate what makes that case special. On might object that by targeting the investigation in this way that the answer to the question is assumed but, I believe, the arguments that follow sufficiently make my case and avoid a vicious circularity.

I begin with what I take to be a simple and, in some sense, uncontroversial claim about the relationship between trusting and trustworthiness: they do no entail each other. While this claim may be widely accepted as obvious the conclusions that are drawn from it are not. The lack of entailment (or a necessary connection) means that acts are not sufficient to explain trust. Trust can be betrayed, failed, and misplaced. We can trust the unworthy and distrust the worthy. Any particular action is not, therefore, a case of being either trustworthy or untrustworthy. And yet, though I argue that the two do not entail each other, that they do not cause each other, there is an inseparability about them.

Trusting does not cause another to be trustworthy and being trustworthy does not cause another to be trusting but being trusting does mean that one believes another is trustworthy. Believing that another is worthy of trust just is what it means to be trusting. From my initial claim that trusting and trustworthiness do not entail each other follow four others: 1) trusting is to believe that another is trustworthy; 2) trust is risky in that it is uncertain and makes one vulnerable; 3) believing that another is trustworthy is to believe that she has a good will towards you; and 4) trust is a two-part relationship consisting in one who trusts and one who is trusted. These four claims lead me to conclude that trust is a framework for interpreting the past, present, and future actions of others, for understanding others. Before I outline the plan for the following chapters I want to address the problem differently. That is, I want to talk about the problem in very practical, everyday terms. I want to retell an experience in order to show why the investigation of trust is important.

Description

Recently, I needed to take my car into the shop because it had stopped running. Although, to be honest, I do not own a car at all, I own a truck. My truck is nearing its 20th birthday, so the fact that it needs repairs comes as no surprise; I can hear it now, rusting away in the driveway. Truly, it needs a complete overhaul, or possibly even a rebuild of the entire engine, at the very least it needs a tune-up. But I lack the funds for such an extensive project. My truck, therefore, must suffer through the piecemeal replacement and repair of only those parts that are most desperately in need, which generally amounts to fixing only those that are too broken to function any longer, which

is how I came to be looking for a shop the other day. To add insult to the injury of this humiliating process, I think that calling it a car rather than a truck, an affectation that I find as amusing as it is useless, in this story is much simpler. My car, then, needed to go to the shop.

Unfortunately, at the point when my car finally stopped running altogether and *needed*, rather than just desired, to be taken to the shop, I had only recently moved to Decatur and, being new to the area, had relatively few resources for locating a shop that both would be reasonably priced and would perform reliable repairs. Turning to the yellow pages is surprisingly unhelpful in doing anything other than simply providing the names, locations, and phone numbers of mechanics; Honest Abe's Reliable Automotive Repair might be a catchy name meant to inspire confidence but it does not, in actuality, provide any indication of the quality of the work done there; claiming to be honest is not, itself, an indication that one is in fact honest. I was presented, therefore, with a problem that should be familiar to anyone who has ever moved to a new town: I was not sure who would treat me fairly. My situation, it occurred to me, was a perfect example of the dissertation that I had been working on. That is, my search for a mechanic provides a familiar touchstone in a discussion of trust. The everyday experience of trying to figure out to whom I should take my car for repairs makes a useful introduction to a consideration of trust and so here we are: me, reflecting on how I decided where to take my car, and you, reading my reflection.

Who is a trustworthy and trusting? What is the nature of trust? When is it appropriate to trust another? Where is trust located? Why is trust a good thing? How is trust recognized? These are all questions that find relevance in choosing a mechanic.

Which mechanic would be trustworthy? How could I know whether to trust any particular mechanic? Would there be a way to ensure the success of my trusting? Could I protect myself and my car from betrayal? Can I choose to trust the mechanic or is that beyond my control? Conversely, can the mechanic choose to be trustworthy? All are questions worthy of answers. I use this quasi-fictional case of trying to locate a mechanic in order to preface and explain the position that follows. To present this case, I will follow, in form if not in direct argument, the plan of the dissertation but, insofar as this is a story and not an argument, I will not present a series of premises that lead to a conclusion. Instead, I will tell it like I might tell a friend.

Separation

Obviously, the first place I should start is with the problem. My car, the poor thing, has had a number of troubles of late. Leaks in the exhaust system, broken oxygen sensors, and fouled plugs were just some of the obvious mechanical problems that my car was having. On the whole, it simply was not running well. Things had gotten so bad that the car barely had enough horsepower to pull itself up a hill. And, on a drive of any decent length, the car would begin to overheat, which, for some reason, caused the oil pump to cut out, which led to no oil being circulated and that is a problem that can kill your ride. The car had been to see a mechanic a few months earlier but that shop found nothing obviously wrong. It left having gotten a tune-up but not much better off. It was drivable but needed a lot of tender care. Whatever was ailing my vehicle had gotten worse in the intervening months until now it could barely make it to the store. Finally, it

died a mere two blocks from my house. I had no choice but to have some serious work done, work that I could not do myself. I had to put my car into the care of another.

But where should I take my car? As I mentioned, I was new to the area and did not know of any good mechanics. Had I known any regular people in town, I might have been able to ask for advice. Unfortunately, everyone that I knew was a recent transplant as well. This meant that I was on my own. In other circumstances, I might have taken it to the dealership. There is a certain uniformity of service that comes from being employed and trained by the same company that manufactures the car. Sure, the dealership is generally more expensive but, in the scheme of things, the extra cost is worth the comfort that comes from predictable service. However, I was a little tight on cash, so the extra cost would have been a real problem when I needed my car fixed. And, mechanics at the dealership are often constrained by rules about how they can affect repairs; they must repair or replace parts rather than jury-rigging or removing them. Besides, the only dealership was located on the far side of town; not the easiest place for me to get to with my car broken. Instead, I wanted to find a local mechanic.

Finding a local mechanic, however, is not easy, least ways not finding a particularly good one. Lacking a recommendation I turned to the internet for help. I searched for reviews of the various shops in my area. Unfortunately, I found that this was not much help either. To be sure, several shops were eliminated right off the bat but generally speaking the problem with online reviews is that they are testimonials. Testimonials, unless you know something about the person giving the testimony is really no different than taking the name of the shop itself as some kind of indication. For all I knew, the folks writing the positive reviews were the friends of the shop owner. But that

is not the only problem with testimonials. Suppose that each testimonial is true, has a positive testimonial really given me any information that could be used to make a decision?

A single positive review is only an indication of what the mechanic has done in the past, it is not a guarantee that the mechanic will perform that way in the future. Having done good work in the past does not entail doing good work in the future. This works just as well for negative reviews as positive reviews. A single failing review is not, by itself, enough to mark a shop as being unreliable but neither is a single positive review enough to mark the shop as reliable. Certainly, some preponderance of reviews will give an indication of the general quality of work but even if the shop had hundreds of positive reviews it might turn out that in my case, the shop would fail to do a good job. Sadly, it is the nature of the beast. So, while it might help a bit to consider the reviews, in the end the reviews would not be able to decisively make the decision. Something else had to be considered.

Risky

One reason that making a decision was even an issue is because the knowledge that I had about the various mechanics was uncertain; it came from sources that I could not verify and related to concerns that I had no knowledge of. When the shop failed to fix the high-pitched whine that someone complained of, was it because of a failure on the part of the shop or on the owner. Did the leak in the engine really start only after the car had come in? Was that miraculous repair that was testified to really as extensive a job as the mechanic made it out to be or was it completed by simply turning a knob?

It turns out that there is a great deal of uncertainty in finding a mechanic. Online searches could only help me so much. Any repair would be expensive; I could lose a lot of money, which is certainly a harm. But, in addition to the cost, I could lose a lot of time if the repairs were not done efficiently. And, if the shop were lackadaisical they might make faulty repairs or even fail to notice things that were important; that is, they might fix a superficial problem while leaving an underlying issue to fester, which might result in the total loss of the car. The uncertainty about the shop and its employees would leave me vulnerable and so I needed a kind of reassurance, one that I could not get from looking at reviews online. I needed to actually talk to the folks at the shop.

I know a little something about cars. I do not know as much as mechanics know about cars but I know more than the average person who happens to see you broken down on the side of the road and stops to provide help. I appreciate the gesture, I really do, but I have, indeed, checked to make sure that the battery cables are not loose. I turn to mechanics when either I cannot solve the problem or when, having figured out what is wrong, I realize that I lack the tools or the skill to fix the car. Talking to a mechanic gives me an idea of how the mechanic will treat my car, which might make me a bit of an expert but I am convinced that one does not have to be an expert to learn something about the shop from talking to the mechanic.

Goodwill

It is not that I want to make sure that the mechanic knows about automobiles when I talk to them. I take it as a given that mechanics know their field. No, I like talking to the mechanic because it is important to me to learn how the mechanic responds to

questions. I want to know how the mechanic will treat the customer. If the mechanic is dismissive or evasive, if she ignores my questions or tries to distract me with other issues or concerns, I become immediately wary. Thankfully, the shop that I called did not treat me as simply a means of generating income. I choose this particular shop as the first one to call because they happened to be the closest shop, I was able to walk there. But, they are also generalists; they work on everything from exhaust systems to suspension issues and they do it on both foreign and domestic cars.

When I called I asked first about whether my problem was one that they could handle and then I asked what kind of time frame they would need to look at it. I was told "yes" and "now" in response. This was a good start. After they had a chance to look at it, they were able to tell me what they thought was wrong with the car. I asked questions. I asked about alternatives. I asked about what would have caused this problem. I asked how much had to be repaired right away and how much could wait. The mechanic was patient with me; he took the time to answer my questions and treated each as if it was a question worthy of answering. But perhaps most importantly, when we started talking about the extent of the repairs and the costs he made it clear how we could work out the problem in the least financially painful way possible.

What the mechanic was showing me was that he cared about me. My car was broken and at his shop, he could have told me that it needed all kinds of repairs. He could have claimed that parts whose names I did not know were bent or broken. He could have said that everything that was broken was absolutely necessary. But he did not do any of this. Instead he told me about ways that we could manage short term fixes and ways that we could avoid large repairs altogether. Even though he gets paid to repair stuff, which

gives him incentive to repair as much expensive stuff as possible, this mechanic was helping me take care of my interests perhaps even at his own expense.

I know what you are thinking: a good salesperson knows when to sell hard and when to sell soft, she knows when she can get a long term benefit from a short term loss. You think that the mechanic was giving me a good deal now in order to ensure that I returned to the shop for my future repair needs. Admittedly, this would explain the behavior of the mechanic but it would do so only so long as my car needed repairs or so long as it was financially cost effective to spend time working on my car. Certainly the mechanic could have been hoping that I would spread the good news of his forthrightness. But this kind of motivation would only get him a superficial kind of concern for me and my car; it would not cause him to care about me. I could not be certain, of course, but I had the impression that the mechanic cared about my interests not because they coincided with his interests or because it was in his interest to consider my interests but instead because he was a good and decent person. He cared about my interests because they were my interests, in this case because they were the interests of a human being. This mechanic was a good person and he had shown good will towards me. He wanted me to have a well running car at a reasonable price and that was good thing.

Two-part

Because the mechanic had such good will towards me I was willing to do more than simply let him repair the particular thing that I knew was wrong with my car. My car was not starting, I knew that. It had lost horsepower and its exhaust smelt of unburned fuel, I knew that. The repairs for that would require the replacement of a few parts, I

knew that as well. Had the mechanic only been concerned with what I would pay him for affecting those repairs he would have been inclined to find more things to fix and there were definitely more things that needed fixing on my car. But at least one of the things that the mechanic found wrong with my car he fixed without either telling me in advance or charging me for. And one thing that he fixed was to remove a part that was, I am told, unnecessary.

Had I thought only about the mechanic in regards of the particular repair in question these latter two repairs would have caused me much disappointment. I had not left my car in his care so that he could remove parts – I had left it so that he could replace parts. And, I had not left the car in his care so that he could replace things that were unrelated to what was currently broken – I had left the car in his care for specific repairs. But because I believed that he meant well for me and my car, I accepted that these things were also good for my car. In fact, I thought so highly of his opinion that I asked about some other repairs that I knew my car was in need of. Much to my surprise he told me that he could not do the work in question. His concern for me and my car had led him to admit that there was a point at which he would not be able to help me. This is just the kind of thing that I expect from someone who is really concerned with my interests. In believing that he had a good will towards me it was not enough to simply think that under a certain set of conditions, in regard to some limited set of objects or events, that he would be willing to help me care for my interests, I had to think that he would help me even if that meant telling me that someone else would have to do the work. My belief in his good will is a belief about our relationship.

Interpretive

At the end of all this I showed up to retrieve my car and it was not ready. Strangely, I was not disturbed by this. I could have thought that they had lied to me, or maybe just mislead me. I could have thought that they had been honest in their assessment but then failed miserably to do the agreed upon work. But I did not think any of these things. Instead I assumed that the work had taken a little longer than anticipated. Sure enough, it had. But it had taken longer because another issue had arisen – this is the one that they fixed on their own discretion. The work that they had done and the impression that I had of them had informed my assessment of their failure to get the work done in a timely fashion. I imagined those folks who had posted negative reviews. Had they been as positively inclined as me when they arrived at the shop to get their cars or had something happened to lead them to think the worse? I worried that my positive assessment was a general condition, that I might just think the best of everyone. I suspect that I am more inclined to think positively than many people but certainly not all the time and certainly not more so than any other people. But interestingly, in this case, my overall opinion of the mechanic importantly affected the way that I understood the tardiness of the repairs, which in turn re-informed my overall opinion of the mechanic.

Had I been thinking that the mechanic was deceiving me then when I arrived to find my car still on the jacks with the hood open and parts still lying about, I would have taken that to be further evidence of the shoddiness of the shop. I would even take the mechanic's earlier assessments of when the work would be completed as either purposefully misleading or ignorantly uninformed. But instead I let it slide that the car was not ready when I had been told that it would be and I took their affecting unasked for

repairs to be a further sign of their concern for me. My interpretation was reflexively informing my belief in the good will the mechanic had towards me. Had the car broken down immediately after I left the shop or had I discovered that the missing parts were valuable on the black market than my opinion of the mechanic would change and my interpretation of all those events would change as well. For now, however, I am pleased with the work done there and will return there when my car needs more work that I cannot avoid.

Terminology

Thinking about when I went looking for a mechanic reminds me that the terminology used in talking about trust can often be misleading. Trust, trusting, trustworthy, and trusted all seem to apply to the same thing while often referring to different things and in different senses. *Trust* can be used to describe the duo of trusting and trustworthiness or it can be used as a kind of demand, as in "I trust that you will be home at a reasonable hour." It can be used to refer to things about which we cannot seriously believe have the capacity to be trustworthy, as when we say "I trusted the chair to hold my weight when I stood on it." And it can be used to describe both an act and an attitude, as when we say either "I am trusting you with the keys to my car" or "my trust caused me not to question your motives." In short, there is a great deal of confusion about the usage of the term. The first distinction that I want to make, one that is, in fact, argued for in Chapter Two, is that there is a difference between trusting and entrusting. I use "entrust" to refer to the action of turning over control while I reserve "trust" to describe particular mental attitude. I therefore avoid using derivatives of trust that make it into a

kind of action, like the example which treats it as a demand or expectation; when we use trust to describe what we are doing instead of the reason that we are doing it, we fall into talking about trust as a kind of action, which I am trying to avoid. Further, I focus only on interpersonal relationships, the reason for which becomes obvious in Chapter Four: I argue that trustworthiness requires a particular motivation, which is only attributable to people. But, I expect that, no matter how hard I try, there will be some slippages in usage within the text. I beg the reader's indulgence on this score, old habits, especially ones that are so often repeated in our daily lives, are hard to break.

Outline

This brings me to a discussion of the plan for the dissertation. The first chapter has been divided into four sections: the introduction, the descriptive story, the terminology, and now the outline. The outline is basic, meant only to give some idea of where the dissertation is heading, where various parts of the argument can be found. But, I do want to make one suggestion about the way the chapters and the argument are organized. First, I treat each chapter as itself a premise in the larger argument. But, more interesting at this stage is the particular breakdown. Chapters Two and Three essentially limit the scope of the investigation by arguing against particular ways of understanding trust. They do each provide a positive account but they set up reader for the claim that is made in Chapter Four. Chapter Four is the marrow of the position; it is center of my argument. It is, truly told, the only chapter in which I seek to make a direct and specific claim about the nature of trust. Chapters Five and Six are consequences of the position defended in Chapter Four and follow from the claims made in the previous chapters. So,

first I tear down our conception of trust in order to present a position that I then expand upon. Before I begin in earnest, a more thorough explanation of what is to come is in order.

Chapter Two

In Chapter Two I argue that the actions associated with trusting and being trustworthy are not themselves cases of trusting or being trustworthy. From these arguments I conclude that the trusting and being trustworthy are not actions and further that trusting is to believe that the other is worthy of trust. The position begins by recognizing that one can fail to act in a trustworthy manner either intentionally or unintentionally and that in each case that action may or may not result in the trusted person being considered to have broken the trust that marked that one as trustworthy. Put simply, trusting does not entail acting in a trusting manner and being trustworthy does not entail acting in a trustworthy manner. Conversely, failing to be trustworthy does not entail being distrusted and failing to trust does not entail untrustworthiness. This lack of connection is obvious, simple even, but the ramifications are significant.

Because trusting and trustworthiness do not entail each other it is a mistake to assume that any particular case of one tells us anything about the other. It does not even tell us anything about the individual who is doing either the trustworthy or trusting action. The actions that we engage in can be motivated in many different ways and not all of them correspond to the ways of trust. Chapter Two's goal is to make this point, to separate action from attitude, and to show that, once the separation is made, a new connection can be found: believing that the other is worthy of trust is trusting the other.

If we hold to the idea that trusting is an action it appears as if there is a separation between these two things. It appears that I can believe that the other is worthy of trust while not trusting that person. But that distinction is between the act of entrusting and trusting not between trusting and trustworthiness. So long as I believe you are trustworthy I trust you. The lack of a necessary connection between our actions and our motivations is important in another way though, which is the topic of Chapter Three.

Chapter Three

In Chapter Three I argue that trust is risky. Trusting is founded on uncertainty and makes one vulnerable. Because our actions are separate from our trust there is an inherent uncertainty and vulnerability that comes with trust. The bulk of this chapter is spent clarifying the various ways that risk is understood in order to try to bring consensus to positions on trust. By and large I think the disagreement here lies in a misunderstanding. Some people use risk exclusively in the sense of harm while others use it to mean uncertainty. However, having tried to find a common ground I still argue two points. Following Chapter Two, I maintain that trust is not an action but I contend that actions are not the only way we suffer harm.

The harm that I refer to in Chapter Three is only fully understood in relation to the argument of Chapter Six, which holds that trust is a framework for interpretation. But in principle that harm is obvious at this stage as well. What we believe can hurt us. Attitudes that we have can do us damage. Not only can they limit our behaviors they can destroy our confidence. If I believe that others are plotting against me, that they have an ill will towards me, then I must expend a great deal of effort working to thwart their plots.

The uncertainty inherent in trust and the vulnerability that comes with trusting greatly affects the options that we see as possible and even probable. The decisions that we make about which course to pursue and how to proceed are influenced and affected by what we take our situation to be right now. If I distrust those who could possibly watch over my children then I cannot comfortably leave my children in their care. Trust is risky. And, it is because trust is risky that not just any motivation will do as an explanation for your behavior.

Chapter Four

In Chapter Four I argue that trust is built on good will. Trusting is a belief that the other has good will towards you while trustworthiness is to have good will towards another. This follows from the vulnerability argued for in Chapter Three. Because I am vulnerable I look for those who will not try to take advantage of my situation. If it is in your interests to consider my interests this can only be because your interests give you reason not to take advantage of my situation, which is a condition which leaves me ever more vulnerable; from this perspective not only have I opened myself up to another but I have done so because you have something to gain from not harming my interests, which would be cancelled when your interests conflicted with mine. Instead, I would rather open myself to the care of one who is concerned with my interests because she is concerned for me. This way, when her interests conflict with mine I can have some hope that she will not abandon me in pursuit of her own good; which is to say that she will not take advantage of my vulnerability when it suits her.

This is a point that I think is often overlooked when considering trust. The motivation which marks one as being worthy of trust is of a very particular kind. It cannot be one of self-interest because, almost by definition, self-interest is not interest in another. It might seem from this that if you trust a person that you will believe that they will sacrifice themselves for you but this, also, misunderstands what it is to believe another to have good will. The kind of single-minded devotion that leads one to sacrifice themselves entirely for the needs of another is the sort which does not consider how the relationship can be harmed by any particular action. Sometimes the most trustworthy thing is to do nothing at all. Sometimes the best way to show that you are concerned for another's well-being is to let that person take care of himself. A desire to please another that is not directed by good judgment is not a good will at all. It is this revelation that allows me to argue that trust is a two-part relation.

Chapter Five

In Chapter Five I argue that trust is a two-part relationship in which A trusts B (or B is worthy of A's trust) instead of the more common three-part conception in which A trusts B in regards to some thing C. This is, perhaps, my most controversial claim. It seems obvious, when we consider those instances when trust has been important to us, that it has always involved a three-part relationship: I have been concerned with how B will act in regards to some thing C. But this focus on the moments when trust has been most in our minds misleads us about its nature. Particular moments and actions drive this way of understanding trust but Chapter Two shows those to be a mistake. We overlook how often trust resides in the background of our lives, how often we look right past some

event or action because we trust the person doing it. Trusting another means not having to consider that person in relation to some object or action. Since trusting is a belief in the good will another has towards you that good will should not be dependent on any particular object in question.

When we consider another in relation to particular objects it is because we do not believe that he has a good will towards us and so must engage in a cost-benefit analysis in order to see if the thing in question is worth more than what we can afford to lose in allowing that other to control it. Three-part relationships are the purview of entrusting and expectation not trusting. Being trustworthy is being concerned for another's interests no matter what those interests are, it is recognizing when you should not be in charge of those interests and doing your best to make sure that someone with the skills required is in charge. If a desire to impress or make yourself look good leads you to keep the reins of the wagon when you know that you have never been on this trail before then you are acting not from concern for those in the wagon with you but from a particular kind of self-aggrandizing self-interest; you would not be acting from good will towards them. If your friend puts her life in your hands, it is because she believes that you will do what is in her best interests not what is in yours. The realization that trust is a two-part relation leads to another realization: that we interpret the behaviors and motivations of others.

Chapter Six

In Chapter Six I argue that trust, following the conception that I have established, is an interpretive framework for understanding others. Actions are not knowable independent of an interpretive model, which limits the number of possible ways of

understanding the actions to a manageable set of positive, neutral, or negative interpretations. Trusting, on this conception is, therefore, to interpret another's actions favorably while distrusting is to interpret another's actions unfavorably.

Some of the most obvious examples of this kind of interpretation involve relationships. Friends who trust one another overlook the evidence, obvious to others, that they are each deviously plotting behind each other's backs. Or a couple, trusting fully each other, slides easily past obstacles that wreck relationships on less stable grounds. Simple oversights, like a forgotten dinner date, become evidence of deep rifts in a relationship when trust is absent but are completely overlooked when trust is robust. Innocuous misunderstandings become mountains of insurmountable rage when distrust looms large while they become sources of strength for distrustless relationships.

Although my interest has in some sense waned in the years that I have worked on this project I am excited about the position and the effect that it could have on discussions of trust. Treating trust as an action sets us up to view our relationships like contracts, seeking to find some balance of desire and punishment, some way to provide ourselves with incentives to not break the trusts we have. But if we can re-conceive or our interactions as part of an ongoing process, as part of our way of being, and, at the same time, realize that our trust is never settled, that it is also open to revision and interpretation, then we can begin to see ourselves as participating with others instead of acting in opposition to them. We can move away from treating each other as self-interested machines that are constantly looking for advantage. We can start to see institutions as collections of individuals who each need our encouragement to become trusting members of a relationship instead of cogs in an impersonal organism that seeks

to generate trust through equally impersonal mechanisms. This project has value, so we might as well get to it.

Chapter Two: Trusting and Trustworthiness

On its face the connection between trusting and trustworthiness might seem obvious. If the two notions seem inextricably linked together, it is for no other reason then they both reference "trust." We should not, however, so easily accept the linkage; the way in which we understand the connection between these two will dominate our conception of trust, which can lead to misunderstandings about the nature of trust. Is the connection necessary or conditional? Are they causally or conceptually related? In what ways are they *not* connected? How we respond to questions like these will go a long way towards clarifying our understanding of trust. It is therefore necessary to consider the ways in which trusting and trustworthiness are related as a prelude to an investigation of trust. I argue that trusting and trustworthiness do not entail each other but that trusting another is to believe that other to be worthy of trust.

Authors that tightly connect trusting and trustworthiness tend to overemphasize the link such that they begin to fear the rise of a general skepticism about the behavior of others; for them, any untrustworthy behavior threatens the possibility of all trusting behavior. Bentley Glass is a good example of an author who stresses the connection between trusting and trustworthiness to the detriment of a conception of trust. Glass is worried that wide-spread distrust will lead to the destruction of science.[1] He believes that untrustworthy behavior by some scientists will lead to all scientists not being trusted but

[1] Bentley Glass, "The Ethical Basis of Science," *Science* 150, no. 3701 (1965).

all scientists need to be trusted and so every scientist *must* be trustworthy. Thus, on Glass' account trustworthiness, as a moral maxim, finds its foundation in the need for trusting. I argue later in the chapter that this justification is problematic because placing so much emphasis on the need to be trusted can actually drive scientists to act in ways that are not trustworthy. The need to get a particular grade in a course, in my experience, is the most common motivation for students who have been caught plagiarizing. Their overwhelming focus on what they need from the course, like the singular focus of Glass' account of what scientists need, actually sets up the conditions for the worst kind of violation. When a person absolutely must pass a course the worst thing he can do is to act in such a way that he ought not pass the course. When a person absolutely must be trusting the worst thing she can do is to act in such a way that she ought not be trusted. Glass is not the only author who is concerned with the effect of individual untrustworthy actions, however; Marcel LaFollette, in considering instances in which data and evidence in scientific experimentation were revealed to be fabricated, writes that "[t]he discovery of fraud and deception threatened to tarnish science's pristine image."[2] LaFollette's concern is that the revelation of a few "bad apples" will ruin the ability of science to do the job that it is meant to do. Although LaFollette suggests a highly idealized notion of science his concern is not without merit. Like Glass, he recognizes that science is balanced precipitously on the edge of a cliff; its practice requires interaction between people who must depend on each other's honesty. If scientists can no longer depend on each other than the practice will tumble into the abyss. Recognition of this dependence is *not*, however, the issue that directly concerns me; whether or not science really is

[2] Marcel C. LaFollette, "The Evolution of The "Scientific Misconduct" Issue: An Historical Overview," *Experimental Biology and Medicine* 224, no. 4 (2000): 211.

dependent in the way suggested by Glass and LaFollette is irrelevant to my immediate position. Instead, I am concerned with what this perceived dependence reveals about our conception of trust.

We might consider the concepts associated with trustworthiness in order to develop an account of trust but it is not clear that considering these concepts will necessarily lead to a greater understanding of the connection between trusting and trustworthiness. For example, many authors equate trustworthiness with integrity. LaFollette claims that "the painful reality of deception by a colleague has usually been followed by conscientious investigation of the situation and a renewed sensitivity to the importance of integrity in science."[3] Simply put, LaFollette sees integrity as an unwillingness to deceive, which, on his account, is central to trustworthiness. Similarly, Glass writes that science "rests upon the scientist's integrity," which he sees as a willingness to be truthful, the bedrock of trustworthiness.[4] And Morton Deutsch claims that integrity is the personality characteristic of being responsible which, for him, means "that the trustworthy person is aware of being trusted and that he is somehow bound by the trust which is invested in him."[5] For Deutsch then, integrity means to have the personality characteristic of trustworthiness. But conceiving of integrity as trustworthiness leaves open the meaning of trustworthiness and its connection to trusting; it fails to provide sufficient explanation of the relationship between trusting and trustworthiness because it simply changes the question to one of the relationship between trusting and integrity. Glass' claim that science rests upon integrity is the same as saying

[3] Ibid.: 212.
[4] Glass, "The Ethical Basis of Science," 1257.
[5] Morton Deutsch, "Trust and Suspicion," *The Journal of Conflict Resolution* 2, no. 4 (1958): 268.

that science rests upon trustworthiness, which is simply saying that science needs scientists to be trustworthy, which is what we already understood his position to be saying. We must still consider the connection between trusting and trustworthiness if this investigation is to be fruitful.

I begin this chapter by examining Glass' account of value in science. His position provides a framework for understanding the ways in which our conception of the connection between trusting and trustworthiness is central to our understanding of trust. His account of the importance of trust exhibits a subtle reliance on a necessary causal connection between trusting and trustworthiness. Additionally, his position provides an early indication of the importance of trust to ethics, specifically regarding our judgments of right and wrong. After explicating Glass' position, I move to a discussion of the various ways in which trusting and trustworthiness are not causally connected. This important step narrows the field of possible contenders for an account of the connection between trusting and trustworthiness and allows me to present a positive account of the connection. It also lays the groundwork for several criticisms that will appear both in this and later chapters. Finally, I move to a discussion of the Prisoner's Dilemma. This classic paradox of decision theory presents several problems to any account of trust, problems that a clarification of the connection between trusting and trustworthiness can help to resolve. I argue, contrary to prevailing opinion, that the Prisoner's Dilemma is not particularly useful in investigating the phenomenon of trust because it fails to be able to distinguish between trusting and trustworthiness.

Bentley Glass

The account of science put forth by Bentley Glass is not, strictly speaking, a discussion of trust but it can still be informative as an example of a practical application of a conception of trust; by considering his position we can see how important our understanding of the link between trusting and trustworthiness is to our conception of trust. His concern for trusting and trustworthiness is intimately related to his conception and understanding of science. Science is not merely some useful way of gaining knowledge about the world, for Glass it is *the* capacity of humanity which will allow us to control our destiny. Rather than being subject to the whims of chance and mutation, humanity, through science, will be able to select those traits which are most beneficial. He does not mean simply that science will allow us to manipulate the genetic code, though he does not seem to be averse to that prospect, rather he means that science will allow us to uncover the truths of right behavior.[6] Further, by uncovering those truths, science will provide us with the manner of selecting for them. Science is, on his account, the ultimate panacea for human frailty because it allows us to eliminate our faults.

Glass claims that "by examining critically the nature, origins, and method of science we may logically arrive at a conclusion that science is ineluctably involved in questions of value, is inescapably committed to standards of right and wrong, and unavoidably moves in the large toward social aims."[7] He is responding to the criticism that science, as a field, is neutral with regards to value, that it neither creates nor reveals

[6] In what is perhaps more than mere coincidence, Robert Heinlein's *Starship Troopers*, which was published in 1959 (6 years before Glass' piece), evidenced a similar conception of the value of science to our understanding of ethical reasoning; science will reveal those actions and attitudes that were objectively right and those that were wrong.
[7] Glass, "The Ethical Basis of Science," 1254.

any particular ethical rules but instead can be used to fulfill any end, good or bad. While it may be true that science can be used for bad ends Glass denies that science is neutral with regards to value and in doing so his goal is ambitious: to show that science does not simply uncover our values but is itself a source of them. He argues that one of the most important social values that can be recognized through science is trust; the moral imperative to be trustworthy can be revealed, on Glass' account, by considering science itself. It is important to remember that even though he argues for the importance of trust to science Glass' ultimate aim is to show how human values can be revealed through science; we can see the importance of trust through science but its importance is universal to human existence. He reaches this conclusion by following two parallel lines of reasoning. In one line of argument he first shows that human values have evolved over time to fit with our environment, and then he shows that it is right to hold those values. In the other line of argument Glass defends trustworthiness as imperative to the practice of science by first showing that science depends on verification and then by showing that verification demands trusting, which, for him, means that science demands trustworthiness. Both lines of reasoning combine in Glass' account to show that we have a moral imperative to be trustworthy, specifically in the practice of science but more generally in the practice of living as human beings; for Glass what is right for science is "what is right for the entire community of life on earth."[8] The imperative to be trustworthy then is not simply an imperative for science but for all human interaction and all life on Earth.

[8] Ibid.: 1255.

The Importance of Trust to Science

Glass reveals his intention to present an evolutionary account of the imperative to be trustworthy when he claims that "[h]uman values have themselves evolved."[9] He does not mean simply that human values have changed like one's taste in wine might change over time but instead means that those things that are important to humans are important to us because concern for those things has helped us to survive. His notion of evolution is the biologist's notion of evolution. Values that contribute to a person's success are more likely to be passed on while values that are detrimental to the success of the individual are less likely to be passed on because that person is less likely to survive and those values will therefore also not survive. Tempered in the fiery crucible of an unforgiving world, our values have developed over time in response to the demands of our environment. These values need not be inherited genetically; it is possible for our values to be passed down as adopted practices so that, at this stage in the discussion, we should not assume that our moral beliefs are inherent, genetic traits.

Glass begins by claiming that "[m]an's own values grew out of his own evolutionary origins and his struggle against a hostile environment for survival."[10] His position is that our ethical values have developed by the process of natural selection, that they are a result of environmental pressures. In order to survive we must be successful in our actions; beliefs and values that led to actions that were unsuccessful were not passed on. Although unavailable to Glass at the time he wrote, this position is developed more fully by Robert Axelrod's studies about the possibility of altruistic behavior occurring in

[9] Ibid.: 1254.
[10] Ibid.

a community of self-interested individuals.[11] Axelrod uses an iterated Prisoner's Dilemma in order to model the process of natural selection. I will return to a discussion of the Prisoner's Dilemma later in this chapter but for now it is enough to recognize what Axelrod and Glass have in common: a belief that behavioral traits are heritable characteristics whose fitness is subject to the process of natural selection. Indeed, we find even contemporary authors working on this same problem. For example, Michael Shermer presents a view of ethics that is based on evolutionary psychology whose "fundamental premise…is that human behavior evolved over the course of hundreds of thousands of years."[12] Shermer's goal is the same as that of Glass and Axelrod, to use science to show not only that human moral behavior has evolved by to reveal what our moral rules should be. I focus on Glass here because Shermer's account does not directly focus on the importance of trust; in fact, his account of trust is secondary to the account that he is presenting.

Our values are, for Glass, like our biological traits and faculties; they have evolved over time through the process of natural selection. He gives us examples of the kinds of values that he is talking about; "the love of a mother for her child and of the man for his mate, the willingness to sacrifice one's own life for the safety of the family or tribe, and the impulse to care for the weak, the suffering, the helpless," which he claims all have "the same primitive beginnings."[13] These values all trace their origins back to being a type of behavior that was slightly more successful than its alternatives. It is easy to see how one could hold the position that he does. For instance, mothers who did not

[11] Robert M. Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).
[12] Michael Shermer, *The Science of Good and Evil : Why People Cheat, Gossip, Care, Share, and Follow the Golden Rule*, 1st ed. (New York: Times Books, 2004), 9.
[13] Glass, "The Ethical Basis of Science," 1255.

love their young would be less likely to expend extra effort in protecting their children and so those children would be less likely to survive. The trait of 'not loving' is, therefore, less likely to be passed on to future generations. At the same time, mothers that loved their children were more likely to protect those children even to the point of sacrificing their own lives. On the aggregate then mother-love, so the argument goes, actually results in a higher survivability rate for off-spring and so the trait of 'loving' is more likely to be passed on to future generations. Loving is a successful strategy for dealing with the uncertainties of the world.

We might worry that seemingly altruistic actions like mother-love actually have alternate, non-moral motivations; human moral activity, if it is a result of the process of natural selection, might really be aimed at furthering our genetic code rather than the right or the good. Simon Blackburn recognizes this as a possible threat to ethics, what he calls a Grand Unifying Pessimism, but he argues that such an origin, even if true, fails to account for what is happening in the instances in question.[14] The threat to ethics is that evolutionary theory unmasks ethics as a cunning ploy or merely an adaptive trait, a kind of trick; that there is no such thing as mother-love, there are only genetic predispositions that work to further our strain of DNA. If trust is part of a successful strategy for the propagation of genetic material, then we might worry that trust is merely a mask for further one's genetic code and not an independent human value. Blackburn claims that the fear of evolutionary theory contains three confusions: 1) that our apparent concerns are not our real concerns simply because there is an evolutionary explanation for them; 2) that a perceived trait does not exist because there is no evolutionary explanation for it;

[14] Simon Blackburn, *Being Good : An Introduction to Ethics* (Oxford: Oxford University Press, 2001), 37.

and 3) that it reads psychology into nature and then back into our selves. The first confusion suggests that because there is an alternative, evolutionary explanation for some particular value that our experience of that value must not be meaningful. The second confusion suggests that if no evolutionary explanation can be found than our value does not actually exist. And the third confusion suggests that we take our current understanding of human behavior as a way of understanding the world in order to return that conception back upon ourselves; in effect, this is to use our understanding of human behavior to justify itself, which is hopelessly circular.

Blackburn's response to the first confusion found in the evolutionary criticism of ethical value is that a trait's status as adaptive has no impact on the fact that we are feeling it. Mother-love may be an adaptive trick that helps those with it be more successful at passing on their genetic code but that does not deny that the mother loves her offspring. Her "real concern" is not to promote the species (a common misunderstanding of evolutionary theory) it is her love of her child. Successful traits are more successful than the alternatives; they are not masks for some other successful behavior. Being hungry is an adaptive trait that lets you know that your body needs nourishment; being hungry allows animals with that trait to be more successful than animals that cannot recognize the need for nourishment. But, it is a mistake to think that being hungry is anything other than how we describe it; it is not a desire to further our genetic code, it is a desire to eat. Hunger does not cease to affect us, to twist our bellies in pain, simply because there is an evolutionary explanation for why we feel it. A mother who loves her young happens to be more successful at helping her offspring to survive than a mother who fails to love her young, but the mother still feels love. The criticism –

that mother-love is just an evolutionary trick – actually denies the initial condition; it has to deny that the mother was feeling love in order to explain why a mother feels love.[15] We see, later in this chapter, that trust is especially vulnerable to this kind of confusion because trust is often seen as being within our cognitive control; a person's apparent trustworthiness is often for purely strategic reasons.

The second confusion, that a perceived trait cannot exist if there is no evolutionary explanation, also misunderstands the process of natural selection. Many traits have obvious advantages over competing alternatives: for example, binocular vision, especially for carnivorous hunters, has decided advantages when compared to non-binocular vision (depth perception being perhaps the most prominent). But not all traits have obvious or even obscure advantages; the pattern in which hair grows on the head, either a clockwise or counter-clockwise cowlick, has no apparent advantage and yet it is a genetically heritable trait. The mistake here is assuming that the only avenue for the continued inclusion of a trait in the genetic code is its participation in the overall success of the individual and the species. It should be clear however from the example provided that traits need not participate in the success of either the individual's ability to procreate or the success of the species; one trait can piggyback with another. A trait that is not detrimental may well continue in a population even when it provides no particular advantage; that there is no evolutionary explanation for a trait cannot therefore function to deny the existence of that trait. It may well turn out that there is no evolutionary explanation for trust and yet trust could still be a value in society and, even, a heritable characteristic.

[15] Ibid., 39.

The third confusion, that we read psychology into nature in order to return it to ourselves, takes a metaphor that was meant to help us understand and explain certain biological processes and uses it as a description of the nature of our being. In order to make this point clear Blackburn presents the example of Richard Dawkins' *The Selfish Gene*. Dawkins uses psychological metaphors throughout his discussion; "A body doesn't *look* like the product of a loose and temporary federation of warring genetic agents who hardly have time to get acquainted before embarking in sperm or egg for the next leg of the great genetic diaspora."[16] In this sentence Dawkins attributes the ability to make war and the ability to become acquainted with others as descriptors for the behaviors of genes but these abilities are really only applicable to individual people who can hold opinions, have intentions, or have personalities. By giving psychological characteristics to genes it is easy to then transfer those characteristics back onto the creatures that carry those genes. So, when Dawkins uses 'selfish' to describe a gene that works to protect and to promote itself, and when that gene is carried by all people, rather than being a trait that only some people carry (like selfishness), it becomes easy to interpret the behaviors of those people as being motivated in the same way; if a gene selfishly promotes itself, then the person with that gene must be selfish as well. But, as Blackburn points out, this is a serious mistake. Genes do not have attitudes or even motivations; we anthropomorphize them in order to more easily explain their behavior. We use the familiar traits and characteristics of people to describe the actions of genes. To then use those descriptions of genes in order to explain our behavior confuses a metaphor for an explanation.

[16] Richard Dawkins, *The Selfish Gene*, 30th anniversary ed. (Oxford ; New York: Oxford University Press, 2006), 234.

Given Blackburn's responses, the fears raised by evolutionary theory are not substantial enough to cause real trouble to ethics. It would be a mistake, therefore, to dismiss Glass' account of trust simply because he presented it as having its basis in evolutionary theory; he might very well be able to provide significant insight into the origin of our values without, at the same time, diminishing the importance of those values to us. Glass' position, however, is not out of the woods yet.

Accepting the possibility that our values have evolved has not yet shown that those values are anything more than subjectively determined by the environmental conditions of our existence; that a value has evolved does not obviously make it good or right. Since Glass is concerned with showing that science is "inescapably committed to standards of right and wrong" and not that it is committed to subjective interpretations of right and wrong he must show that our evolved values are in some sense objectively good; valuing that is purely subjective, that is subject to whim, fails to achieve the status of a standard that Glass is so keen on pursuing. His position might simply show that in the environment of the United States, or the larger environment of the Anglo-Saxon scientific community, a certain set of values has come to be successful. These values might, in another community, be detrimental in much the same way that a Broad Beech Fern is well suited for the medium-light, high moisture environment of the arboreal forest floor but is poorly suited for the intense light and low moisture of an equatorial desert. Glass must therefore still defend our values as being good if he is to achieve his goal of showing not only that trustworthiness is a value that we should hold but also that science can reveal this to us. He approaches this, as we might expect of him given his background in biology, from a teleological perspective. He defends our values as good insofar as

those values are targeted at some end, an end that those values evolved to fulfill, which ultimately enables us to be more successful than those who do not share that value.

Glass uses an analogy about the human eye to make his point about the rightness of our values; "[t]he eye," he claims, "is for seeing."[17] While this might seem like a simplistic point, for Glass it carries substantial meaning; the purpose of the eye will ultimately defend its use as good. What counts as good depends on the function of the thing in question. This point should not be too revolutionary for those familiar with Aristotle: "just as the good…for a flautist, a sculptor, and every craftsman…seems to depend on its function the same seems to be true for a human being."[18] Glass' argument follows this same line of reasoning. The eye has a specific function, one which it has evolved to fulfill; "[s]ight conveys information about…vitally important matters."[19] The eye's evolved function helps us to be successful in our environment. Avoiding danger or finding water, examples of the kinds of vitally important matters suggested by Glass, are necessary to our continued existence and the ability to see helps us to do these types of things. Because the eye has evolved to fill this role, which gives it a purpose, Glass argues that it is right for the eye to do this; "[t]o see is right; not to see is wrong."[20] 'Right' is an organ fulfilling the end which it evolved to have as its purpose; this evolved end, by virtue of its successful continuance and in some sense dominance, is one that we recognize is of worth. Sight allows us to avoid danger and to find food which helps the survivability of the species and is therefore valuable to us. The eye is designed by the process of natural selection to see and so Glass says that sight and the organ which gives

[17] Glass, "The Ethical Basis of Science," 1255.
[18] Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis, Ind.: Hackett Pub. Co., 1985), 15-16.
[19] Glass, "The Ethical Basis of Science," 1255.
[20] Ibid.

it to us are right, that seeing is good. If trust can be shown to be a successful strategy for continuing the species than trust will have been shown to be right.

Fitness, which describes successful traits, does not however, by itself, clearly and easily get us to traits that are 'right.' Traits that have been selected for fitness are not necessarily 'good' traits; Glass is aware of this and he points out that "[f]itness, like it or not, in the long run meant simply the contribution of each trait and its underlying genes to survival."[21] Fitness, instead of being some sort of measure of the objective value of the trait in question is really just a description of the trait's value in promoting survival in relation to its alternatives. No trait is ever universally fit because no trait is ever fit indefinitely or independent of situation. The trait of being flightless was once the fittest trait for the Dodo bird; in an environment with no predators, being able to fly, with all of its high energy demands, is not terribly efficient. However, being flightless was the trait that sealed the Dodo's fate; what had been a very fit trait became a very unfit trait. The introduction of new species, new traits, or new conditions can radically affect the fitness of any given trait. If our values, specifically our moral values, are due to our evolutionary origins then our values are only fit within or relative to some specific set of circumstances; they could, and mostly likely would, become a danger to us. Glass recognizes this point though; "these ethical values are always, in the evolutionary scheme of things, relative, and never absolute."[22] He is not claiming that the current success of a trait is an indication of its overall or universal fitness, because there is no universally fit trait.

[21] Ibid.: 1254.
[22] Ibid.: 1255.

Glass must then account for how the relativism of fit can lead to an objective statement of value. And, because "[o]ur highest ethical values – the love of a mother for her child and of the man for his mate, the willingness to sacrifice one's own life for the safety of the family or tribe, and the impulse to care for the weak, the suffering, the helpless – all those things had the same primitive beginnings" Glass must account for them as well.[23] He is attempting to show that human values are traits have evolved to be the values that they are because they have helped us to survive in an unfriendly world against competing species, traits, and values. Trust is, for Glass, one of our highest ethical values.

Glass measures traits in terms of the successful continued existence of the humans who carry that trait. This position might seem to fly in the face of Hume's law, that one cannot derive an ought from an is, because it attempts to identify what we ought to do by reference to what it is that we do. But, Glass has no problem proceeding in this manner; "[t]he evolutionist is quite prepared to admit the existence of right and wrong in terms of the simple functions of biological structures and processes"[24] He is perfectly willing to admit that right and wrong are determined by fulfillment of purpose, and purpose can only be determined by looking to what exists in the world. His discussion of the rightness of sight and the human eye was an attempt to make this point clear. Again, Glass is not alone in holding this position; Richard Dawkins' *The Selfish Gene* and E.O. Wilson's *Sociobiology* both purport positions similar to the one suggested by Glass.[25] So, on Glass' account, our values have evolved and those values are, in some sense, good; loving is

[23] Ibid.
[24] Ibid.
[25] Edward O. Wilson, *Sociobiology : The New Synthesis* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1975).

right and not loving is wrong because loving is a strategy that has been shown to be successful in promoting human life. The goodness of those values, however, depends entirely on the specific function for which they evolved; the conditions that led to the selection of these values over competing values, the conditions for which these values were more successful, determine the fitness of these values. Being dependent in this way makes the values subjectively good and not objectively good. But, Glass is not claiming that the current success of a trait or these values is an indication of its overall or universal fitness; there is no universally fit trait. Instead, he must then account for how the relativism of fit can lead to an objective statement of value and the objective importance of trust. This leads Glass to turn his attention to the power of science.

Trustworthiness Is Imperative

Glass' intent in writing the article was to focus on issues within science; he wanted to address how scientists should behave. But, he wanted to give a justification that relied on the practice itself rather than appeal to anything external to science. That is, he was trying to show that science can reveal to us what values we should hold. In order to do this his account would, eventually, have to turn its attention to science in order to explain how people should behave. So, he begins to pull his account together by focusing on the objectivity of science, which, he claims, "depends wholly upon the ability of different observers to agree about their data and their processes of thought."[26] This dependence has two faces; (1) it needs external observers in order to recreate and verify the results and (2) because of its increasing complexity it needs separate individuals involved in different aspects of research in order to proceed. Research cannot be

[26] Glass, "The Ethical Basis of Science," 1256.

considered objective unless the results and the process can be verified by someone else but research also depends on previous work, work that individual scientists have neither the time nor the resources to complete on their own. Science, by being dependent on the reports of others, is especially susceptible to fabrications and lies. Glass recognizes this when he claims that "we must be able to trust the word of others."[27] If we cannot trust others then the process of science will fail and so Glass is arguing that being trustworthy is a moral imperative. Glass' assumption of a necessary connection between trusting and trustworthiness allows him to turn a demand for trusting into a demand for being trustworthy. I will challenge this position later in this chapter but for now it is enough to see how it functions within his overall position.

Trust s an Inherited Trait

Glass then turns to the development of science; "the first simple steps in the development of science: observation, reporting, written records, communication."[28] Science, which developed over time through trial and error, success and failure, transitioned from a singular action to a community event. He sees this as parallel to the development of human reason. Glass argues that, initially, observations were done by particular individuals. One could make a number of observations and could be very successful because she was able to keep track of a number of relevant facts during her lifetime. But by incorporating the reports of others, presumably by exchanging observations, she could improve her success rate. A group, therefore, by working together can vastly improve the success of their endeavors. Reporting, as part of an oral tradition,

[27] Ibid.
[28] Ibid.: 1255.

is subject to faulty memories but when reports begin to be written down, they become much more stable and lasting; we can refer back to generations long since past in order to compare our observations, which creates within each of us the possibility of a substantial body of evidence for making inductive judgments about the world. Finally, with the advent of stable and ever-increasing communications, our observations can be transmitted and compared untold times. Already we can see that Glass' position, which began with knowledge that was maximally subjective, has taken strides towards objectivity. Science had its beginnings in the individual process of observation in a substantially subjective experience but it has become a collective process that is capable of revision when confronted with the incompatible observations of several people. It is important to remember however that Glass believes that "we cannot break the bonds of our subjective interpretations of the physical events of nature."[29] For him we are always embodied creatures who respond to the world based on our experiences. We can struggle to achieve some kind of objectivity but understanding of the world will always be interpreted through out subjective sense of being.

Science can, so Glass contends, be used to make humans conform to their environment and to make the environment conform to humanity. One might suggest that science is the trait of adaptability, one that humans have and one that should be considered fit. But, adaptability can be understood in two ways. One way of understanding adaptability is to consider it as a kind of malleability; that is, a species that can mold itself to its environment would be adaptable. Humans do not have this kind of adaptability; we are not able to develop gills when exposed to water or a thick hide when

[29] Ibid.: 1256.

exposed to heat. Humans simply cannot make their bodies match their environment. The other kind of adaptability would be the ability to function within multiple environments; humans do have this kind adaptability. Our skin, when exposed to sunlight, does darken in order to protect us. We do produce more hemoglobin when we move to high altitudes. And, our capillaries do restrict blood flow to our extremities when we are in cold weather. That is, we are able to function in multiple environments. But, even more important that our ability to remain homeostatic is our ability to make our environment conform to us and this kind of adaptability is only possible because of our intelligence. We adapt to the cold by wearing clothes, to the water by using a snorkel, to high altitudes by carrying oxygen; our intelligence allows us to adapt to various environments. Adaptability in either of its forms, however, does not come without a cost; it is an energy intensive trait. Humans give up a lot in order to be adaptable in the way that we are. We are, in one sense, the most fragile animals on the planet; Immanuel Kant says of humans that were "given neither the bull's horns, the lion's claws, nor the dog's teeth, but only hands."[30] We can survive in any environment but only when we force that environment to conform to our needs.

In addition to the role that heritable characteristics play in our survival Glass also believes that our survivability is culturally transmitted. Removing our access to electricity would throw contemporary human culture into chaos. Eliminate our ability to transport goods over vast distances and many cities would fall in a matter of weeks. Humans are like tame housecats, throw us into the wild and we are lost. So, the cost of being fit can be seen in two ways: 1) the amount of energy it takes to be intelligent and 2) the role that

[30] Immanuel Kant, *Perpetual Peace, and Other Essays on Politics, History, and Morals*, trans. Ted Humphrey (Indianapolis: Hackett Pub. Co., 1983), 31.

living in a society plays in our culturally transmitted knowledge. There is a necessary consequence to the manner in which we are fit: another trait could be more fit. For example, ants lack what we recognize as individual intelligence and so ants lack the success trait that we take to be so valuable. Ants are however, looking strictly at population size, far more successful than humans. A species that had an ant's group, rather than individual, intelligence but lacked the restrictions of an ant's circulatory system could be far more successful than humanity because humans pay a high price in energy expenditure in order to support our individuality. What we imagine to be an incredibly valuable trait (individuality) – a trait that we think makes us the fittest of all species – could turn out to be the very trait that makes us a dead-end of evolutionary history. If our morality functions on these evolutionary rules, then we are not only tied to our morals in a way that removes them from all cognitive control but we are also doomed to be usurped by some other group with a more successful moral code. That is, if trust is merely a kind of inherited trait then it is susceptible to this same kind of supplanting but if our naturally evolved values could be augmented by knowledge gained through science and transmitted through culture then we would have a way of maintaining our status at the top of the fitness pyramid.

The Need Does Not Obligate

We find then, in Glass' account a conception of trust that takes trusting and trustworthiness to be linked through a necessary, causal connection which is justified by appeal to a human need, which Glass takes to be primordial in origin and unavoidable in human experience. But when trustworthiness ends up serving the good that trusting gets

us, when trustworthiness is needed only in order to make trusting possible we confuse a practical necessity for an ethical imperative. Being trustworthy may be good because, as a practical matter, it fulfills our need for trust but it is not clear from the fulfillment of a practical need that trustworthiness is itself morally good. It leaves open the possibility that we may, in at least some instances, not trust others. One might, as a scientist, need to be able to trust other scientists but that does not mean that one must trust all scientists at all times or even any particular scientist at any particular time. This means that for any given interaction the only reason that I ought to trust is that the interaction is worthy of trust and not that there is some necessity to be trusting. My need to trust others is a general claim about existing within society but no particular interaction must instantiate the needs of society; we need the capacity to trust others but when we should exercise that capacity seems to be particular to the conditions of specific interactions. However much I may value meaningful social interaction or even the possibility of scientific advancement in a complex and diverse society those values can be obscured, diverted, or subverted by my interest in other things. The capacity to trust others cannot demand that I trust others in all instances. The capacity to be trustworthy cannot mandate that I be trustworthy in all instances.

It might seem that our need to trust can justify the obligation to be trustworthy, certainly Glass argues this point but privileging the neediness of trust does not clearly lead to the conclusion that Glass is advocating. We should consider the claim under different circumstances. We need air to breathe and so we must have a right to that air; it is necessary to our continued existence. This explanation can justify various protections for the environment like the Clean Air Act. It is useful here because the atmosphere and

air are common metaphors in trust discussions. Sissela Bok suggests that "*whatever* matters to human beings, trust is the atmosphere in which it thrives."[31] Annette Bair, following Bok, says that "[w]e inhabit a climate of trust as we inhabit the atmosphere and notice as we notice air."[32] Glass' account of the need to be trusting is very much in line with the human need for air. It is not simply that scientists need trust, although they do, rather it is that all humans need trust and science can make that point clear to us. The position itself is problematic, however. I need air to breathe, so not only to I have a right to that air but, on Glass' read, I have an obligation not to do anything to harm the ability of that air to provide me with breath. Smoking would be morally prohibited on the grounds that it damages the atmosphere's ability to be inhaled. Many non-smokers would agree but notice that no single act of exhaling smoke will sufficiently hamper the usability of air. So, although I need air, no particular moment of misuse will justify no longer being able to breathe. Further, the very act of exhaling, which is a necessary counterpart to breathing, damages the air that I breathe in exactly the same way. Trust does not function in this way, trusting does not harm trustworthiness, but the air analogy dramatically makes the earlier point. The need to breathe does not justify an absolute prohibition in regards to particular acts involving air. Neither can the need to be trusting justify an absolute prohibition in regards to particular acts of violating trust. Glass' justification of trustworthiness in untenable.

Further, putting the good of trustworthiness in the service of a necessary trust leaves open the possibility that the particular good of being trustworthy can be replaced

---

[31] Sissela Bok, *Lying : Moral Choice in Public and Private Life*, 2nd Vintage Books ed. (New York: Vintage Books, 1999), 31.
[32] Annette Baier, *Moral Prejudices : Essays on Ethics* (Cambridge, Mass.: Harvard University Press, 1994), 98.

by another act which serves the same end. On Glass' account there is nothing specific about trustworthiness that is to be commended except that it gets us to a thing that we admit that we need: the trust of others. Trustworthiness is merely a means to generate trusting. Suppose that I am not very good at handling money and I am honest about it, you will not trust me to manage your finances. In order to get you to trust me, which Glass claims that we need, I am presented with a strong incentive to lie to you about how well I have handled other people's money. In this case a lie, which is not generally the act of a trustworthy person, is helping you to trust me, which fulfills Glass' imperative. The appearance of trustworthiness, not actually trustworthiness, would suffice to generate trust given Glass' account.

The need to be trusting cannot make trustworthiness an obligation in the sense of being something that we cannot not do; being obligated in that manner would demand trustworthiness in all situations. This would include those situations that Annette Baier recognizes as most problematic for an account of trust. Baier points out that trusting is not always a good thing; she reminds us that even thieves can trust each other and that trust helps to further practices that the rest of society finds harmful.[33] For example, compatriots in a shady investment deal depend on each other for the success of their enterprise; they each have a role to play in the scam that any particular member of the group could violate for her own gain. Trust seems to be necessary; trustworthiness to the goals of the endeavor and trustingness to the other participants. In order to get the end that they all admit that they need – money – they must trust the others. The criminals *need* to trust each other just like every other human being but their need to trust each

[33] Ibid.

other will generate a set of behaviors that will be harmful to society so trust does not seem to always be a good thing. Further, being obligated to trust would require us to be trusting even in those situations where we should not be trusting, i.e. when the other was not worthy of trust. We must, therefore, consider the relationship between trusting and trustworthiness. By doing so we can highlight areas where Glass, and others, have gone astray.

Connection

As the Glass case demonstrates, misunderstandings about the connection between trusting and trustworthiness create problems. In order to make the particular connection explicit it will be helpful to review the ways in which trusting and trustworthiness fail to be connected both between and within persons. There are at least three ways to show that trusting and trustworthiness lack a necessary connection in interpersonal relationships: betrayal, failure, and misplacement. In addition to the separations visible in interpersonal relations we can also see a separation within particular people; I can trust without being trustworthy and I can be trustworthy without being trusting. In this section I investigate these divisions in order to lay the foundation for an account of trust.

Against a Necessary Causal Connection

The most dramatic evidence against a necessary causal connection between trusting and trustworthiness is betrayal. I take "betrayal" to mean willfully behaving in a way that you know is going to harm another or her interests when you had previously encouraged her to believe that you would not do such a thing. Betrayals are cases in

which we have been deceived but more than that, a betrayal is a case where our trust has been intentionally broken by one who has benefited from that trust. Con-men betray others on a regular basis; the dishonest and insincere need to appear honest and sincere in order for their deception to be complete. It may seem overly dramatic to say, as Sissela Bok does, that "trust is a social good to be protected just as much as the air we breathe."[34] Or it might seem hyperbolic to say that society could not function without our trusting some others at some times but it is clear that our involvement with each other would be vastly different were we not able to ever trust anyone. In a distrusting society, interactions with others would always be complicated exchanges meant to guarantee against the possibility of our interests being defeated; a great deal of energy would be wasted. In terms of Glass' evolutionary account, this would be extremely inefficient; a distrusting society simply requires more energy and will, as a result, not be as successful as a trusting one. But, and it is important to recognize this, betrayals cannot occur in *dis*trusting societies, they are an outgrowth of a society that trusts. Betrayal of trust requires there to be at least some trust between individual members of a society. Far fetched though it might be, we can imagine a collection of individuals living in constant fear of the ways in which others might stymie their interests. Indeed, Thomas Hobbes' state of nature describes an existence in which all are distrusted.[35] It takes, on his account, a sovereign to provide the kind of assurances necessary for one to take the promises of others seriously. There can be no betrayal in the state of nature because in a state of nature there is no trust. The separation evident in betrayal reveals that our trust of another

[34] Bok, *Lying : Moral Choice in Public and Private Life*, 26.
[35] Thomas Hobbes and J. C. A. Gaskin, *Leviathan*, *The World's Classics* (Oxford ; New York: Oxford University Press, 1996).

is not an absolute indicator of his trustworthiness; our trusting can be taken advantage of. Betrayals can be painful and often terribly harmful but a betrayal involves an intentional attempt to deceive and, so, is not truly a failure of our assessment, though what this assessment might consist in we do not yet have a clear enough conception of trust in order to say.[36] That we can betray reveals a separation between our self and others, between our assessment and their freedom.

A Breach Does Not Entail a Betrayal

However, not all breaches of trust are betrayals; some breaches are inadvertent or unintentional. Our trust can be let down in many ways and not all of them will be in the form of betrayal. Often, what appears as a betrayal in one instance may be later revealed to not have been a betrayal at all. For example, Aaron might have made a good faith promise to do something on Bob's behalf only to later be presented with an obligation to Charles that conflicts with what he owes Bob. If Aaron fulfills the obligation to Charles, Bob could, not knowing the conditions of Aaron's choice to ignore his obligation, see Aaron's actions as evidence of Aaron's betrayal. The conditions under which Aaron makes the determination to ignore his obligation to Bob, and what his reasons are for doing so, will effect the decision as to whether Aaron has betrayed Bob. Aaron might have promised Charles before he promised Bob, or the obligation to Charles might have been far more substantial (Charles might have needed a Kidney while Bob needed a ride to the store), or, even more obscurely, Charles might have threatened Bob's life if Aaron fulfilled his promise so that Aaron's neglecting of his promise to Bob was in actuality for

[36] This account becomes more fully developed as our understanding of trust is deepened in the chapters that follow.

Bob's benefit. The important revelation here is that betrayals require a certain intentional attitude on the part of the betraying agent and not simply the failing to fulfill the trusting of another.

Further, finding a particular connection between being trusting and being trustworthy is more difficult than betrayal alone reveals to us because, as Bob's trust in Aaron indicates, even well-placed trust does not guarantee the fulfillment of that trust; it does not take willful action on the part of the other to separate trustworthiness from trusting. Being deceived by the untrustworthy is an understandable failure to recognize the other's untrustworthiness; the goal of the deception is to create the impression of being worthy of trust. But a breach of trust by those who are truly worthy of being trusted is a failure of the trusted individual to fulfill trust in that instant. Failure of trust calls into question the connection between trusting and trustworthiness because it shows that being worthy of trust is no guarantee that the trust will be fulfilled. If breaches of trust do not necessarily label one as untrustworthy than it seems obvious that trustworthiness is not concerned with particular actions and must therefore be something more like an attitude or an intention but a fuller account of trustworthiness must wait until my account of the relationship between trusting and trustworthiness is itself more clearly spelled out.

Failures of trust reveal other things as well; trustworthiness is not always affected by the vagaries of trusting. Being trustworthy is not simply an additive measure of one's fulfillment of trusts. There is no magic number of fulfilled trusts at which point one can be said to be trustworthy. Likewise, individual failures of trust do not make one an untrustworthy person. For a breach of trust to count as a betrayal it must impugn the

trustworthiness of the other.[37] I may intend to fulfill the trust you have in me and yet fail to do so but that failure, depending on the circumstances, may say nothing about my worthiness of being trusted. This, in turn, reminds us that even a single failure can be sufficient to mark one as untrustworthy; some violations of trust are so devastating, whether intentional or not, that the trust cannot be maintained.

Misplacing Trust Without Betrayal

Misplaced trust presents a problem for the possibility of a connection between trusting and trustworthiness as well; we often trust those who we should have distrusted and distrust those who we should have trusted. Our trust does not have to be betrayed or failed in order for us to realize that we should not have trusted. Suppose I have a sitter who has taken care of my children on numerous occasions. I have found this sitter to always be responsible. I am pleased with the sitter's work and have never found any reason to doubt him. I trust the sitter with the care of my children; I am comfortable leaving my children in his care. And then, one day, while speaking with him, I discover that his position on child rearing and punishment are diametrically opposed to mine my own. Or, I discover that my sitter has no real interest or concern for me, my children, or my interests; his concern is for things that make me uncomfortable (world domination, white power, or maybe even increasing the market share of cheese related products). Whatever we take to be reasons for trusting there will turn out to be a corollary set for reasons for not trusting. So far, we have not addressed what will justify trusting and so it seems that even otherwise ridiculous things might fall into this category. The revelation of those counter-reasons will provide me with the justification for not trusting. Therefore,

[37] See chapter 4 for a discussion of what constitutes impugning the trustworthiness the other.

a conversation with my sitter about the larger set of things that he values could cause me to realize that I had misplaced my trust in him by revealing that he values things that I take to be reasons for not trusting. And, I would suddenly think I was wrong to have trusted even if the sitter had never done anything in particular of which I disapproved. In the example, what I had mistakenly taken to be reasons for trusting the individual were not in fact related to the sitter but the sitter had done nothing to try to hide his true feelings from me; this mistake would have been mine alone. Deception is often easier to handle emotionally than the recognition that our trust was misplaced; when I am deceived I can comfort myself that I was misled but when I (mis)place my trust in one who has not tried to deceive me the onus of the mistake rests solely on my shoulders. The mistake remains a mistake even though the trust was not broken because the error resides with my judgment and not with the consequences of that judgment.[38]

Worthiness Does Not Entail Trusting

Even if betrayals, failures, and misplaced trust were not enough to question the idea of a necessary connection between trusting and trustworthiness there would still remain the problem of failing to trust those who are worthy of trust. We can explain the error of deception, we can excuse the error of failure, and we can even justify the error of misplacing our trust but if being trustworthy does not result in being trusted then the idea of a necessary connection between being trusting and being trustworthy seems irreparable. The lack of a connection between our trustworthiness and how we appear to others has a long history in ethics, existing as early as Plato's discussion of justice in the

[38] Although it makes sense to ask what my judgment is in reference to at this point, the question is slightly premature. What has just been shown about the error of our judgment is leading to, not from, a conception of trust.

Republic; it forms the crux of the discussion of justice. The goal, for Plato, was to determine whether it is better to be or merely to appear just. We can see in this search exactly the issue at question here; appearing trustworthy is not the same as actually being trustworthy. Because appearing trustworthy does not entail being trustworthy we often fail to trust, or we go so far as to distrust, those who are actually worthy of trust. Misplaced distrust, then, reveals to us that being trustworthy is not guarantee of being trusted; the connection between trusting and trustworthiness is again shown to lack a direct or causal connection.

Generalizing the Case

We have seen that trusting and trustworthiness are not directly and inextricably linked. However, an obvious criticism of the discussion presented so far would be to point out that my argument, so far, only deals with trusting and trustworthiness between persons. It may be suggested that within a single individual the two are connected; I have only shown that trusting and trustworthiness are separated between individuals. This response holds that the moral fiber required to be a trustworthy person would lead one to be trusting as well, that a lack of one would lead to a lack of another: The Prisoner's Dilemma, the subject of the third section of this chapter, treats trusting and trustworthiness in just this way. As appealing as intrapersonal trusting and trustworthiness might be it is not obviously the case. In fact, it would seem that it is obviously not the case. A single person can, in the span of even one hour, exhibit a surprising variation of trusting and trustworthy behavior. That I am trusting does not mean that I am trusting to all people equally or even that I am trusting to everyone. I may

trust some while distrusting others. I may even be distrusting of one person at one time while being trusting of the same person at another time. Likewise, that I am worthy of trust does not entail that I will be trustworthy to everyone or in all interactions, which is something that my final account of trust must take into consideration. For now, however, we realize that even within individual people trusting and trustworthiness are not inextricably tied together. But saying that trusting and trustworthiness do not have a necessary connection within the individual is merely to point out that we have the capacity to be multifaceted in our responses. This must be true, otherwise our actions would all be uniform and unvaried; we would behave like machines, with no possibility of judgment and no room to alter our responses.

These distinctions show us why Glass's account fails beyond even the problems listed previously. Since not all breaches of trust count as betrayals, not all failures to fulfill the scientific imperative to be trustworthy can constitute a challenge to the scientific system; if the breach will not necessarily destroy trust then it cannot necessarily destroy science. Therefore, Glass' fear turns out to be misguided. Since trustworthiness is not always affected by the vagaries of trusting, individual actions do not, on their own, cause one to be either trusted or distrusted, the threat to science posed by cheating is overstated. For example, a scientist might lie in reporting her results by understating her case. She might, as a matter of course, want to hedge her bets about the data in question and make a much less substantial claim. Given a simplistic, action-oriented account of what trustworthiness is, she would fail because she would not be telling the truth about her data. That individual case would be, if Glass were correct, enough to discount not only the conclusions in question but all the experiments that she had ever been involved

with. Further, from his perspective, it should cause us to start doubting science in general. However, far from causing us to distrust the scientist it might actually encourage us to trust her results because we know that she does not *overstate* her results; we can assume that her conclusions are not unsupported by the data. There might be other problems with reporting data in the manner that she does but surely it is not because it causes a specific and general distrust of her results.

The problems that we find in Glass' account however are merely a subset of the problems that we find within the Prisoner's Dilemma, which is far more important to current discussion of trust. In the next section I investigate the Prisoner's Dilemma and discuss its relevance to an analysis of trust.

Prisoner's Dilemma

The Prisoner's Dilemma is commonly used as a way to test and measure trusting. It is part of the foundation for discussions of trust, which leads authors like Russell Hardin to talk about trustworthiness as "holding your interests in my interests."[39] But the Prisoner's Dilemma is a very special kind of trust relationship. It, unlike most other instances of trust, is reciprocal in a very direct way: when I am trusting of you and cooperate, I am, at the same time, being trustworthy. Within the model of the Prisoner's Dilemma an action is simultaneously trusting and trustworthy. This leads one to assume that there is a direct and immediate connection between trusting and trustworthiness that I argue does not extend to other trusting encounters and is not, therefore, integral to our conceptual understanding of trust. When we consider those other non-reciprocal

[39] Russell Hardin, *Trust and Trustworthiness*, The Russell Sage Foundation Series on Trust ; (New York: Russell Sage Foundation, 2002).

encounters, and the ways that trusting is broken, we can re-interpret Prisoner's Dilemma interactions so that the act of cooperating need not be seen as trusting; cooperating may be simply an act of being trustworthy without any direct concern for being trusting on the part of the trustworthy individual. Further, the act of cooperating, or defecting, in the Prisoner's Dilemma might be strategic and not connected to trust at all. In this section I examine the Prisoner's Dilemma. I then show how the Prisoner's Dilemma confuses the various motivations for acting into a single, simplified, explanatory motivation. Finally, I conclude that, as a model, it is insufficient for explaining and understanding trust.

Prisoner's Dilemma in Context

In the 1950's and 1960's tensions between the United States and the Soviet Union were running high. The Soviet ideology was different enough from American sensibilities that citizens were uncertain whether or not "the Russians" would lie in order to get their desired ends; a common way for stating this concern was to ask whether or not we could *trust* the Soviets. Fred Schwarz, in his book, *You Can Trust the Communists*, tries to show that communists are indeed worthy of our trust. He writes that "[o]nce we accept the fact that Communists are Communists, and understand the laws of their thought and conduct, all the mystery disappears, and we are confronted with a movement…which is perfectly understandable and almost mathematically predictable."[40] We can see from the quote what he means by trust: the mystery surrounding Communist behavior disappears and they can be seen as predictable and even regular. Schwarz takes trustworthiness to be the completion of some individual or set of actions; specifically, actions that the other has

[40] Frederick Charles Schwarz, *You Can Trust the Communists* (Englewood Cliffs, N.J.,: Prentice-Hall, 1960), 1.

claimed they will fulfill such that we can predict his behavior. 'Trusting the Communists' amounts to believing that they will act in the manner which they have claimed they will act; fulfillment of a particular action or a particular behavior is the target of trust in this case. Schwarz's commitment to this position is evident in the subtitle of his book, which is "to do exactly as they say."[41] That is, Schwarz believes that we can expect the Soviets to do exactly what they have claimed they will do, which, given events of the time period, was of paramount importance.[42] I will argue in Chapter 3 that the lack of mystery and predictability are themselves problematic for a conception of trust, but for the moment I point them out only because they illuminate Schwarz's conception of trust, which will allow me to make my final distinction about the relationship between trusting and trustworthiness: that trusting and entrusting are not equivalent.

Schwarz, in considering how to deal with the Communists, is worried about the consequence of being wrong; believing that the Communists will not follow through on their claims could have disastrous consequences but believing that they will act in a particular way, when in fact they will not, can still be damaging. While Schwarz is not alone in this concern, conceiving of trusting as fixated on the fulfillment of specific actions misses the point of the separations discussed in the second section of this chapter: 1) untrustworthy people can act to fulfill certain instances of being trusted when it suits a greater deception; 2) trustworthy people can fail to fulfill trust by mistake or accident; 3) one can realize that he should not have trusted even though no breach of that trust has

[41] It is interesting to note that the 1962 paperback edition of his book carries a different subtitle; "to be Communists." While this title de-emphasizes the truth-telling of the Communists it retains the predictability that is so important to Schwarz's position.

[42] In 1958 Nikita Khrushchev famously said "We will bury you!" Taken out of context the reported statement was a frightening threat to most Americans that caused the populace to wonder whether the Soviets really would.

occurred; and, 4) one can fail to trust those who are most deserving of it. Schwarz emphasizes the importance of actions to trust, which is exactly what these separations deny. If actions were the determining feature of trust then none of the separations discussed would carry much weight; the rate with which one fulfills trusts would be the most relevant feature when determining whether to trust or not. We should, therefore, be wary of linking trustworthiness to the fulfillment of specific actions because doing so fails to capture the relevant connection: it mistakes entrusting for trusting.

Schwarz's position, both the topic of concern and the resulting conclusion, is part of a greater project, a kind of movement of the period in which he was writing. His concerns about the Communists and his response to those concerns are an extension of the motivations behind and the use of the Prisoner's Dilemma. In 1950 Merrill Flood and Melvin Drescher developed the Prisoner's Dilemma as a way of modeling decision-making behavior in the relationship between the United States and the Soviet Union.[43] As a model, it quickly became ubiquitous in decision theory. It is therefore not surprising that Schwarz, writing a decade later, would utilize the concepts and the terminology inherent in that discussion, especially since his concerns are largely the same as theirs. As an area of research, the Prisoner's Dilemma is daunting; the sheer volume of articles published about it makes any addition to that literature a monumental task. Additionally, the number of disparate fields that deal with it as a topic worthy of investigation makes any attempt to summarize opinions on the Prisoner's Dilemma an easy target of criticism. It is not, however, my goal to provide a comprehensive discussion of the issue. Instead, I want only to show how concern for the Prisoner's Dilemma structures our understanding

[43] Steven Kuhn, "Prisoner's Dilemma," The Metaphysics Research Lab Center for the Study of Language and Information Stanford University, http://plato.stanford.edu/archives/win2007/entries/prisoner-dilemma/.

of trust; it focuses our attention on specific components of decision-making in interpersonal relationships. Further, I want to show that the Prisoner's Dilemma is, as a model, inadequate for explaining trust.

Cooperative Matrix

The simplest description of the Prisoner's Dilemma is any decision scenario in which one individual can improve his situation by defecting and sacrificing the position of the other but in which cooperation produces the best combined result. Provided that the particular values used in the discussion conform to certain parameters they do not carry the weight of the argument so whichever values we use are irrelevant to the position being presented. Morton Deutsch provides one of the earliest examples in which the Prisoner's Dilemma is used to explain trust but, personally, I find his formulation to be somewhat confusing.[44] Instead, I use the more common formulation found in Robert Axelrod's book *The Evolution of Cooperation* that focuses strictly on rewards rather than on a mixed set of penalties and rewards.[45] In Axelrod's formulation of the decision matrix rewards appear thusly:

	Cooperate	Defect
Cooperate	3, 3	0 ,5
Defect	5, 0	1, 1

The important aspects of the Dilemma are that defecting can produce the best result for the individual actor while cooperating can produce the best result for both actors taken

[44] Deutsch, "Trust and Suspicion." His formulation uses rewards and penalties that requires the reader to consider positive and negative values, which, in some real-world situations, are hard to translate.
[45] Axelrod, *The Evolution of Cooperation*, 8.

together. Alternately, defecting can produce the worst result for both actors while cooperating can produce the worst result for the individual actor. Broken down according to the scheme above, this means that choosing to cooperate brings with it the possibility of receiving a reward of either 0 or 3 while defecting brings with it the possibility of receiving a reward of either 1 or 5. A successful strategy is one that produces the most points. The most successful strategy is to always defect when your partner will cooperate because it gives you 5 points per round. The least successful strategy is to cooperate when your partner defects because that gives you zero points per round. It is important to note that trading defections does not improve one's situation; alternating between defecting and cooperating produces an average score of 5 for every two rounds of play, provided that each person participates in the trading defection scheme.[46] Consistent cooperation, in comparison, produces an average score of 6 for every two rounds, which makes consistent cooperation a more successful strategy than trading defections. Additionally, both trading defections and consistent cooperation strategies are more successful than continuous, combined defection; each player only gets 2 points for every two rounds if they continuously defect. In fact, even if one player cooperates in the first round so long as the defection is reciprocated on the second turn then a two-turn Prisoner's Dilemma produces the same result as cooperation but it quickly becomes outdistanced by cooperation on any iteration larger than two. Said differently, the utility

[46] Participating in a scheme to trade defections though would, on the Prisoner's Dilemma model, count as a trusting relationship.

value of a mixed-choice pair (one player defects while the other cooperates) is 2.5 while the utility value of a cooperative pair is 3 and a defecting pair is 1.[47]

It is often argued from the facts of the Dilemma that, regardless of what your partner does, the best move for the individual is to choose to defect; if my partner chooses to cooperate, then by defecting I get the best possible result (5 points) compared to the second best result (3 points) that I would get for cooperating under those conditions. If, however, my partner chooses to defect then I would be foolish to not also defect; cooperating when my partner defects gives me the worst possible payout (0 points) while defecting when my partner defects gives me the second worst payout (1 point). It seems then that the best move is to always defect because I am always better off if I defect than if I cooperate. No matter what my partner does it seems that I am better off defecting. If my partner cooperates then I do better by defecting and if my partner defects then I do better by defecting. The dilemma arises however when we conclude that the 'best' move is to defect. Each person in the decision-scenario is capable of coming to the conclusion described above, which should mean that both players defect. If that happens though, then the second **worst** reward is given to each person; they each get only 1 point. Gordon Tullock draws our attention to this; "if both parties make the same decision, they are better off if that double decision is 'don't squeal' than if it is 'squeal'."[48] Taken together then, the participants could be much more successful by cooperating; that is, the problem raised by the dilemma is that *our* best move is to

[47] If we assume that the partner will act randomly then the utility value of choosing to cooperate is 1.5 while the utility value of choosing to defect is 3. This seems to indicate that defecting is by far the better choice; it has double to utility value of cooperating. But, we must remember that this assumes that the other person is going to act randomly and we do not normally think of agents as being chaotically determined. This will have important implications for the value of using the Prisoner's Dilemma in discussions of trust.
[48] Gordon Tullock, "The Prisoner's Dilemma and Mutual Trust," *Ethics* 77, no. 3 (1967): 229.

cooperate and not to defect while *my* best move is to defect and not cooperate. The apparent dilemma, then, is between acting on individual self-interest and acting on group interest.[49]

Invented at the beginning of the Cold War as a way to make predictions about the rational behavior of another, the Prisoner's Dilemma was intended to help us make the best decision in our dealings with others; specifically, within the context of America's concerns for the use of nuclear weapons.[50] It was initially believed to describe the conditions under which people made decisions and that decisions under these conditions occurred at every level of our lives. We can find examples of the Prisoner's Dilemma in relationships as common as the one we have with our local grocer and we can see the influence of the Prisoner's Dilemma on scales as large as the policy of Mutually Assured Destruction, America's official policy during the 1960's.[51] Since its invention however, the Prisoner's Dilemma has become more than a model, more than merely descriptive; it has become normative. Since the 1950's several generations have been raised on the idea that the best strategy for decision-making is to consider how the other is going to behave first rather than deciding independently which action we believe is correct; to decide how to get the best for ourselves in the interaction. This decision strategy leads to the conclusion that the 'best' move in the Prisoner's Dilemma, and in life, is to always

[49] Discussions of the Prisoner's Dilemma consider one-shot interactions, limited iterations, and indeterminate or unlimited iterations. For each area of discussion there is disagreement about what the rational choice is. The disagreement though is what makes the issue interesting. But, it is important for my purposes not because it lacks a convincing answer but rather because it continues to be used in relation to trust. Although there is disagreement about the response, there is no disagreement about its use.

[50] "Best" in these discussions is understood as 'most successful' at producing some economic or military result.

[51] MAD attempts to create conditions under which the penalty for defection is so great that it outweighs any possible benefit to be gained from that defection. Stated this way the Prisoner's Dilemma actually resembles Pascal's Wager.

defect.[52] This result is based, in-part, on the assumption that agents are, and should be, risk averse; that our driving motivation is to avoid risk or harm. In the prisoner's dilemma, in order to avoid the worst possible scenario, we must defect; defection, while bad if both parties choose it, is still better than the consequences of cooperating when the other does not. But the conclusion that defection is best, in addition to assuming that humans are risk-averse, assumes that self-interest is a natural imperative that functions in opposition to group-interest. On this model if we believe that the other is going to cooperate, we are led to the conclusion that we should defect and thereby get the most for ourselves. The model, which initially was concerned with decisions made under the worst possible conditions (where failure might mean nuclear annihilation), became standard for common, every-day decisions.

The difficulty in resolving the dilemma provides us with an indication why so many authors have trouble with the place of trust in society; defecting is treated as a rational choice and cooperation is irrational. Returning this decision-making strategy back into a conception of trust we find that trusting others is a bad idea, it is irrational; in trusting we set ourselves up to be taken advantage of. Many authors fear exactly this result; Russell Hardin argues that a trusting person, if he is trusting in all situations, will necessarily be taken advantage of.[53] The fear of skepticism suggested by Bentley Glass and Marcel LaFollette, the fear that stems from a causal connection between trusting and trustworthiness, is generated by a concern for the defection of others and either a desire to get the most for ourselves or to avoid the worst for ourselves. It should be no surprise

[52] In any limited iteration of the Prisoner's Dilemma logic seems to dictate that we should defect because the smart move is to defect on the final turn. But anyone can recognize this and so the smart move regresses to defecting on the first move.
[53] Hardin, *Trust and Trustworthiness*.

then that once we have accepted the Prisoner's Dilemma as a model for human behavior, trust becomes an issue; there is no good justification for trusting in this model. What needs to take place, in order to alter our current poor opinion of trusting is to change the model for rational human behavior. This, it seems, is, in part, what Francis Fukuyama is striving to achieve in his book on trust.[54] All of which is part of the complexity and difficulty in discussing the Prisoner's Dilemma. Discovering the rational response to the dilemma, however, is not my direct interest. Focusing on the rational response, I argue, is part of the problem; the Prisoner's Dilemma is inappropriate as a model for trust.

Not a Good Tool For Conceptualizing Trust

Whatever problems arise from trying to solve the Dilemma, those problems are mostly tangential to my current concern, which is to elucidate the reasons that the Prisoner's Dilemma should not be used as a tool for conceptualizing trust: it both conflates three separate motives for acting and it conflates trusting with actions that follow from trusting (as well as trustworthiness with actions that follow from trustworthiness). These problems are related to but separate from the distinctions already drawn in this chapter, some discussion of this issue is therefore in order.

There are three motivations that can be inferred from a decision to cooperate in the Prisoner's Dilemma: 1) an actor can be trusting that his partner will cooperate; 2) an actor can be trustworthy to his partner; and 3) an actor can be strategically choosing in his self-interest without concern for being either trusting or trustworthy. The first and second motivations, trusting-that and trustworthy-to, are necessarily conflated in the Prisoner's

[54] Francis Fukuyama, *Trust : The Social Virtues and the Creation of Prosperity* (New York: Free Press, 1995).

Dilemma because any decision made within the scenario can be interpreted to be motivated by both trusting and trustworthiness; nothing about the choice to cooperate in the Prisoner's Dilemma allows us to discount the alternate motivation. Being trusting in the Prisoner's Dilemma is choosing to cooperate while being trustworthy is also choosing to cooperate. Consequently, we cannot know whether a prisoner who keeps his mount shut is doing it because he wants to be trusting of the other or because he wants to be trustworthy to the other. In addition to being indistinguishable the two motivations are inseparable when considered from within the Dilemma's scenario; it gives the impression that any trusting action is also a trustworthy action; because we cannot tell which concern is causing the prisoner to act we mistakenly assume that it is because he is concerned with both. This inseparability can be traced back to Deutsch's initial study. He stipulates a necessary connection between the two; "[t]he essential psychological feature of the game is that there is no possibility for 'rational' individual behavior in it unless the conditions for mutual trust exist."[55] By suggesting that each person in the game must be trusting of the other in order to make a rational decision Deutsch limits the scope of the Dilemma in relation to explaining trust. His focus on a case of mutual trust assumes that trusting and trustworthy behavior are the same and so he neglects the idea that one prisoner may be trusting while the other is being trustworthy, which presents as being a case of mutual cooperation.

What seems to be overlooked by most, however, is Deutsch's suggestion that the trust must be *mutual*. The Prisoner's Dilemma, as it is described in Deutsch's 1958 paper, requires *both* parties to be trusting. That is, he sees this as an example of reciprocal

[55] Deutsch, "Trust and Suspicion," 270.

relationship. If the Prisoner's Dilemma were an example of the common conditions in trust relationships then Deutsch would not need to use 'mutual' in order to describe its essential feature but being mutual is, as he describes it, absolutely necessary for understanding how people respond in the Prisoner's Dilemma. That is, 'mutual' does not usually describe the conditions of a trust relationship. This suggests that the two forms of trust, trusting and trustworthiness, rather than being one and the same are really, perhaps necessarily, separate within the individual. If the dilemma did *not* require mutual trust, if trusting did *not* entail trustworthiness then we could reasonably expect trusting and trustworthiness to be linked within the individual in all cases; there would be no obstacle to them being one and the same. In the Prisoner's Dilemma, however, choosing to be trusting and to cooperate, which is to accept the possibility of getting the worst possible result, is also a choice to be trustworthy, which is to give up the possibility of getting the best possible result.[56] The Prisoner's Dilemma, according to Deutsch's argument, requires that each person be trusting **and** trustworthy, which as a particular requirement indicates that it is not present in all trust relationships. It is the fact that the Prisoner's Dilemma requires *mutual* trust, and not simply trusting or trustworthiness, which has been overlooked. Meanwhile, the simultaneity of trusting and trustworthiness which is part of the mutual trusting of the Prisoner's Dilemma has been accepted as part of trust relationships in general. Our conception of trust is therefore misled by discussions that rely too heavily on the Prisoner's Dilemma because even though we see both trusting and

[56] I take this description of what it means to be trusting and to be trustworthy as merely an extension of how the terms are being used in discussions of the Prisoner's Dilemma. I am presenting these, at this point, only as tentative explanations. While I am beginning this investigation with the least amount of conceptual baggage possible, it seems that we have to accept some things in order to develop our understanding, whether or not we continue to hold those first suppositions.

trustworthiness exhibited in the Dilemma we forget that they are separate; we take the one action, cooperation, to be an example of both simultaneously.

This explanation, however, is still not correct. The Prisoner's Dilemma seems to require us to be both trusting and trustworthy at the same time, it seems to require mutual trust, but that requirement is an illusion. Suppose that Knox and Larry are caught shortly after committing a crime. The police, as good interrogators, separate the two and try to get each to turn on the other thereby creating a Prisoner's Dilemma. We can imagine Larry thinking that he does not want to be a snitch but still being afraid that Knox is. Larry might then choose to confess in order to protect himself, deciding that he was not trusting enough *of* Knox in order to be trustworthy *to* Knox. Larry is showing exactly the kind of mutuality that Deutsch is describing; if Larry wanted to be trustworthy then he would have to be trusting as well but he is not prepared to be trusting and so he cannot be trustworthy. Knox, however, may not need such a convoluted decision process in order to figure out what he should do. Knox might simply decide that no matter what happens he will not squeal. He can be trustworthy without any concern for being trusting; he is choosing not to pursue the possibility of the best result and ignoring the possibility of getting the worst result; Knox does not have to consider the entire relationship in order to be trustworthy or even to appear to fulfill the behavioral conditions of trustworthiness required by the game. The Prisoner's Dilemma does not, therefore, require mutual trust at all, it only appears to. By appearing to require both trusting and trustworthiness it fails to distinguish the motivation of the person acting; it fails to distinguish between trusting and trustworthiness. And, since we have already seen that trusting and trustworthiness are not inextricably, causally linked then failing to distinguish between them presents serious

problems for a conceptual analysis; it cannot tell us which of the two motives is driving the decision. The Prisoner's Dilemma should not be our model for understanding trust. But, even Deutsch overlooks the revelation that can be gleaned from the conditions of the Prisoner's Dilemma; "one can say that a co-operative orientation will produce trusting (and trustworthy) behavior."[57] Deutsch's suggestion that a co-operative orientation will lead to both trusting and trustworthy behavior is incorrect because the game misleadingly fails to distinguish between those activities.

Failure to distinguish between trusting and trustworthiness is not, however, the only problem with the Prisoner's Dilemma, it also fails to distinguish between trust(ing) and strategic motivations. This failure is far more damning; thinking strategically is not trusting. We have already seen that acting strategically does not require one to be trusting but thinking strategically is a natural response when confronted with the rational choice presented by the Dilemma. This is because the Prisoner's Dilemma itself leads one to think strategically. The focus and concern in deciding how to act turns to the numerical values inherent in each decision rather than anything particular to the individuals involved. The way in which I presented the Prisoner's Dilemma above is evidence of this kind of thinking. It is hard, perhaps impossible, to consider the Dilemma outside of this context; the consequences, rewards and penalties, of the decision drive the decision-making process. Attempts to answer the question "how should I act" or "what should I decide" are driven by the attention drawn to the rewards of each decision pair rather than the relationship of the agents.[58] By looking at the decision *sans* relationship between the

[57] Deutsch, "Trust and Suspicion," 272.
[58] I am not actually arguing that the frame of the question is (mis)leading us to an incorrect answer but it is a useful analogy for making the point. What I am suggesting is akin to the kind of misdirection apparent in

participants we are necessarily removing a consideration of trust from that decision. This leads attempts to solve the dilemma, like Axelrod's, to become cases in which the decision is being made strategically instead of revealing a concern for either trusting or trustworthiness; it forces us to make the decision without regard to the personal feelings one has towards the accomplice. Even when we consider the strategies involved in Axelrod's multi-iteration, generational study, like TIT-for-TAT, we are still faced with what is, in the end, a purely strategic decision; although the decision is made with reference to the past relationship of the participants, the goal of successfully getting the greatest reward still provides the criteria for evaluating that relationship.[59] Gordon Tullock, while arguing that most authors imply that the issue of the Prisoner's Dilemma is one of trust, actually denies that trust is the issue at all; he reminds us that "one prisoner's opinion about the probable behavior of the other is irrelevant to his own decision, since his payoff will always be better if he confesses."[60] That is, the strategic move is to act without reference to how the other will behave; in a strategic response to the Prisoner's Dilemma trust is necessarily not an issue. Yet we continue to have authors who conceive of the Dilemma and its resulting choices as an issue of trust, trust that is generated by a consideration of rewards and penalties. For example, we see it in the conception presented by Bhattacharya et al. when they say "trust follows from the ability

---

riddles and magic. By focusing your attention on the rewards and penalties I draw your attention away from another aspect of the decision, an aspect that I argue is more important and relevant to the issue of trust.

[59] TIT-for-TAT, the most successful strategy in Axelrod's multiple iteration Prisoner's Dilemma study, is to respond in kind to the last way you were treated. If, in your last interaction, the other defected, then you should defect. If, on the other hand, the other cooperated, then you should cooperate. This strategy seems to move beyond a purely mathematical decision because it pays attention to how we have been treated in the past but the criteria for making the decision remains the same; it forces us to answer the question by considering how to get the most points possible instead of using other aspects of the decision.

[60] Tullock, "The Prisoner's Dilemma and Mutual Trust," 229.

to structure contracts or rewards and punishments so that individuals behave in a prespecified manner."[61]

Not only is it the case that the Prisoner's Dilemma fails to distinguish between trusting and trustworthiness but it also fails to distinguish between trust and acting strategically. This brings us to the penultimate point of this chapter: when we consider how the Prisoner's Dilemma can mislead us in our conceptual understanding of trust and we accept the distinctions made earlier about the connection between trusting and trustworthiness we can see that the biggest and most common mistake made in any discussion of trust is to mistake behavior with trust itself.

If we return to the example of Knox and Larry the trust/action distinction will become clear; Larry had a convoluted process through which he decided how best to respond while Knox simply chose not to snitch. The decision that each made was not whether to be trusting or distrusting; the choice that each made was the result of trust (or a lack of it), not an example of it. The Prisoner's Dilemma fails to distinguish trusting, trustworthiness, and strategic actions; it fails to distinguish actions from the motivations that spur us to do those actions therefore the Prisoner's Dilemma cannot reveal my motivations for acting. This inability is a problem because once we accept the ways that trusting and trustworthiness fail to be connected we realize that actions are not the determining feature of trust. If the Prisoner's Dilemma cannot reveal anything about our motivations then it cannot illuminate the concept of trust.

That I can be betrayed means that a person can mask his motivations. Masking my motivations necessarily means that the reason that I am acting is not apparent in the

[61] Rajeev Bhattacharya, Timothy M. Devinney, and Madan M. Pillutla, "A Formal Model of Trust Based on Outcomes," *The Academy of Management Review* 23, no. 3 (1998): 460.

action itself. For betrayal, my motivations can only be revealed be the consequences of the action itself. Even then, the betrayal is only revealed in the moment that you realize that I intended those consequences. If you accept that those consequences are not intended by me then you will take those consequences, which constitute a breach of trust, to be merely a failure of trust and not an actual betrayal.[62] The agent's intentions differentiate a failure from a betrayal. Realizing that I have (mis)placed my trust in someone who does not deserve it, when that person has not actually broken my trust, also shows us that it is the intention of the agent that is important to trust; even though the other has done me no wrong, no action that would constitute a betrayal, I can decide that I was wrong to have trusted that person. This is only possible if we consider the motivations of the person in question and not the actions of that person. We should not take this to be a denial of the value of paying attention to the actions of others; actions can provide clues to the motivations of others. What I deny is that the actions alone reveal those motivations or that the actions are synonymous with motivations. If the separations between trusting and trustworthiness reveal that our motivations for acting are important for trust and the Prisoner's Dilemma only reveals our actions and not our motivations then the Prisoner's Dilemma does not tell us much about trust.

Connecting

Severing the connection between trusting and trustworthiness helps to avoid problems and misunderstandings but we need to be careful about overstating the case; if

[62] There is a mistake inherent in speaking of trust in this way. A result of the separation that I am arguing for in the present chapter and which I make clear in subsequent chapters is that the consequences are never a breach of trust but are instead a breach of entrusting. It is, however, convenient to hold onto the colloquial way of speaking for the time being.

we deny any connection whatsoever then we are left without the means to explain the association. Thankfully, the various ways that trusting and trustworthiness are not connected lead us to recognize how they are connected. Betrayal shows that trusting does not entail that the other be trustworthy. Failure shows that being worthy of trust does not entail the fulfillment of that trust. Misplaced trust shows that our assessments are not always accurate. But in each of these cases we can see that when we do trust it is because we believe the other is worthy of that trust; trusting is inseparable from an assessment of the trustworthiness of the other. The nature of this assessment remains to be seen and is considered in Chapter Four but for now it is enough to recognize that this is a connection that we cannot sever. It will never be the case that we are trusting of another while believing that they are not worthy of trust. Our trusting of the other is connected not to the trustworthiness of the other but rather to our belief in the trustworthiness of the other. That I trust you does not mean that you are worthy of trust but rather that I believe you are worthy of trust. My trust in you is necessarily connected to my assessment of your trustworthiness. And this, it will turn out, is incredibly important to understanding the concept trust.

In the next chapter I show how these distinctions further our analysis of trust by revealing that trust is risky. This is an important step in the investigation because it will motivate the foundation for what it means to be trusting.

Chapter Three: Trust is Risky

In the last chapter I distinguished the various ways that trusting and trustworthiness failed to be connected. Put simply, trusting another does not necessitate that either she or you are trustworthy. Neither does being trustworthy lead another to be trusting of you nor does being trustworthy entail that you will be trusting of others. These separations reveal that trusting and trustworthiness are not themselves actions and so the act of entrusting is neither an example of trusting nor an example of trustworthiness. Acting *as if* I trust another is to treat the person like one who is worthy of trust, which seems to be acting in a particular manner but, as we saw in the last chapter, actions, even ones that seem to be done in a particular manner, can be done for reasons that are contrary to the aims of trust; con-artists survive on this exact premise by acting *as if* they are trusting of a mark (the target of the con) in order to lure the mark into engaging with the con-artist. It seems then that entrusting or 'acting as if' are, at best, indications that the other may be motivated by trusting or trustworthiness but are, at worst, outright lies. We cannot, therefore, continue to treat trusting, trustworthiness, and entrusting as being directly and causally linked together.

The linkage between trusting and trustworthiness can be shown to be distinct in a number of ways, however, there remains a close relationship between the two, not a necessary causal relationship but something else altogether. I have suggested that the two are connected conceptually via our understanding of what trust means, that to be trusting

of another simply is to believe that the other is worthy of trust . But, this explanation fails to fully account for trust; it still requires much in the way of clarification. If trusting is to believe that the other is trustworthy then it seems we must uncover what it means to be trustworthy, or to believe that another is worthy of trusting, in order to fully understand trust. It is the goal of the next several chapters to expand on this very basic premise by focusing on specific aspects of the trust concept which follow from the distinctions made so far. This chapter is devoted to showing that trust is, indeed, risky.

We can see from the presentation of Frederick Schwarz's account in the last chapter that, for him, trust is risky.[63] Trusting the Communists, on Schwartz's conception, had very serious consequences, so serious that many were unwilling to do so. He takes trusting to be an activity and we can see that perspective in the goal of his work. His book was an attempt to show that although interaction with the Communists was risky they could still be counted on to do what they had claimed; we did not have to *dis*trust them. His main concern was for our interaction with the Communists, which is action, and he considers this by focusing on whether the Communists will *do* what they have claimed that they will, which is also about action. I have already argued that trusting and trustworthiness are not actions so it might seem strange to use examples that treat trust as an action in my effort to show that trust is risky but it is important for the discussion to consider both those who think trust is risky and those who do not and many of those authors take trust to be an action, whether overtly or not.

Schwarz' account of trust as risky is by no means the only one that recognizes this feature of trust. Morton Deutsch, in his early work on trust, suggested a connection that

[63] Schwarz, *You Can Trust the Communists*.

so closely linked trust and risk that "[r]isk-taking and trusting behavior are thus really different sides of the same coin."[64] Kim Giffin says that trust is the "[reliance] upon and object, event, or person in an attempt to achieve a desired but uncertain objective in a risky situation." [65] We can see that one of the key elements of her definition is risk. Russell Hardin, while discussing a case in which a seemingly unsecured emergency loan was given to a ship captain, says "the background possibility of sanction greatly enhances this trustworthiness and justifies Hambros's general risk-taking."[66] And, Jonathon Adler says of trust that it "seems sufficient for a seeker to be extending trust that the seeker is at risk."[67] Trust as risky seems therefore to be fairly well recognized in discussions of the topic. Trust that is risky also matches our common understanding of what it is to be trusting. Statements like "I just can't trust you anymore" express the idea that the demands of trust are more than the person can bear, that the possibility of failure is either too great or relates to something that is too important to the (non)trusting individual; the risk inherent in the act of trusting overrides the capacity of the individual who is wanting to or being asked to trust. With so much agreement about trust's risky nature it may seem strange to devote an entire chapter to the topic but the consensus is not nearly so uniform as it seems.

Russell Hardin adamantly denies that it is trusting which is a risky: "to act on trust is to take a risk, although trust is not itself a matter of deliberately taking a risk."[68] On Hardin's account it is not the trust we have in another that is risky, rather it is how we

---

[64] Deutsch, "Trust and Suspicion," 266.
[65] Kim Giffin, "Interpersonal Trust in the Helping Professions," *The American Journal of Nursing* 69, no. 7 (1969): 1491.
[66] Russell Hardin, "Trustworthiness," *Ethics* 107, no. 1 (1996): 33.
[67] Jonathan E. Adler, "Testimony, Trust, Knowing," *The Journal of Philosophy* 91, no. 5 (1994): 266.
[68] Hardin, *Trust and Trustworthiness*, 12.

choose to act based on that trust that is risky. His position is based on the idea that trust is a form of knowledge that we have about another, as such it cannot properly be called risky. He argues that those who consider trust as risky are mistakenly attaching risk to trust because they treat trust as the subject of choice. But, he points out that "it is incoherent to say I choose to trust you" because I cannot choose to know or not know based on the evidence presented to me and so I cannot choose to trust or not trust based on the evidence presented to me.[69] Hardin is not the only one who objects to trust being risky, however. Carlos Parales-Quenza says that "[a]stuteness is related to risk while trust is linked to certainty."[70] He separates trust from risky not by making a distinction between acting and trusting, like Hardin, but rather by arguing that trust is connected to certainty. Further, Parales-Quenza sets up a direct opposition to the idea that trust is risky in his conception of astuteness. He suggests that astuteness is "a meta-representative ability that enables individuals to transcend personal perspectives in order to adapt to situations of inter-action," which, he argues, *is* risky.[71] It turns out then that there is a long standing discussion about whether trust is risky or not but when we look back with fresh eyes to the earlier statements supporting trust as risk we begin to notice minor discrepancies: Schwarz is suggesting a conception of trust that sounds like accepting a promise, Deutsch is referring to behaviors, and Giffin's position seems to take trust to be goal directed. Even those that accept trust as risk do not entirely agree about what it means for trust to be risky.

---

[69] Ibid., 10.
[70] Carlos Parales-Quenza, "Astuteness, Trust, and Social Intelligence," *Journal for the Theory of Social Behaviour* 36, no. 1 (2006): 43.
[71] Ibid.: 41.

The status of trust as being risky is therefore in question, both because it is unclear what is meant by "risky" and because some deny that trust is, in fact, risky. I begin this chapter by returning to the Prisoner's Dilemma. I argued in Chapter 2 that current uses of the Prisoner's Dilemma fail to provide us with useful information about how and why people trust. However, it can still be quite informative in other regards provided that we are careful to observe its limitations. In this chapter I use the Prisoner's Dilemma to show that trust, by demanding that we be both trusting and trustworthy, is commonly perceived to be risky. The common understanding helps us to become clear about which elements are important to an account of risky trust. I then move to Hardin's criticism of this position. In denying the riskiness of trust he presents several arguments that are not only worth noting but adopting as well; most importantly he argues that trusting is not an act in the same way that entrusting is an act and it is only our actions that are risky. Although I eventually deny Hardin's claim that trusting is not risky, the positive account of trust as risky that I present must respond to Hardin's criticisms; I must make sense of how trusting can be risky without itself being an action. The discussion then moves to Niklas Luhmann, who gives us an account of trust as a reduction of complexity. Luhmann provides us with the basis for one half of the justification of my account of risky trust. He is concerned with social interaction and interpersonal trust, the overwhelming complexity of the world is his starting point, which leads him to claim that trust is a result of uncertainty about the world. The other half of my argument for risky trust comes from Annette Baier, whose account of the vulnerability inherent in trust follows the Luhmann discussion. Her position accepts the uncertainty of trust but also admits to the possible harm inherent in trust. Finally, I bring

these positions together in order to show how trust can be risky without being synonymous with entrusting. But, before I begin to argue for trust as risky a general discussion of risk would be in order.

Risk

"Risk", in some ways like trust, suffers from overuse. In the vernacular it is somewhat undifferentiated, while in the technical literature it has a precise and clear meaning. The various uses of "risk" confuse the issue and lead to misunderstandings about the relationship between trust and risk. For my account to be convincing I must appeal to and make sense of each way that the term is already understood. In this section I argue that in its vernacular and technical uses, risk has divergent meanings while still sharing a conceptual base: both treat risk as uncertainty.

One way of understanding risk, which captures our folk conception, treats risk as potential for harm. Although the common usage appeals to the probability of harm associated with action I argue that it is the harm itself that marks an action as risky. We can see this in our everyday speech. "He took a risk telling her how he felt," appeals to the harm that would be caused to him had she not responded reciprocally; what if she laughed in his face? "She made a risky move quitting her job," focuses on the harm involved with no longer having a steady source of income; what if her new venture failed to work out? Likewise, skydiving is risky because there is a serious threat of injury (or, a threat of serious injury). In each of the examples, there is a possibility of harm but the possibility is not as important as the harm. For example, it would be strange to consider an activity to be risky when it had the recognized and regular potential to result in some

good end; "sleeping is risky because you might end up well-rested" seems silly because it runs counter to our standard use of the term; being well rested is not something that most people are worried will happen to them such that they would work to avoid it. Being "worried" that something will happen admits to the possibility that it will not happen but the concern in that situation is caused by the harm involved and not the possibility. The notion of harm, therefore, drives our common understanding of risk. Russell Hardin's distinction, that only actions are risky, treats risk as potential for harm. For him, because trusting is an epistemological state, a belief, there is no harm associated with merely trusting; we cannot be harmed by our beliefs. Certainly, we can be harmed by actions that we take because of our beliefs but the belief, in this case trusting, cannot itself risky.[72] For Hardin, we put ourselves at risk only when we act. When we consider risk as "potential for harm" we are led to the idea that trusting is not risky because harm is not directly the result of beliefs; it is not our trust that makes us vulnerable, but our actions.

The technical way of understanding risk, meanwhile, treats it as uncertainty; it involves a potential for loss but that loss is not, strictly speaking, equivalent to harm. J Frank Yates describes the risk construct as consisting in 1) potential losses, 2) the significance of those losses, and 3) the uncertainty of those losses.[73] The third point quite clearly shows risk as related to uncertainty but we might still wonder about harm given the focus on losses. It seems that "potential for losses" has the same meaning as "potential for harm" since we usually see harm as consisting in a loss of some kind. This supports the idea that risk is harm, which would make the technical use that I am describing no different than the vernacular use. However, loss is described in the

[72] I will expand on this point later in the chapter.
[73] J. Frank Yates, ed., *Risk-Taking Behavior* (New York: Wiley, 1992), 4.

literature as an outcome that is less than some reference point. Reference points are our starting position or baseline for evaluating alternatives; if my reference point for a given lottery was winning $100 then winning only $50 would be considered a loss. If I expected to get a promotion and a big Christmas bonus but got only the promotion then not getting the bonus would seem like a loss. Loss does not always mean the removal of some thing sometimes it is the removal of the possibility of gaining some thing. What becomes important on this view of risk is the probability of some particular outcome; risk involves the uncertainty inherent in probabilistic outcomes.

Although the technical conception seems to sit in opposition to our folk understanding it turns out that we can find the risk as uncertainty conception within our folk understanding as well. The risk that he took in telling her how he felt was the *possibility* that she would not reciprocate instead of the certainty of it. The risk that she took in quitting her job was the chance that her alternative plan would fail to work, not the certainty of it. Skydiving carries the *threat* of injury, not the certainty of it; there is a probability that your parachute will fail. If skydiving carried with it the certainty of injury, call it skyfalling, we would not consider the activity to be risky because the harm would be guaranteed. But we cannot assume that having a definite harm associated with any particular activity would necessarily lead to avoiding that activity; if the pleasures of skyfalling sufficiently outweighed the pains, we might still jump out of the plane. We often pursue activities in which harm is guaranteed. Running, for instance, will always contain some amount of pain; the pounding on the pavement, the exertion, the straining of muscles all have some, perhaps minor, pain associated with them and yet the joys and pleasure of running are enough to give some people a reason to do it daily. Or, to use a

different example, we might decide that telling the truth is what we should do, even though it causes some pain to oneself or to others. Jeremy Bentham recognized this important point about admixtures of pleasure and pain in his Felicific Calculus; his system suggests weighing them in such a way that pains need not be absent but merely overridden. So it is that Bentham tells us to "[s]um up all the values of all the *pleasures* on the one side, and those of all the pains on the other. The balance, if it be on the side of pleasure will give the *good* tendency of the act upon the whole."[74] It is fairly common then to consider that humans will endure some amount of pain in order to gain some greater amount of pleasure. Because we do not consider the certainty injury associated with an action to be risky we can see then that it is the *possibility* of injury and not the injury itself that makes skydiving risky; guaranteed harm does not stop us from acting.

The folk conception of risk, therefore, is not, strictly speaking, concerned only with harm even though its use is driven by harm; "there is a risk of getting hurt" is suggesting that there is a possibility of getting hurt and not that there is a *harm* of getting hurt. And, this point becomes obvious when we consider that, often, reference to the risks inherent in an activity also reference the harm as well. "He risked losing his life when he jumped off that bridge," describes both the object put in harm's way (his life) and that being in harm's way is uncertain. Because both the folk conception and the technical conception of risk hold that it involves uncertainty I focus my attention on showing that trust involves uncertainty in order to show that it is risky. In addition, I show that trust carries the possibility of harm, a point which will be expanded in Chapter 4's discussion of the importance of goodwill to trust.

---

[74] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, Dover ed. (Mineola, N.Y.: Dover Publications, 2007), 31.

Common Understanding

As we saw in Chapter 2 the Prisoner's Dilemma creates a situation in which an agent must make a decision about how to act. The Dilemma is designed to focus our attention on how our partner will behave, thereby giving the impression that cooperative behavior is motivated by trust. I have already argued that the Prisoner's Dilemma fails to provide the kind of insight into trust that we might have hoped for because it fails to distinguish between trusting and trustworthiness. However, it can still be useful to our investigation if it can help to indentify general properties of trust, properties that trusting and trustworthiness share in common. If the Prisoner's Dilemma is, as I claim in this chapter, confronted on all sides by uncertainty then trust must also be uncertainty. Even though we cannot identify a specific motivation from the decision to cooperate, any of the motivations for cooperating in the Dilemma cannot lack what is required of the Dilemma itself. Therefore, if the Prisoner's Dilemma is risky then strategic, trusting, and trustworthy motivations must all be risky or else they cannot be considered as viable explanations of action within the Prisoner's Dilemma. So it seems that, in order to reveal the manner in which we are at risk in the Prisoner's Dilemma, a further discussion of the Dilemma is required.

Robert Swinth, while discussing the sequence of developing trust says "[o]ne person acts first and takes the ambiguous path which exposes him to the risk of personal loss."[75] The literature on risk may not be obvious but it is fairly clear in its usage. Consider the Allais Paradox: Maurice Allais in discussing risk presented a case in which people, when presented with a choice between two different sets of lotteries, made what

[75] Robert L. Swinth, "The Establishment of the Trust Relationship," *The Journal of Conflict Resolution* 11, no. 3 (1967): 336.

seemed to be very different choices.[76] The risk in the Paradox is not one of harm but one of loss in relation to a reference point. In his experiment people were presented with two sets of choices. The first choice was to pick either Gamble A, which had a 100% chance of winning $500,000 or Gamble B, which had a 1% chance of winning $0, 89% chance of winning $500,000, and a 10% chance of winning $2,500,000. The second choice was to pick either Gamble C, which had an 89% chance of winning $0 and an 11% chance of winning $500,000 or Gamble D, which had a 90% chance of winning $0 and a 10% chance of winning $2,500,000. Most people would pick Gambles A and D, thereby violating the independence axiom. But the specifics of how and why this happens are not especially important for my present purposes; rather, it is that the Paradox is a classic case within decision theory about choices made under the conditions of uncertainty, i.e., risk. More importantly, there is a loss involved but there is no harm. Michael Weber points out that "in the choice between A and B on must sacrifice a *guarantee* of $500,000."[77] It is because we feel as if we are giving up the $500,000 that choosing B would be considered a loss. But the example points to something further: Gamble A is a sure-thing while Gamble B is risky. Gamble B, however, is risky irrespective of A yet is a loss only in reference to A. But we should not be misled by Gamble A being a sure thing, though considering it will help to clarify the issue. Because Gamble A has the certain outcome of $500,000 it is not a risky option. Though it sounds redundant let me emphasize the point: certainty is opposed to risk. I have argued, therefore, that risk means uncertainty but Swinth's claim highlights that trust is not only risky but also that trusting

[76] M. Allais, "Le Comportement De L'homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes De L'ecole Americaine," *Econometrica* 21, no. 4 (1953).
[77] Michael Weber, "The Resilience of the Allais Paradox," *Ethics* 109, no. 1 (1998): 98.

is always a risk of something because trusting is always in relation to its alternative: not trusting.

Swinth argues that the risk inherent in trusting is about personal loss, a kind of harm, while I argue that it is a risk of being let down, which is a loss of trust; on my account the harm caused by a breach of trust is not physical but is to trust itself. Swinth reminds us that trusting choices contain risks of harm and possible benefits; he continues, "the participants should expect trusting behavior of each other and should be able to stand a test period of simultaneously making trusting choices (as in the usual PD game) in situations in which the risks and payoffs are fairly large"[78] Swinth's position then, in discussing trust, treats trusting as synonymous with acting in the Prisoner's Dilemma and, according to him, both are risky. Having returned to the Prisoner's Dilemma and in light of the objections to trust being risky we are faced with the question; "why is the Prisoner's Dilemma risky, anyway?" The answer might seem obvious but it will help our understanding of the issue to consider it more directly. I begin this section by briefly recapping the structure of the Prisoner's Dilemma. Then, I show how it is commonly interpreted to be risky. This is followed by a discussion of a debate between Virginia Held and Gordon Tullock on the status of the Prisoner's Dilemma in trust research. Finally, I end this section by showing how the Prisoner's Dilemma can be understood to be risky even without appeal to the harm associated with mismatches of defection and cooperation.

[78] Swinth, "The Establishment of the Trust Relationship," 336.

Restating the Scenario

The basic structure of the Prisoner's Dilemma puts two people in a situation in which the best possible outcome for the individual can only be achieved by defection (rather than cooperation) and in which the worst possible outcome can only appear by cooperating but in which the best possible outcome for both parties simultaneously can only be achieved by both parties cooperating (rather than defecting) and in which the worst possible outcome for both parties simultaneously can only appear by defecting. The rewards that I used on page 58 were 5 for a successful defection, 3 for a successful cooperation, 1 for an unsuccessful defection, and 0 for an unsuccessful cooperation.[79] If the two parties are going to respond in the same way, then they can get the most points by cooperating whereas if the two parties are going to respond differently one can do substantially better by defecting.

Choice in the Dilemma

The harm that is often associated with the Prisoner's Dilemma can be seen clearly when we focus on the effects of cooperating; choosing to cooperate in the game, which many consider to be a choice to trust, puts the player in the position of getting the worst possible outcome. On either the colloquial understanding of loss as harm or on the technical use of loss in relation to a reference point, getting the worst possible outcome will surely count as a loss. The decision to cooperate when faced with a Prisoner's Dilemma is risky because cooperating is the option in which the most harm can befall the

[79] Here I am using "success" to mean the fulfillment of a desired goal or end. In the case of defection this would be to be choosing differently than your partner while in the case of cooperation it will be choosing the same as your partner. But, this sense of success is concerned entirely with getting the most points. One could argue that a defect was successful only when your partner *also* defects, since that move would then justify the decision not to cooperate. I leave that discussion for others to pursue.

player. Since most take cooperation in the Prisoner's Dilemma to be a choice to be trusting it seems as if trusting is the option in which the most harm can befall the player.

Deutsch reminds us that "[i]f each player chooses to obtain either maximum gain or minimum loss for himself, each will lose."[80] This position, unlike many others, recognizes that there are two separate motivations that can lead to the same action; a desire for the most gain should lead to defection the same as a desire for the least loss, which reveals that the Prisoner's Dilemma fails to differentiate even between risk-seeking and risk-averse behavior. Choosing a sure-thing over a risk is risk-averse behavior while choosing a risk over a sure-thing is risk-seeking behavior. When there is no sure-thing, risk-averse behavior seeks to reduce the amount of uncertainty in a decision by causing each possible outcome to be closer to some specific value while risk seeking behavior seeks to increase the amount of uncertainty in a decision by causing each possible outcome to be farther from some specific value. For example, in a lottery that had only two outcomes one would be risk-averse if one increased the value of the lower ticket, thereby decreasing the difference between the expected utility and either possible outcome; risk-seeking behavior would be the reverse, increasing the difference between the expected utility and any possible outcome. Deutsch's claim, then, besides revealing that the Prisoner's Dilemma remains ignorant of risk-averse or risk-seeking behavior also points out that no matter how the choice is framed, whether of loss or of reward, the decision can never avoid the loss associated with the scenario. Choosing to seek the greatest reward for your self is not, in the case of the game, going to be a decision to never be hurt or to incur a loss. Choosing to minimize the loss openly admits

[80] Deutsch, "Trust and Suspicion," 270.

to being motivated by the thought of that loss. We can see then that the Prisoner's Dilemma, on the standard interpretation will always, in some way, reference the loss contained in the decision. This means that the Prisoner's Dilemma is risky even on the account of risk as loss. Further, the Prisoners Dilemma is risky on the account of risk as uncertainty because it is a stipulation of the game that you cannot be certain what the other player will do. Since the Prisoner's Dilemma is necessarily risky, any explanation of choice made in the Dilemma must account for that risk. Each of the motivations for acting in Dilemma must be risky or else there would be some cases of the Dilemma which are not risky. Therefore trust, at least in the Dilemma, is risky. I will eventually return to the idea that the Prisoner's Dilemma fails to achieve most of what we want from a model of trust but this much, at least, it gets right: trust is risky. Although, as we shall see later in this section, it gets it right for the wrong reasons.

Held and Trustability

Virginia Held, who also uses the Prisoner's Dilemma as a model for trust, wonders "is it *rational* or not to take a chance on the goodwill or 'trustability' of another human being?"[81] Her question is perfectly natural given the trajectory of the discussion surrounding trust and the Prisoner's Dilemma which leads up to her paper. When decisions to trust and trusting are subjected to the kind of analysis that immediately precedes this part of my paper it easily turns to the strategic mode of understanding the decision. Since 'taking a chance' draws our attention to the harm that will come from a poor outcome, her account also considers trust to be risky. But more than that, Held

[81] Virginia Held, "Rationality and Social Value in Game-Theoretical Analyses," *Ethics* 76, no. 3 (1966): 216.

considers whether the risk associated with trust is justified. We can see in her concern the importance of risk to our understanding of trust; the harm associated with trust tempers our decision by becoming a factor in the analysis of available options. Are we being reasonable when we take a chance? Is it logical to jeopardize our current situation by leaving it in the uncertain hands of another?

Held, in describing the work of Jon von Neumann and Oskar Morgenstern, tells us that they prescribe the "'best of the worst possible outcomes'"[82] which is a suggestion to minimize your possible losses. On her account, a Von Neumann-Morgenstern agent is risk-averse and would probably defect if given half the chance. But this suggestion raises the question: what is the best of the worst possible outcomes when it comes to trusting and/or the Prisoner's Dilemma? Since there are four possible outcomes in the Prisoner's Dilemma, we can imagine that two of the outcomes be in the "worst" category while the other two would be in the "best" category. In each of the categories there would be a better and worse outcome. It seems obvious that getting nothing is the worst possible outcome overall that one could get in playing the Prisoner's Dilemma, which means that it is the worst outcome in the worst category. An instance of simultaneous defection, getting 1 point, should be the second worst outcome overall because it is in the bottom 50% of points possible, which should make it the best of the worst category. This would mean that getting 3 points would be the worst of the best category while getting 5 points would be the best of the best category. But, there is a problem with this conception. Getting the worst outcome of the worst category results from cooperating but there is also a best outcome that results from cooperating, which gives 3 points Therefore, the best

[82] Ibid.

possible outcome under the conditions that can also lead to the worst possible outcome could also be the best of the worst. This should rub readers the wrong way since it implies that cooperating is the worst category.

Held then refines her point in the simplified question "*is it rational or not to trust*?"[83] Here again, we see the importance of the risk associated with trusting; this would indicate that 'trusting' contains within it the concept of risk, since 'taking a chance on the trustability' is replaced by trust. However, and although it is perhaps unfair to characterize it quite this way, Held overlooks the inherent ambiguity of the Prisoner's Dilemma trust relationship, which I described in Chapter 2; one cannot distinguish between an act to trust, an act to be trustworthy, or an act to be strategic.

Held may take the Prisoner's Dilemma to be about trust but Gordon Tullock certainly does not. He points out that most authors involved in the discussion "make statements implying that the problem is one of *mutual* trust."[84] It is actually worth noting that Tullock thinks of the Prisoner's Dilemma as an attempt to present a case of mutual trust; precious few who follow in Morton Deutsch's footsteps are actually clear on this account. What is even more interesting, however, is that Tullock denies that the Prisoner's Dilemma actually *is* a case of mutual trust; he believes that it does not matter what you think about the other when making a decision about how to behave in the game. Tullock supports his claims about mutual trust with quotes from three articles; one each from Robert Wolff, George Thompson, and Virginia Held. However, what each quote supports is that the authors are concerned with the rationality of trust and not whether it is mutual in nature. Held's quote I have already used above, she asks whether it is rational

[83] Ibid.
[84] Tullock, "The Prisoner's Dilemma and Mutual Trust," 229.

to trust or not. Her position, at least as quoted by Tullock, is not clearly concerned with the possibility of mutual trust but rather the rationality of trusting at all. Thompson says "If A had good reason to trust B (if, for example they were 'on to the trick' ahead of time an had made an agreement about it) his rational move would be to not confess."[85] His position, also, is not clearly concerned with the possibility of mutual trust. Although, in Thompson's case, he does suggest, in the parenthetical, that there was a kind of mutual agreement between the two actors in the Dilemma. A prior agreement, however, about how to behave in the Dilemma violates the game's conditions and, it turns out, this is part of Thompson's point; "deciding whether it is rational *to trust*, in this case, is not a decision to be reached by a game-theoretic strategy appropriate to the example but a decision to be made on the basis of data independent of the conditions of the 'game'."[86] Thompson's point then is that mutual trust is outside of the context of the Prisoner's Dilemma. In fact, from the quotes presented by Tullock, only the quote from Wolff obviously suggests that trust in the Prisoner's Dilemma is a mutual affair; "each criminal has to be able to trust the other not to pull a double-cross."[87] Tullock's concerns about the mutuality of trust are important, I return to them shortly, but for the moment I will skip over this seeming oversight in order to consider his criticism of Held's position.

On Tullock's account, the Prisoner's Dilemma is not a case in which trust is an issue, at least not in a manner in which a decision is necessary; "one prisoner's opinion about the probable behavior of the other is irrelevant to his own decision, since his payoff

[85] George Thompson quoted in Ibid.: 230.
[86] George Thompson, "Game Theory And "Social Value" States," *Ethics* 75, no. 1 (1964): 37.
[87] Robert Paul Wolff, "Reflections on Game Theory and the Nature of Value," *Ethics* 72, no. 3 (1962): 173.

will always be higher if he confesses."[88] He argues that my decision how to act need not consider what possible course of action you will take because no matter whether you defect or cooperate, I will be better off by defecting: if you defect I score more points by defecting than by cooperating and if you cooperate I score more points by defecting than by cooperating. So, defection seems to be the clear solution and it requires no reference to trust in order to come to that decision. If your behavior, or what I anticipate about that behavior, is irrelevant to my decision-making then the only risk would be the positive risk of a potential for the better of the two outcomes. Further, if your probable behavior is irrelevant to the issue then there is no need for trust at all.

Held, in response to Tullock, wants it to be clear that the Prisoner's Dilemma truly is a risky venture; she asks "what course of action may be deemed to be the rational one when one can *not* know that the other fellow will do"[89] It is, she suggests, the uncertainty of the actions of the other that makes it trust; "if one can make an accurate prediction either that he will or that he will not confess, one can decide in accordance with the usual recommendations for rational behavior."[90] She denies Tullock's contention that accurate prediction is the issue, which, were he correct, would result in the kind of rational response he suggests. Held seems to take the issue too far however; "in this situation it is a pure chance, unenlightened by the possibility of bringing calculations of probable success to bear upon the decision."[91] Pure chance, the possibility that the actions of others are being governed by random processes is surely too much even for trust. The problem that Held raises by responding in this manner is that it seems to make trust an

[88] Tullock, "The Prisoner's Dilemma and Mutual Trust," 229.
[89] Virginia Held, "On the Meaning of Trust," *Ethics* 78, no. 2 (1968): 156.
[90] Ibid.
[91] Ibid.

issue only between strangers; when you make an accurate prediction about the behavior of the other, its simply not going to be trust. Held's claim is that the "question is whether to take a chance on his behavior being co-operative when *no* prediction is possible" seems to take the position too far.[92] Held is presenting a conception that denies the possibility of predicting the behavior of the other. But, by claiming that the behavior of the other can be no better than chance she seems to overlook what is important about trusting: the interpersonal relationship we have with others. Instead she favors something that is much more akin to optimism: moving ahead on the assumption of a positive result while recognizing that the positive result has no better chance of occurring than anything else. But her basic point remains the same; "trust seems to have more to do with situations of *uncertainty* than with situations of *certainty*."[93] So, when she returns us to the point of inquiry we realize that trusting really is the kind of thing that is risky; "*should* we or *should* we not take a chance on furthering a common interest while risking an individual interest?"[94]

Cooperating Can Be Harmful

The Prisoner's Dilemma can now be understood to be risky because of the harms associated with cooperating; by cooperating we open ourselves up to harm of defection. Choosing to defect seems rational since it appears to be the only way to protect ourselves from the harm; it seems to be the most protected from uncertainty. However, this is only an illusion. Clearly, the choice to defect necessarily avoids the worst possible payout and in this sense it is a form of protection but when we consider the cost of that protection it

[92] Ibid.: 157.
[93] Ibid.
[94] Ibid.

should become obvious that it is no protection from loss; the second worst point value will be a loss when compared to either of the payouts of the better category. Further, the choice to defect in no way reduces the uncertainty of our situation; we know nothing more and have become no more certain about either the final outcome or our partner. Defection fails to reduce complexity any more than cooperation, a point that will be taken up in section V of the present chapter.

Consider what a choice to defect gets the actor, it removes the worst possible outcome and it allows for the best possible outcome. A loss has been avoided but not all loss has been avoided. What remains is still the second worst outcome, the case in which both parties defect. While this, on the formulation we are using, still brings with it some points those points are less than could be achieved by simultaneous cooperation. What is sacrificed then is the possibility of agreement; in choosing to defect you must hope that your partner behaves differently than you but the rational choice, so it appears, is to defect. If all agents acted rationally, then it should be that both parties defect. It would seem then that the 'rational' choice is one in which at least one person must act irrationally; this should be a problem for any account which suggests that the rational choice is the choice to defect.

The Prisoner's Dilemma is not risky, however, simply because of the harm but also because it involves uncertainty. It is a necessary condition of the Dilemma that each player cannot know how the other is going to behave. Each can develop predictions or beliefs about the other, they can anticipate whether the other is going to cooperate or defect but there cannot be a condition of epistemic certainty about how the other will behave, which is not to say that the behavior of the other is no more than chance. If I

know how you are going to act then the decision is not going to be the same kind of choice at all. This would not solve the problem though. If I knew that my partner were going to defect I could still choose to cooperate; I might choose to cooperate because I know that others will hear about it and so my future interactions will be made better. Or, I might choose to cooperate because I believe that, regardless of my partners actions, cooperation is the morally right thing to do. Either way, removing the uncertainty of the other's behavior removes what is most important about the Dilemma; it removes uncertainty. So, regardless of whether or not there is harm involved, the Prisoner's Dilemma will contain an uncertain element. In fact, no choice in the Dilemma can entirely avoid the uncertainty, which is why the Prisoner's Dilemma is risky in the technical sense.

Actions

When thinking about trust we can easily conflate an account of entrusting with an account of trusting and yet there is a subtle but important distinction to be made about the difference between the two. It seems, at first glance, that trusting and entrusting should be terms that are inextricably linked; when a thing is entrusted to me, it seems obvious that I am being trusted with regards to that thing, that someone trusts me with it. We view these two concepts as interchangeable because we believe that being trusted is simply to be entrusted with something and that being entrusted with something is an indication of being trusted. However, trusting and entrusting are conceptually different, and we need to be careful not to allow the discussion to drift aimlessly between them. A failure to recognize a move from one to the other could lead to confusion and misunderstanding.

Trusting and entrusting, though related concepts are not coextensive; there are four major differences between them: 1) Entrusting is an action, which trusting is not; 2) one can trust another with something that has not been entrusted to the other; 3) one can trust another with something that could not be entrusted to another; 4) one can entrust something to another without trusting that other with the thing. Russell Hardin contends that trust is not risky because trusting is not entrusting. I explicate his position so that I may respond and show that trusting is indeed risky and yet is also not entrusting.

Trusting Is Not an Action

There is a similarity between trusting and entrusting but the similarity does not entail equivalence between them. The crux of Hardin's position about the separation between these two concepts is that entrusting is an action, whereas trust is not. Entrusting an object or activity to another person is something in which we actively participate; we must be engaged in the process of entrusting. I entrust my car to a valet when I hand over the keys. I entrust my wellbeing to a surgeon when I lay down on the operating table. I can even entrust a choice to you: about dinner, or route, or who to invite. Entrusting may be a matter of signing over my authority but this exemplifies Hardin's point; entrusting is something that we *do*.

Trusting, on the other hand, is not an action. My trust of you, does not consist in my giving the valet the keys to my car, rather my trust of the valet is a particular belief I hold about his commitments such that I can give him the keys to my car. My trust of the surgeon is not that I allow her to operate on me, rather my trust of her is the knowledge that she holds it within her interests to be concerned with my interests which, in this

instance, should result in successfully operating on me. Trust is, as Hardin presents it, not something that we do; trust is something that we know. And, perhaps obviously, we do not use every piece of knowledge that we have; we do not act on all knowledge, which means that even though I have knowledge such that I can trust the valet, I am not required to give the keys to my car to him – I can park it myself.

Trust Without Entrusting

Hardin maintains that "I can trust you to do something that I have not (could not) entrust to you."[95] We can make sense of the first part of Hardin's point fairly easily; it is possible to trust someone and yet not have entrusted them with anything. For some, this point will be obvious; I trust my mother with my car even though I have never had the opportunity to leave my car in my mother's care. But for some, this point will be a challenge; if the act of entrusting, engaging in trusting behavior, simply is trusting then without the specific action there is no trust. One might argue that claiming to trust another without ever entrusting anything to that other is nuts, that "without action, the claim of belief is empty."[96] I follow Hardin's lead in calling this account of trusting behavioral. He argues that relying on the action of trusting is problematic because it really would mean that we fail to trust our friends and loved ones up until we actually put something in their control; he points out that "I could trust you very much without having occasion to act on my trust."[97]

---

[95] Hardin, *Trust and Trustworthiness*, 65.
[96] I'm not sure how to cite this. It is a quote from a comment/criticism of Hugh's, which, since it was said directly to me, I am not sure that he wants directly attributed to him.
[97] Hardin, *Trust and Trustworthiness*, 58.

Hardin's reply to the behavioral criticism is to consider trusting as a form of knowledge. By considering trust to be something we know, trusting becomes possible even in the absence of the act of entrusting. For Hardin, the important piece of knowledge is "because you're your interests encapsulate mine, which is to say that you have an interest in fulfilling my trust."[98] Hardin says that his position is grounded in an economic epistemology: "The economic theory of belief and knowledge focuses on the individual believer, on the costs and benefits to the individual in coming to have various beliefs, not on the matter of belief."[99] His account therefore is about the costs and benefits of coming to hold the belief that another has reason to consider what is in my interests; I can trust the valet to drive my car even if I have never and will never let the valet drive my car because I can believe that the valet would treat my car as something which is important to me.

Entrusting Requires Right

Someone who holds a behavioral conception of trust might not yet be convinced though; actions might still be important to our understanding of trust. A behavioral approach seems to demand that only present and past actions may be considered because it focuses on what has actually been entrusted. But for events that have not yet happened a behavioralist can imagine that in the future the object or thing will be entrusted the particular person. So the act of entrusting would still be important for understanding what it means to trust because trust will exist only on the condition that I would be willing, and perhaps able, to entrust the thing at some future time. Hardin's second point, that I can

[98] Ibid., 3.
[99] Ibid., 115.

trust you to do something that I could not entrust to you, addresses this response. Hardin's position can be parsed in two ways. First, I can trust you in regards to something which would be impossible for you specifically to be entrusted with; I could trust my friend Bailey to give me his kidney even though we do not have matching blood types because I can know that Bailey holds my interests in his interests. This first point may not be so convincing though; it is analogous to our initial response involving the loaning of a car and it suffers from the same behavioral criticism. Second, I can trust you to do something which would be impossible for me specifically to entrust to you; I can trust my friend Chase to save a drowning child even though that child is not one that is in my care and so not one that I have the authority to give into the care of another. I cannot entrust the child to Chase's care because entrusting requires giving over control, which, in this case, is not mine to give. But it might be in my interest to for the child to survive (perhaps it is my sister's daughter) in which case Chase could hold my interest within her interest. In this case the behavioral account finds a steeper hill to climb. As a friend, I might have a number of expectations on how Chase will behave. Knowing her, I can expect that she will want to save a drowning child. I can even expect that she will be capable of saving the drowning child. This would seem to be enough to allow me to entrust the saving of the child to her. But, since the child in question is not mine I do not have the authority to place that child in Chase's care. In order to entrust I must be in the position of ceding control; the act of entrusting requires me to hand over to another the power to act, it must be an action, but I cannot act in this case because *I* do not have authority or control over that child. I could not trust Chase to save the child on the behavioral account. Chase might predictably save all drowning children, this might even

be something that I find endearing about her, but if trusting it to be concerned for my interests than mere predictability will not account for trusting.

Distrusting the Entrusted

Further, I can entrust something to you and yet not trust you in regards to that thing. For example, I might not believe that Daphne will treat my car with the kind of care and concern I would wish but I have good insurance and I know where she lives, so I am willing to let her drive my car. I will entrust her with my car even though I have no real trust in her. This point, I take it, the behavioral account has no real defense against. If actions are the mark of trust then even con-artists are worthy of trust because, right up until they betray us, they have done the actions of trust. It is because of the difference between what we know and how we act that trusting and entrusting are not and should not be considered equivalent. Hardin points out that many believe that "trust is a significant issue only when there is doubt in the trusted's likely performance."[100] This makes not only trusting to be a case of action but it makes trustworthiness action oriented as well. Virginia Held's response to Gordon Tullock seems to have exactly this kind of structure. Held's idea is that the less you *know* about how someone will behave the more you must trust them to behave that way. Trust, under this conception, requires ignorance, which obviously runs counter to Hardin's conception of trust as a form of knowledge. Held's position seems to capture much of what we experience in our everyday experience of trust; we feel that we are trusting when we put something into the control of another and yet cannot guarantee how that other is going to respond. However, Hardin finds this conception implausible. He points out that on this view we can never trust close

[100] Ibid., 73.

associates as much as we can trust complete strangers because the more we know someone the more certain we are of how they will behave. If this criticism is correct then surely there must be something wrong with the position that trust is a matter of the expectations which we have formed. For example, it would mean that no one ever trusted her or his parents. Hardin suggests that what we find appealing about this position is that an individual may act without sufficient justification; I can jump off a roof into your arms even though I have just met you. What makes this "scant expectations" view so attractive is a concern for our actions. Actions are not necessarily evidence of trust, however, so expectations are not necessarily trust.

Mistakes to Avoid

Another mistake that any discussion of trust needs to be careful to avoid is confusing a justification of promising for a justification of trusting. This confusion is particularly common because promising seems to be a perfect analogy for trust. We often think that any case in which one is expected to keep one's promise is a case in which the individual is being trusted, and we often think of someone who is trustworthy as someone who keeps her or his promises. Linking our conceptual understanding of promising and promising-keeping to our conceptual understanding of being trusting and being trustworthy is not so far-fetched. They do indeed share much in common and yet a justification of promise-keeping will not suffice for a justification of trusting. In the first place, the analogy would place the justification of trusting squarely with a justification of promise-keeping. In the second place, the moral burden is on promise-keeping and being trustworthy not on making promises and trusting. In the third place, to moralize accepting

promises should strike us as strange. If these objections are on target then using promising as an analogy for trusting is a mistake. This becomes especially important for our investigation because it calls positions already discussed into question. For instance, Schwarz' concern for the trustworthiness of communists can be seen as a concern for whether they would keep their promises. Clarifying the trust/promise relationship will therefore help our understanding of trust.

The relationships of trust and promise do have a structural similarity; for trust there is one who trusts and one who keeps that trust, for promise there is one who promises and one who keeps that promise.[101] However, even if we assume that the analogy between trust and promise is justified we still face the same set of problems that plague trustworthiness in relation to trusting. The moral burden in promise relationships rests with the promise-keeper so, in the analogy, the moral burden in trust relationships would rest with the trust-keeper. But this makes the justification for trusting rest with the trust-keeper, the person who is trustworthy, and we have already seen that the justification for trusting cannot be the same as the justification for being trustworthy, so any similarity between being trustworthy and keeping promises is moot. That is, a justification of trustworthiness does not translate into a justification of trusting which means that the analogy between keeping promises and being trustworthy cannot succeed as a justification of trusting. So, even though on its surface any discussion of promises will be similar to a discussion of trust the similarity is really between keeping promises and keeping trusts. Hardin makes this point by discussing the goods of promise and trust. According to Hardin the good of a promise, what makes it good, is that people tell the

[101] We might consider these to be mirror images of each other, images where the direction of flow of the obligation travels in opposite directions; from truster to trustee and from promisee to promiser.

truth but the good of trusting is "that the *trusted* is typically trustworthy."[102] For example, in the case that A promises B, it is A that carries the obligation. If the analogy holds true then when A trusts B, it should be A that carries the obligation but this is clearly not the case; the trusting person incurs no obligation from trusting. Instead it is B who carries the obligation in the case of trust because it is B who is to be the trust-keeper. The obligation is on the one who promises but not on the one who trusts. If we wish to use the analogy of promising to explain trusting then it must be between believing someone to be a promise-keeper and believing someone to be a trust-keeper. The analogy is with believing the promise of the promiser and not with the action of the promiser actually keeping her promise.

We can see the centrality of belief to the analogy by looking at the obligations associated with promise and trust. The obligation of promise attaches to the one who must keep the promise and the obligation of trust attaches to the one who must keep the trust but the one who must keep the promise is the one who has made the promise, the one who promises, while the one who must keep the trust is *not* the one who trusts. Likewise, the one who is promised is the one to whom the promise is owed but the one who is trusted is *not* the one to whom the trust is owed. The moral burden, though it rests with same person in the relationship, the one who must keep the promise or trust, is not acquired in the same manner. Hardin points out that any imperative to keep a promise is based, first, on the prior act of making the promise: "the potential promise keeper…is first the promiser."[103] Individuals who make a promise are, in essence, claiming to be those who keep their promises, they impose an obligation on themselves by voluntarily

[102] Hardin, *Trust and Trustworthiness*, 62.
[103] Ibid., 63.

and knowingly an explicit obligation. This first act of promising is what separates being a promise-keeper from being trustworthy. If the promiser were to fail to keep her promise she would either have lied or would have chosen to abandon the promise that she made. No such obligating first act is present in being trustworthy however. The analogue then, if there is one, between promise and trust would have to include a prior claim of this type, a prior claim to being trustworthy. But Hardin believes that it is hard to conceive of a case of "gratuitously announcing one's trustworthiness in a way that matters apart from promising."[104] Any case in which I vowed to be trustworthy would *de facto* be a case in which I promised because to announce that you can be trusted is to make a promise to do the thing about which you have claimed to be trustworthy. To tell someone that you trust them is to do something wholly different; it is to impose, or attempt to impose, an obligation on them.

Because the example of the morality of promising forces us to be concerned with the morality of trusting and not with the morality of being trustworthy we need to find an analogy that shares all aspects of the relationship. Believing the promise of another most closely resembles the relationship of trust so, the appropriate analogy for trusting is with believing the promise of another. But even this analogy is incomplete because for the analogy to work as a justification for trust there would have to be a discussion about the importance of accepting promises. To moralize the accepting of promises, however, is strange. The justification of accepting a promise seems to be a purely practical matter; we accept a promise when we have good reason to accept the promise. To claim that someone has a moral obligation to accept a promise is to say that there is a reason for

[104] Ibid.

accepting the promise that has nothing to do with the likelihood of that promise being fulfilled; it is to claim that there is a reason for accepting a promise even when there is good evidence suggesting that the promise will *not* be fulfilled. But a moral obligation of this type would be contrary to the usefulness of promising because it holds that a promise is still binding even when that promise is broken. Without a moral obligation to either accept promises or to trust, the justification for keeping promises or being trustworthy is not sufficient to convince me to accept promises or to trust.

Even though what is moral about keeping promises seems at first glance to be the same as what is moral about keeping trusts, it is in respect to our moral requirement to accept the promises of others or to believe that others are worthy of trust that the analogy breaks down. This position highlights the affect of Hardin's claims about trust being a form of knowledge. To accept the promises of others is an activity, it is something we engage in, but to believe others are worthy of trust is to have a piece of knowledge about them, a piece of knowledge that may or may not allow you to engage in behaviors like accepting promises.

Hardin criticizes most discussions that treat trust as an expectation by pointing out that an expectation involves either knowledge or belief on the part of the truster. Most discussions treat trusting "as a matter of behavior,"[105] however, so it would be a mistake for those accounts to treat trusting as an expectation. We form expectations about how others will behave or act but Hardin argues that trusting itself is never an action. Trust is a set of beliefs that we hold not some particular or even general activity that we engage in. If we treat trust as an activity or a behavior then we should not also consider it an

[105] Ibid., 56.

expectation. Treating trust as an action is putting the cart before the horse, so to speak, since we often try to explain actions by reference to trust or expectations. Hardin asks "[i]f trust is an action, then what are we trying to explain in thousands of pages on the topic."[106] Describing one action, turning over my car keys, by reference to another action, trusting another with my car keys, simply pushes the looked-for explanation farther away. An explanation of action E that points to prior action D only suffices to explain the consequences of action D not to justify action E. 'I trusted my accountant' explains why you did not audit your books or 'I expected my ex to cheat on me' explains why your hired a private detective. Promising is not, therefore, useful for explaining trust.

Complexity

If promising is not an effective analogy for explaining trust than another way of understanding it must be found. This understanding must be able to account for riskiness inherent in trust. The discussion of science presented by Bentley Glass in Chapter 2 reminds us that the world is an incredibly intricate place; it is filled to brimming with free agents, each of whom has different motivations, ideals, wants, and desires. Their actions, though perhaps predictable, are far from transparent to us. Glass' fear and the resulting need for a steadfast rule of trustworthiness is an attempt to deal with this complexity. Niklas Luhmann not only recognizes this aspect of living, he fully depends on it. He presents an account which holds that trust is an attempt to deal with the complexity of the world; left without this capacity we would be frozen, unable to respond to a near infinite set of choices and relevant, possibly mitigating, factors. Trust allows us to reduce this set

[106] Ibid., 58.

to a manageable size. Luhmann's discussion is one of the earliest accounts that focuses strictly on trust and its importance to individuals, institutions, and society.

Luhmann's concern in discussing trust is primarily sociological, yet his account is extremely useful for any philosophical investigation of the issue because he spends a considerable amount of time discussing the conceptual understanding of trust. He is concerned with how trust functions in society and the ways we create and maintain trust relationships. Luhmann begins by addressing the term in its most widespread use; "[t]rust, in the broadest sense of confidence in one's expectations is a basic fact of social life. In many situations, of course, man can choose in certain respects whether or not to bestow trust."[107] These seemingly benign claims are incredibly complex and will require substantial unpacking. For Luhmann, trust is a form of confidence and involves expectations, it is a basic fact of social life and therefore pervasive, but it is also necessary for social life; a lack of trust would prevent any kind of structured social order, without trust one could not even distrust. Trust is on this account, at some level, a choice. The key to understanding his position is that trusting is one way, possibly the best, to deal with the complexity of the world. We exist in a world beyond our ability to comprehend and beyond our ability to deal with. We are surrounded by individuals who, like ourselves, make innumerous choices and decisions throughout their lives. The sheer number of choices available to even a single person in the course of a single day is more than any person can conceivably consider as part of a decision-making strategy. In addition to the freedom of individuals, our world is filled an inconceivable number of causally determined objects that each have a near infinite number of possible interactions

[107] Niklas Luhmann, *Trust and Power : Two Works* (New York: J. Wiley, 1979), 4.

with other causally determined objects. Understanding and predicting the movements of these objects are more than any human could hope to achieve. This world is far too complex for us to understand by force of brute calculation so we must reduce this complexity of that world to a manageable scale. Trusting others allows us to reduce that complexity.

A Reduction of Complexity

Luhmann argues that "[the] world is overwhelmingly complex for every kind of real system."[108] In our college physics classes we learn how to predict the actions of objects in space but we make a number of assumptions in order to make accurate predictions: we ignore many features of the real world. Friction is taken to be constant across the surface, surfaces are themselves taken to be uniform, perturbations of air pressure and gravity are ignored; we idealize the conditions under which the interactions of the system in question are taking place. This idealization is necessary because any real system is far to complex to allow for prediction but even an idealized system can have overwhelming complexity: Edward Lorenz found a level of complexity in a simple weather simulation program that showed that small changes in initial conditions could produce drastic differences in final outputs.[109] Idealized systems rapidly become too complex to predict, real systems are overwhelming complex, and any system that involves the freedoms of human beings is intractably complex; this level of complexity makes the range of possible responses within any system far surpass even the system's ability to comprehend let alone individual members of the system. We must reduce the

[108] Ibid., 6.
[109] James Gleick, *Chaos : Making a New Science* (New York, N.Y., U.S.A.: Viking, 1987).

complexity of the system to a manageable level in order to interact fruitfully and Luhmann thinks that trust is one of the ways in which we can do that. His account of trust begins with familiarity. Because we are "conscious of the world's complexity" we recognize, as a basic fact of our existence, that some things are beyond our understanding.[110] Familiarity, which is to have a kind of knowledge or experience, with the world gives us a backdrop for understanding and allows us to avoid wasting time and resources. Our familiarity with the world allows us to ignore certain possibilities that are unlikely or uncommon. Acting as if our future will conform to our past experience limits the extent of the world's complexity. But familiarity is neutral with respect to predictions; my familiarity with you could just as easily give me a reason to trust you as to distrust you. In order to interact with people we must be able to take some things for granted, which familiarity allows us to do, but trusting further limits the future possibilities to a manageable set. "Familiarity is a precondition for trust" and is necessary for human social interaction.[111] Trust, on Luhmann's account, is not merely a pervasive fact about social life it is a necessary condition for having a certain kind social life; trust "presupposes a situation of risk."[112] Trust then, on Luhmann's account necessarily involves uncertainty; it is uncertainty that creates the possibility of trust. Without uncertainty, Luhmann argues that we would not have trust. Trust is risky.

Luhmann's position is not without objectors however. Parales-Quenza argues that "the reduction of information undermines trust."[113] If this is correct then Luhmann's

---

[110] Luhmann, *Trust and Power : Two Works*, 6.
[111] Ibid., 19.
[112] Diego Gambetta, ed., *Trust : Making and Breaking Cooperative Relations* (New York, NY, USA: B. Blackwell, 1988), 97.
[113] Parales-Quenza, "Astuteness, Trust, and Social Intelligence," 43.

position, rather than giving us an explanation of the trust concept, actually obscures the issue. Parales-Quenza's claim is that by reducing the information we have about a situation we make it harder to trust the individuals involved in that situation; he wants us to have more data points about the behavior of others. Luhmann, on the other hand, claims that the only way we can understand the behaviors of others is to trust them, which he takes to mean skipping over the need for collecting an ever increasing data set; we reduce complexity by generalizing not specializing. Trusting so is how we deal with the liberty and complexity of other people.

A Basic Fact of Social Life

One of the justifications of the investigation of trust as a concept is that trust just *is* a basic fact about social life. We find this claim repeated in almost all discussions of trust and it is with good reason; we find ourselves, consciously and unconsciously, engaging in trust relationships every day. The conception of trust may dictate the scope of those relationships but no account is going develop a conception that takes of trust to be rare. Luhmann's position on this, however, is rare. Although he conceives of trust as a basic fact about social life it is how he conceives of it that separates him from an authors like Hardin and Baier. His conception is not justified merely by the pervasiveness of trusting actions in the world. Instead, he defends trust's central position as integral to the functioning of social life; we hear the refrain that the "world is overwhelmingly complex for every kind of real system."[114] It is a central problem for any theory of human interaction that the theory must account for how an individual copes with the autonomy of all the others. Luhmann is showing us that the range of possible responses that any

[114] Luhmann, *Trust and Power : Two Works*, 6.

system might present far surpass even the system's ability to comprehend. "Human beings," Luhmann tell us, "are conscious of the world's complexity."[115] It is the case then that human beings, in all their interactions, recognize the extent to which those actions are beyond their control and understanding. We know, as a basic fact of our existence with others, that we must trust those around us.

It is not yet clear, however, from what has been said about Luhmann how this complexity implies the trust that is a basic fact of our interactions. This clarification will have to wait though as the ground has not yet been fully prepared; there are other things that need to be covered yet. That trusting is basic fact about human interaction does not itself say much about trust. Breathing is a basic fact about human interaction as well and, if we accept claims like those of Sissela Bok, breathing should be a fair analogy to trusting.[116] Hume, on the other hand, reminds us that because something *is* does not imply that it *ought* to be; that trusting is a basic fact about human social interaction does not tell us that it should be necessary to that interaction or, even that we ought to trust. Trust might have a purely accidental relationship to human social interaction.

Necessary for Interaction

Luhmann does maintain though that trust is *necessary* for human social interaction not merely that it is pervasive attribute of the interaction. He emphasizes this point when he describes the relationship between trust and familiarity; "[f]amiliarity is a precondition for trust."[117] In order to trust another we must be familiar with her, the way she acts, and the way she responds. We cannot trust the unfamiliar because we do not

[115] Ibid.
[116] Bok, *Lying : Moral Choice in Public and Private Life*.
[117] Luhmann, *Trust and Power : Two Works*, 19.

have enough experience with it. Luhmann's discussion of this relationship centers on the historicity of familiarity, "the past prevails on the present and the future."[118] We are familiar with a thing when we have had experience with it. Our experience is a product of the past and Luhmann points out that the past is not fraught with uncertainty. The past is a static thing which, though we might come to have different understandings of it at different times, is not subject to the vagaries in future events; it has had its "complexity reduced at the outset."[119]

The familiar world gives us a backdrop from which to make predictions about the future. It is our experience with a thing in the past that gives rise to what we think it will do in the future. Human social interaction that had to treat each and every interaction as if there were no past, that could not use previous experience to predict future experience, would be so inefficient as to not function. It is a necessary precondition then of human social interaction that we have familiarity, at least with humans, but more so with those which we are to interact in meaningful ways. While this point may seem to be overly demanding of requirements of interaction it is hard to conceive of anyone in the position of interaction in the absence of exactly the kind of familiarity that is being referred to.

Familiarity is neutral with respect to the expectations we will form, however. My familiarity with you could just as easily give me a reason to hand over control of that which is important to me as it could give me exactly the reasons I need to refuse you even the slightest control over those things. Familiarity breeding discontent may not be more obviously true than its opposite but it is equally recognizable. Trusting, on the other hand, is to have formed certain positive expectations about how the other is going to act.

[118] Ibid.
[119] Ibid., 20.

Familiarity allows us to narrow the possible responses of other agents to size that such that we can interact in a world which is overwhelming but trust limits those possible responses even further. That is, trust allows us to act "as though there were only certain possibilities in the future."[120] And it is this reduction in the possibilities of future actions that allows us to interact with other human beings in social environments. Trust is not merely basic fact about social life it is a necessary condition for having a social life.

Familiarity and Trust

We manage our familiarity and trust of others in a way such that we can cope with the world we are presented with. This amounts to, under Luhmann's account, choosing to trust others or not. This decision, in turn, affects the way we act. But the 'choice' as described by Luhmann is not strictly a choice to trust or not. It is not coming to a fork in the road with 'trusting' on the right and 'not trusting' on the left. Luhmann's 'choice' is found instead in how we direct our attention. The position here is that we direct our focus on different aspects of what we know about the other, what things we have become familiar with, and that focus leads us to either trust or not in relation to that person. So, when I am confronted with problem of trusting or distrusting someone with whom I have been friends for some time my 'decision' to do follows from which parts of our shared history I focus my attention. I give priority to some aspect of our relationship over others. It is likely that my friend has disappointed me in some way in the past and it is just as likely that my friend has fulfilled my expectations as well; trusting or not trusting is privileging one of those experiences over the other. Luhmann believes that we can control that focus. We can direct our attention in a number of ways. In this case it could

[120] Ibid.

be either that I know my friend is remorseful for having let me down or because I know that my friend's situation at that point was different than it is now. But the reason it is focused is not as important to our immediate concern as the fact that it *can* be focused. I take some of those past experiences as outliers and some as a reliable indicators of behavior; "[t]he actor binds his future-in-the-present to his present-in-the-future" by choosing which aspects of our relationship will dominate.[121] It is a matter of what we choose to emphasize that allows us to direct our trust.

A Positive Expectation

Because Luhmann's position on trust involves a forward looking perspective of behavior, because it takes past experiences as a kind of evidence for future behavior, his position is that trust is a kind of expectation but not of the kind that Hardin objects to. Luhmann claims that "[y]ou cannot trust chaos;" complete uncertainty can never be subject to our expectations.[122] When we have confidence in another our familiarity has allowed us to be confident that the other will perform some action. That is, we have formed an expectation about how the other will act. When we trust another we have a positive expectation about how that other will behave. But we need to be careful not to mistake our expectation about future behavior as a prediction about the behavior. Prediction, Luhmann will tell us, is something "the correctness of which would be measured when the predicted event occurs."[123] If trust were a prediction it would have to be applied to some particular end before we would be able to measure its correctness. "I was right for trusting you with the car" would mean that my trust of you was directed at

[121] Ibid.
[122] Ibid., 39.
[123] Ibid., 88.

the fulfillment of the car being returned intact. But, trust is not understood in that manner by Luhmann because that would make trust the object which it is meant to explain; trusting explains how it is that we lend a car to another not our expectation in lending the car to another – trusting *for* the return of the car is different than trusting *that* the car will return. For Luhmann, rational choice and prediction can replace trust but are not examples of it; "[t]rust is…something other than a reasonable assumption on which to decide correctly."[124]

Confidence Is Not the Same as Trust

Having confidence in the way another will respond to the complexity of the world might appear to be trusting in that individual but Luhmann makes a distinction between the concepts of confidence and trust. Confidence is an expectation about future events about which you lack control but it is not yet to trust. Having confidence is to not have considered the alternatives. Trust, on the other hand, "presupposes a situation of risk."[125] That is, trusting requires that you be aware that there is an alternative: not trusting. We have trust in situations in which we have uncertainty about the outcome of events and still make a decision. Outcomes can be uncertain in situations of confidence as well but we fail to make a decision about those situations, we merely act.

Being disappointed when you have had confidence in another is, according to Luhmann, an "external attribution" while disappointment in trust has an "internal attribution" and regret. [126] In situations where the damage received by the uncertain

[124] Ibid.
[125] Niklas Luhmann, "Familiarity, Confidence, Trust," in *Trust : Making and Breaking Cooperative Relations*, ed. Diego Gambetta (New York: B. Blackwell, 1988), 97.
[126] Luhmann, *Trust and Power : Two Works*, 79.

outcome would be no greater than the positive you sought, Luhmann says, that would not be trust but rather rational calculation. You must have more to lose than you have to gain in order to trust; when you have more to gain than you have to lose in a given situation a rational calculation of the scenario dictates that you should proceed. Trust requires the truster to be open to a great deal of harm, greater than the benefit they receive from trusting. This distinction between the potential harms and the potential gains is what separates trusting another to care for a thing from having confidence that another will behave in a way that is beneficial to you.

More to Trust?

Yet, it seems that there is still more to trust that merely risking more than you have to gain; "[t]rust is only involved when the trusting expectation makes a difference to a decision; otherwise what we have is simple hope."[127] It is merely hope where the choice is unaffected; I hope that you will arrive to work on time but as nothing important to me hangs on your arrival to work I have no need to trust that you will. Trusting therefore requires the ability to choose even if that choice is not specifically about the trust itself. For something to be an instance of trusting I must have had the option to not engage in that activity; my acceptance or denial of you as an agent worthy of trusting had to hang in the balance. This means that cases of reliance could never be cases of trust because when I rely on you it is because I must do so. For Luhmann, reliance means that I must entrust something to another. If I can avoid entrusting something to that person then I am *not* relying on that person. And, I will choose not to entrust something to another if I am not trusting of that person. Trusting another has an impact on the decisions we make.

[127] Ibid., 24.

It would seem odd then to claim that trusting another requires the trusting expectation to affect the choice made while simultaneously claiming that trust itself is choice free. This is reminiscent of a position held by Hugh LaFollette; he claims that trusting requires an action of some kind in order to be considered trust.[128] This explains how Luhmann can and would hold the position that trusting is a choice that we make; we choose where to focus our attention and we choose whether or not to act.

Overwhelming Complexity Produces Trust

Combining all of what has been said so far we begin to understand Luhmann's position that trust is a reduction of complexity. The world is an incredibly complex place filled not only with more things than can be kept track of by a single individual but also with other individual agents whose actions cannot all be reliably predicted. This level of complexity is far too much to allow for successful social interaction though and so it must be reduced. By our familiarity with others through past experience we can begin to form expectations about her or his future behavior. We can act as if our future will conform to our present, which allows us to limit the extent of the world's complexity which allows us to function in an efficient manner without wasting valuable time and resources on a problem that could never be resolved otherwise. Trust is necessary for human social interaction.

The Benefit of Trust

Having recognized the worth and value of trust Luhmann attempts to show us a way that trust can be to our benefit; "[t]rust accumulates as a kind of capital which opens

[128] Hugh LaFollette, *Personal Relationships : Love, Identity, and Morality* (Oxford, UK ; Cambridge, Mass.: Blackwell, 1996).

up more opportunities for more extensive action but which must be continually used and tended."[129] To be honest it unnerves me to use capitalistic examples as our paradigm of human interaction. Capitalism is a product of human interaction, to now use it to describe or explain necessary precondition for human interaction is strange. At the very least doing so could be misleading; it implies that trust is a commodity. And although we may be inclined to treat trust in that way, trust can never be a commodity. As Luhmann points out trust cannot be traded, exchanged or stored. Further, the whole concept of commodities exchange is founded on trust, prediction, and expectation. Trust cannot be an object of itself. And so, even though Luhmann uses a metaphor of capital in describing trust he does not intend that metaphor to be taken as a strict statement of the nature of trust. It is merely a *kind* of capital, something that may open new avenues of interaction but cannot be exchanged for any other type of good.

Risky

Besides uncertainty there is another common element of conceptions of trust: vulnerability. Vulnerability is related to risk but is not identical to it. It is the harm that we open ourselves up to when we trust. I have tried to emphasize that it is uncertainty, as opposed to harm, that makes trust risky. Rajeev Bhattacharya points out that "[f]or most of these theorists, vulnerability is a key element of trust."[130] This position is one that I support but Bhattacharya disagrees and suggests that "[m]uch of the discussion of trust and vulnerability has been, in our opinion, something of a theoretical red herring

[129] Luhmann, *Trust and Power : Two Works*, 64.
[130] Bhattacharya, Devinney, and Pillutla, "A Formal Model of Trust Based on Outcomes," 460.

confusing vulnerability with uncertainty or risk."[131] There is a sense in which I could simply ignore Bhattacharya's criticism; having shown that trust is risky, Bhattacharya and I are in agreement. But, it is important for the position that will be developed in the following chapters to see that that risk and uncertainty are not, by themselves enough to count as trust. Something else must be present and the beginnings of that justification can be found in trust that is vulnerable. This might seem to make the vulnerability account a separate issue, something to be discussed in Chapter 4 but Bhattacharya's concern, that vulnerability is confused with risk and uncertainty makes salient a discussion of the issue at this point.

Annette Baier is perhaps the most recognizable of those who suggest that trust makes us vulnerable. Her full account will be considered in greater detail in Chapter 4. Now, however, I focus specifically on the vulnerability inherent in trust. She suggests that trust brings a specific kind of vulnerability, which is a "vulnerability to not yet noticed harm"[132] This vulnerability comes from the discretionary power of the individual who is trusted; "to trust is to give discretionary powers to the trusted, to let the trusted decide how, on a given matter, one's welfare is best advanced."[133] Trusting is not simply to have there be some uncertainty about how the other will behave it is to give control over to that individual in such a way that the other has the capacity and ability to make determinations regarding the disposition of the trusted thing. This is contrasted with providing strict and unalterable rules for behavior, rules that can be followed mechanistically.

---

[131] Ibid.: 461.
[132] Baier, *Moral Prejudices : Essays on Ethics*, 104.
[133] Ibid., 136.

Baier's position accepts the complexity of the world suggested by Luhmann but focuses on the liberty of others rather than the broader uncertainty of all things. The harm that we are in danger of receiving when we trust is caused, in part, by the autonomy of other individuals. In giving over discretionary powers we allow others to make decisions, decisions which could or could not go in our favor. When I trust another I believe that he will use his discretion in a way that furthers my interests because they are *my* interests and not simply because our interests coincide. Because the other is genuinely concerned about my well-being he will *want* to pursue things that are in my interests, sometimes even when those interests conflict with his own. Already we can see that Luhmann and Baier agree on at least some of the aspects of trust, and yet, there are other aspects about which they disagree. For instance, Baier and Luhmann agree that trust is pervasive but disagree that it is a choice. They also agree that trusting requires that the trustor have more to lose than to gain; for Luhmann this means literally that the loss will be greater while for Baier it means that our vulnerability is such that we could mitigate it by retaining control for ourselves. It is the uncertainty of the choices made by others in situations which we can neither predict nor anticipate that requires us to trust them.

Synthesis

We come now to the conclusion of this chapter, the point at which I must synthesize the ideas that trust is uncertainty, vulnerable, and yet not modeled on entrusting. It will be helpful here to step ahead to the subject of Chapter 6, the account presented by Robert Solomon and Fernando Flores. They claim that "distrust infiltrates every marriage insofar as one is aware of (and grateful for) the contingencies of and

aware that love can never be guaranteed."[134] Couched in a discussion of marriage, this is an important point about the uncertainty of trust, an important point about the nature of trust. Because trust necessarily involves uncertainty it cannot be the case that recognition of that uncertainty dissolves trust; accepting that trust involves uncertainty *means* that the uncertainty of the outcome or the motivations of the other cannot count as a strike against the possibility of trusting. It can seem to us that once a person begins to take note of the liberty of the other then it would be impossible for us to trust that other because, by focusing on how your trusting may be failed, the focus is on the possibility of a breach of trust. But, this denial of trust is evidence of a specific conception of the trust relationship, a conception that takes trust to be about the certainty of prediction; the more certain I am about how the other will behave (or what motivates them) then more that I can trust him. Trust however cannot function on this model, not really. It certainly cannot function on this model if the argument I have presented thus far has been even slightly convincing.

Clearly, if I believe that your motivations are always self-directed then I cannot trust you. Knowing that you will, when it benefits you, abandon me leads away from, not towards, trust. The certainty of knowing how you will respond, in this case, actually leads to less rather than more trust. And, it seems that the more, or the stronger, that I believe that your interests are for me and my interest than the more certain I am about your behavior as well. We might think that this, at least, would be a case where we could truly trust another; surely, one would argue, if I am certain that you will act in my interests then I can trust you. But, wherever certainty resides trust cannot. A mechanical, unthinking response on your part would prove to be useful for prediction and can

[134] Robert C. Solomon and Fernando Flores, *Building Trust in Business, Politics, Relationships, and Life* (New York: Oxford University Press, 2001), 30.

certainly stand as a reason for my acting but this level of certainty actually negates any need for trust. Certainty and trust are not opposed dichotomously; they are not at war, instead the two fade into each other while remaining mutually exclusive. Sadly, opposition to this claim will invoke an argument similar to debates about volunteerism, which I have no intention of engaging. So, I will try another tack.

If I could travel into the future and ***know*** how you would respond in a case where my interests were at stake then I would not need to trust you in order to allow you control of my interest; I would have certain knowledge about the disposition of my interests. Admittedly, I do not need to trust you in order to allow you to control my interests (to entrust you) but the certain knowledge of your behavior denies the possibility of trust.

If I ***know*** you will not betray me even when asked to by another then I still cannot trust you because there would be no uncertainty. But, it seems foolish to deny that, independently of my knowledge about how you would act that I could trust you at particular moments, even that particular moment; my knowledge of how you will respond cannot be the motivation for my trust in you at that moment but I might still trust you simultaneous to having knowledge about how you will behave. That is, the certainty of how you will respond, either in or against my interests, stops me from trusting you in regards to that particular action at that particular moment but it does not necessarily impact my general thoughts and beliefs about you. Because entrusting and trusting are different things, because they can be motivated by wholly different factors they can be occurring simultaneously. Trusting and entrusting, while not identical, are also not mutually exclusive. I can entrust something to you without trusting you but I can also entrust something to you while I trust you. This simple point can be quite profound. The

certainty which abolishes trust only effects a particular conception of trust that relies on entrusting as its model.

In fact, this distinction helps explain the familiarity problem. Many who write about trust point out that trust requires some knowledge of the other while it, at the same time, denies the possibility of perfect knowledge. This seems to suggest that the better you know someone, the less that you (can) trust her. By pointing out that trust is a general attitude you have about the motivations of the other, you can have it both ways; require familiarity and still deny certainty. Certainty would abolish a three-part trust relation but on my account trust is not about the single interaction in question and so certainty does not affect my account.

The problem presented by certainty only relates to the certainty of that particular event coming to be. Familiarity allows us to become incredibly astute at predicting the behavior of others. Normally, that you threaten trust; it would not threaten my ability to trust you but it would threaten any account of trust that recognizes that trust requires uncertainty. It threatens trust because the more you know the less one can trust. But why must trust involve uncertainty? Mostly, the account is phenomenological. Trusting is taking a chance, opening oneself up, engaging in an activity that could cause harm.

When we think about crossing a bridge, I imagine one that spans a gorge over a particularly treacherous river, it makes some difference to us whether or not we are certain that the bridge can sustain our weight. If we have tested the bridge's strength and have found that it could carry 6 times our body weight, we have little trouble in crossing the bridge (unless we are afraid of heights or unless the bridge is very narrow and lacks guard rails). If we could have greater epistemic access we might even be able to *know*

that we would be successful at crossing the bridge. In this case it would be inappropriate to say that we trusted the bridge. Further, we would not experience anything like trust. We might have a great deal of confidence in the bridge and we might have developed all manner of expectations, not only about our crossing but about what we will do after we have crossed, but we will not be trusting the bridge to maintain is structural capacity. It is only when we are uncertain whether the bridge will hold our weight that we begin to trust the bridge. In this case we find that we have little confidence and not much in the way of an expectation. The lack of certainty in the bridge leads to trust. Here, like in many other cases, actually crossing the bridge need not be motivated by trusting, you might be running from corrupt local police like Kathleen Turner in *Romancing the Stone*, or you might be ignorant of the danger of crossing such a bridge, like the folks driving over the San Francisco Bay Bridge during the 1999 Earthquake.

Trust is a matter of believing that the other is trustworthy. That belief is necessarily born of the uncertainty in the world; it is not a bit of knowledge that submits to factual checking, rather it is a bit of knowledge which is subject to revision and revaluation. Trust is risky; the harm of misplacing trust is not simply a matter of possibly being injured but is also, at the same time, a matter of causing us to question our judgments. We saw in Chapter 2 that even when no breach of trust has occurred, the revelation that another was not worthy of our trust can be terribly damaging. Having established that trust is risky, it is time to move on to show which motivations will suffice for forming a belief that the other is trustworthy; not all motivations, as we have seen with the strategic motivations in the Prisoner's Dilemma, are sufficient to give me reason to trust. Discussion of the appropriate motivations is the topic of the next chapter.

Chapter Four: Trust as Good Will

In this chapter I argue, following Annette Baier, that goodwill separates trust other, related concepts like confidence or mere expectation. Baier suggests that "we ought, as moral philosophers, to look into the question of what forms of trust are needed for thriving of the version of morality that we endorse."[135] She is concerned with how trust affects and is related to ethical theory; in what way does our conception of trust support a foundation for our ethics? Her position accepts, as a starting point, that trust is pre-moral. The project of this dissertation can be seen as taking up the gauntlet that she has thrown down: what form must trust take in order to make sense of our ethics? The work done so far has been, for the most part, laying the groundwork for a conception of trust that can be the foundation of an ethics. I have shown that there is no direct causal link between trusting and trustworthiness such that they entail each other but that trusting does entail a *belief* that the trusted person is trustworthy; I cannot be trusting of another *without* believing that the other is worthy of trust. I have shown that trusting and entrusting as well as being trustworthy and being entrusted, though often occurring simultaneously, are neither coextensive nor necessarily causally connected and so we must separate the two ideas in our conceptual understanding; our actions may often indicate our trusting or trustworthiness but they do not demand that we be trusting or trustworthy. Further, I have shown that trust is risky; it requires and is, in a sense, a

[135] Baier, *Moral Prejudices : Essays on Ethics*, 96.

product of uncertainty. Our uncertainty about the behavior of others is, in Baier's terms, a special kind of vulnerability: an opening of oneself to the discretionary judgments of others. Having made these distinctions we are now confronted with the question: if actions are not the determinant of trust, what is? When a particular action can be interpreted as being either trusting, trustworthy, or strategic then there must be some other feature that can help us to differentiate the trust from the non-trust. In this chapter I address this issue; I suggest that it is a particular type of motivation that separates the trust from the non-trust. Care and concern for the other is the motivation that is necessary for trustworthiness and so a belief in the care and concern that another has for one's self is necessary for trusting. Further, goodwill is not simply willing positive things for another, one must exhibit good judgment in the application of that will.

One concern that the preceding chapters might suggest is that, if the analysis is correct, we seem to have very little reason to actually engage in trusting. Focusing on individual instances of trusting or trustworthiness provides little reason to do either. Being open to the uncertainty and harm inherent in trusting is not something that we would otherwise endorse so why do so in these instances? The response, if trust is to be non-instrumentally valuable, must include a justification for trusting that does not ultimately refer to the success of the trusting in question; if a cost-benefit analysis shows that releasing control of our interests can produce a greater benefit than harm then the rational choice is to release control but, as Niklas Luhmann's discussion showed, this is insufficient for trusting. Having more to gain than to lose in an interaction is not trust; and so, Virginia Held asks "[w]hen the probability of whether the other fellow can be trusted is either totally unknown, or exactly .5, is it *rational* or not to take a chance on the

good will or 'trustability' of another human being?"[136] Her question probes when trust is appropriate given the conditions that a cost-benefit analysis *cannot* be calculated. The goal of this chapter is to show why we trust; this account is meant to be descriptive. Understanding the kind of motivation necessary for giving over control of our interests to another or for accepting responsibility for controlling the interests of another will help us to understand when we should be trusting or be trustworthy.

I suggest, following Baier, that what provides us with a motivation for trusting, what answers Held's concern, is the belief that the other is acting with goodwill towards us; our interests are important to them *because* they are our interests and not because of some other good end that will be achieved for the other. The response that I present is very similar to the account of the good will presented by Immanuel Kant when he says "inasmuch as reason has been imparted to us as a practical faculty, i.e., as one which is to have influence on the will, its true function must be to produce a will which is not merely good as a means to some further end, but is good in itself."[137] That is, the good to be found in trust is not in the ends that are achieved by being either trusting or trustworthy but rather from the will that guides the person. On my account more specifically, we are only truly trusting or trustworthy when we believe in the goodwill of the other or are motivated by our goodwill and not when doing either will achieve some other good end.

I begin by returning to a consideration of the risk inherent in trusting. This is important because it is the nature of that risk that leads us to our explanation of trust; vulnerability to the decisions of others demands that the judgments of those others be

[136] Held, "Rationality and Social Value in Game-Theoretical Analyses," 216.
[137] Immanuel Kant, *Ethical Philosophy : The Complete Texts of Grounding for the Metaphysics of Morals and Metaphysical Principles of Virtue* (Indianapolis, IN: Hackett Pub., 1994), 9.

salient to trust. Next, we consider the relationship of actions to trust in order to show that it is the motivation for action, not the action itself, which indicates trust. Then, the role of motivations for trust are considered and I argue that goodwill is the only motivation that carries explanatory force. We then look at the importance of judgment to trust; our vulnerability to others may lead to concern for the judgments of others but the nature of those judgments must itself be considered in order present a full account of trust. I conclude with the claim that trusting another is to believe that she makes judgments with goodwill towards us, that being trustworthy is to be motivated by goodwill towards others.

It is important to understand, given the topic of this chapter, how Baier conceives of what is at stake in an investigation of trust, doing so will help us to hone in on our present concern. She is laying out a position that takes trust to be not only necessary for the functioning of our morality but also as important to the functioning of any cooperative activity; involvement with others, on her account, requires some degree of trust in order to proceed. Baier believes that we make ourselves vulnerable to the whims of another when we accept the risk of placing something important to us under her or his control. Trust, on her account, is interpersonal and our motivations play a central role in its justification. Her position is that we give over discretionary control because we believe the other has goodwill towards us. I adopt Baier's position but with one major caveat: since actions do not fully determine trust I cannot say that any instance of giving over discretionary control is an instance of trusting.

Baier's position is not without its critics or detractors however. There are those who suggest that trust is something else entirely. Hugh LaFollette points out that

"[s]ociologist Bernard Barber claims this trust is evidenced when we assume (1) that others will act within a persistent moral order, (2) that they will perform their technical roles competently, and (3) that roles that require a special concern for others will be faithfully fulfilled"[138] On this account the motivation for the action is conspicuously absent; trust seems only to require fulfilling certain actions. Barber's position, like many others, is one in which Baier's account, and the account that I present, must contend in order to be accepted.

Vulnerability

There is a significant difference between the relationships we have with people and the way that we interact with other entities and institutions: animals and institutions are not capable, strictly speaking, of having a bad will. People, unlike institutions, have the capacity to make decisions and choices, to follow whims and desires, to be mistaken and be led astray. People, unlike animals, have motivational states that we can understand, which are not merely anthropomorphized versions of motivations that we attribute to them; people can be motivated self-interestedly, altruistically, or by any intermediate combination of those two extremes. Interacting with people is unlike interacting with other things and, further, beyond these differences, our human interactions present us with a special kind of uncertainty. It is not simply that we lack a predictive capacity for the future, which is just as true for bridges as it is for people; there is always some probability that a bridge will fail, we lack the capacity to perfectly predict the success of crossing a bridge. Instead, the uncertainty of dealing with people involves their discretion. We often *feel* as if an institution is using some form of discretion by

[138] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 115.

discriminating either against or for us but it seems that, really, it is some particular person in that institution that is using her discretion to our detriment or advantage. Even when an institution systematically treats one group of people differently than another it is the individual members of the institution and not the institution itself that are responsible for how that group is treated. My goal here, however, is not to present an argument suggesting that it is inappropriate to ascribe belief states to institutions but rather to separate a discussion of trust in institutions from a discussion of trust in other human beings; I am concerned here with interpersonal trust.

Trudy Govier points out that "friends cannot control each other. When we trust, we accept our vulnerability."[139] Her point is that people, our friends, are not mechanical devices whose actions are predictable, regular, or even probabilistic; people are free to act in any number of ways. When we interact with people, most specifically when we are trusting of them, we become vulnerable to their judgments. Trusting, as Govier shows us, is to be vulnerable. This vulnerability shows why it is critical that another has a goodwill towards us. Govier is not the only author who holds this position, Hugh LaFollette makes the same point when he says that in relationships in which one person reveals something sensitive to another "[t]he revealer trusts that the listener will not harm her or abuse her welfare. She thereby makes herself vulnerable to the listener: she exposes herself to the possibility of exploitation."[140] The threat or the harm attached to opening yourself to another human being is not only the uncertainty of how that person is going to act but why that person is going to act the way she will. There are a number of ways of talking about the uncertainty of human life; Alastair MacIntyre discusses the uncertainty attached

[139] Trudy Govier, *Dilemmas of Trust* (Montreal ; Ithaca: McGill-Queen's University Press, 1998), 49.
[140] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 111.

to personal interaction when he describes the "four sources of systematic unpredictability in human affairs."[141] I, however, am presently concerned with only two kinds of uncertainty. One type of uncertainty that affects our daily lives is a generalized uncertainty that we have about all future events, which maps onto MacIntyre's discussion of the 4th type of uncertainty: "trivial contingencies can powerfully influence the outcome of great events."[142] This is an important source of uncertainty but ultimately it is not the form that is most important for showing that trust requires goodwill. The other type of uncertainty is a localized uncertainty about the decisions of other (particular) people, which maps onto MacIntyre's description of "the game-theoretic character of social life."[143]

The dangers and uncertainties of being involved with others need not be tied strictly to their immediate actions. Our friends and loved ones may cause us harm by failing to act or failing to tell us something. Robert Goodin says that "it makes perfectly good sense to speak of someone's being vulnerable either to harms that come about through others' omissions or neglect or to harms that come about through others' positive actions."[144] Omissions, and what they reveal about the attitudes of those around us, can be extremely damaging. And, because we do not have perfect access to the internal thoughts of others we can never be certain of our place within their interests. Are we, to them, a tool for some other end or end in ourselves? Is it for our sake that they are concerned or for the sake of something else? Consider the thought experiment generated

---

[141] Alasdair C. MacIntyre, *After Virtue : A Study in Moral Theory*, 2nd ed. (Notre Dame, Ind.: University of Notre Dame Press, 1984), 93.
[142] Ibid., 100.
[143] Ibid., 97.
[144] Robert E. Goodin, *Protecting the Vulnerable : A Reanalysis of Our Social Responsibilities* (Chicago: University of Chicago Press, 1985), 110.

by an episode of Star Trek: Voyager: it is suggested by an alien race, the Devore, that telepaths cannot be trusted; they cannot be trusted because, for the telepaths, there is no uncertainty about the motivations of others; telepaths *know* what drives others to act.[145] The concern voiced by the Devore leader Kashyk is with how society is structured and how positive interactions between peoples can best be achieved; knowing the feelings and motivations of others bars a particular form of interaction, trust, which is deemed to be important to the functioning of a society. My position disagrees with a simply statement of the assessment of this fictional alien race. I argue that it is not the case that telepaths cannot be trusted or rather it is not the case that they are to be *dis*trusted; knowing the motivations of another does not immediately make one more inclined to take advantage of others. Rather, my position is that there is no need for trust when there is no uncertainty about motivations; knowing for certain that you will or will not ignore my interests in favor of your own interests eliminates not only the need for but the possibility of trusting. Kashyk then is correct when he claims that trust cannot exist without uncertainty about the motivations of others (normal non-telepathic interaction) but he is wrong about the effect such a lack of uncertainty will have; it does not produce distrust but rather obliterates the possibility of trust (either trusting or distrusting). It is not, however, the place of this work to engage in a discussion of whether a better life would be one without trust; I am, honestly, unprepared to answer this question. Uncertainty is part of our existence, it is therefore important that we understand the effect that this uncertainty has on all our lives; as Goodin points out "vulnerability implies more than

[145] Les Laundau, "Counterpoint," in *Star Trek: Voyager* (United States: Paramount Television, 1998).

susceptibility to certain sorts of harm, however, broadly conceived; it also implies that the harm is not predetermined."[146] Uncertainty creates the conditions necessary for trust.

Vulnerability and our response to it have a profound effect on our lives. Robert Swinth suggests that a "person, by exposing himself, effectively makes a statement and asks a question."[147] The statement that we make by exposing ourselves is that we (on my account) believe that the other is motivated by a concern for our interests. It will not be enough to believe that the other is acting from a concern for either duty or right; these motivations, while noble, fail to differentiate me from any other person – a friend who acts solely from duty is just as good a friend to every member of society and therefore is no particular friend of mine. Trusting another is not to believe that the other will be untrustworthy to others while being trustworthy to the one who is trusting but it must contain some belief that the individual person who is trusting is relevant to the relationship.[148] Acting from duty or from a concern for the right can lead to the trusted person acting counter to the interests of the one who is trusting. The question we ask is whether the person will fulfill that trust or take advantage of it. In chapter 5, this statement/question interpretation will be shown to provide us with a reason for acting; being trusted is itself a reason for acting in a trustworthy manner. At this point, however, it is enough to recognize that we are exposing ourselves whether in the form of a statement, a question, or something else. Trust involves uncertainty and that uncertainty is important for understanding the nature of trust and its relationship to actions.

---

[146] Goodin, *Protecting the Vulnerable : A Reanalysis of Our Social Responsibilities*, 112.
[147] Swinth, "The Establishment of the Trust Relationship," 335.
[148] I make clear later that there are instances in which the trusted person can act against the immediate interests of the trusting person and still not be violating the trust but that explanation must wait until this portion of the account is fully explicated.

Motivation

Risk and uncertainty create a situation in which we can never know, beyond all doubt, how the future will turn out. While this is true of every decision it is especially true of decisions that we make about our involvement with others. In this sense, every decision involves trust; the fullest account of this, however, must wait until chapter 6 when all the necessary ingredients have finally been gathered together. It might seem from the pervasiveness of uncertainty that the next best alternative is to play the odds. That is, if actions of the sort that we are interested in result 60% of the time given a specific set of initial conditions then it would be a good bet to try to create those initial conditions. At the very least it would seem that expecting those actions to result when those conditions are met is a reasonable expectation. I do not want to give the impression that I am opposed to a generalized account of inferential reasoning but the issue that takes center stage when considering trust from this perspective is the likelihood of the fulfillment of an action; actions and behaviors become the determinant of trust but entrusting has already been shown to be a different than trusting. Playing the odds is concerned with entrusting not trusting. The question for us, however, is whether trust really is something that we should, or even do, base on actions in the first place; if trust is not entrusting then perhaps it is some other action. A number of authors consider the actions associated with trust to be trust but I contend in this section that actions are not trust; that what trust requires is not behaving in a certain manner but rather being motivated by the proper motivations. Promising is an action that is often considered to be parallel to trusting and so it is useful as a model for what trust, as an action, would look like. In this section I consider promises and show that trusting cannot follow promising's

lead since promises themselves rely on trust. I consider possible motivations for trust in the third section of this chapter.

Promising's Disanalogy

Promises, though not directly analogous to trusting, can be useful for showing why actions, by themselves, are incapable of accounting for trust. I suggested in Chapter 2 that a failure to keep one's promises would indicate either that one lied about what was promised or that one abandoned the promise that was made. It is possible that the promiser might construe her actions as not really abandoning the promise, in which case she would see no failure, but an explanation or justification would still need to be given to the promisee or else there would still be, at least from the perspective of one party to the exchange, a broken promise. The apparent change in the actor, from the condition in which the promise was made, that results in a failed or broken promise presents a very different scenario than what can be said of failures or breaches of trust. This difference stems from the conditions under which promises and trusts are entered into; Russell Hardin points out that in order to be a promiser, one must commit a prior act of promising. [149] But a prior claim of this sort can never be required of being trustworthy such that it is significantly differentiated from being a promise; that is, the prior act of committing to being trustworthy simply *is* a promise.

Hardin considers the nature of commitment at length when he discusses *The Ring of the Nibelung*, which is part of his treatment of dispositions as they relate to trust.[150] He focuses on the character of Alberich and the fact that this otherwise despicable individual

[149] Hardin, *Trust and Trustworthiness*, 63.
[150] ———, "Trustworthiness," 27.

was able to commit to a future set of actions that would require him to be a [fairly] decent person. The concern, for Hardin, is that it seems improbable that Alberich actually *could* commit to or even to predict his own future actions, especially actions that would be so divergent from his nature. Alberich, in the first of the four operas in The Ring cycle, vowed to not lust after the Rhinemaidens. Hardin's contention is that commitments of this sort are simply beyond our capacity as human beings; Alberich must both deny his nature *and* be prepared to refuse to act in a particular way regardless of any future conditions or events. That, Hardin suggests, is not possible. We cannot predict what things will confront us and so we cannot commit to not responding in certain ways. Since we cannot make these sorts of commitments we cannot, in these situations, ever be trusted in regards to those things. Hardin claims that we cannot expect Alberich to keep his vow and so Hardin asks "could one reasonably trust his commitment?"[151]

As an aside, before I continue with the discussion of promise and what it reveals about the separation between trust and actions, I have some concern about Hardin's characterization of Alberich's action in acquiring the Rhine gold. Hardin suggests that "the *Ring* turns on a bit of psychological nonsense" when it expects us to accept that a man, especially the character of Alberich, could fully renounce all future sexual desires in a single moment.[152] I suspect, contrary to Hardin, that Wagner has given us a far better explanation of human psychology than he is being given credit for. The focus is on the moment when Alberich forswears love of which Hardin says that "[b]y committing himself forever, Alberich gains extraordinary power."[153] I want to suggest, however, that

[151] Ibid.: 26.
[152] Ibid.
[153] Ibid.

this is not what has happened; Alberich did not gain power because of his commitment. The context of the curse tells a different story. The Rhine-maidens have been tasked with protecting the gold by their father; it is, in fact, their reason for being there and yet Alberich takes the gold from them without a fight. As protectors, they are seemingly inept until we consider what they have to say about the gold. In response to the question of who has the power to forge the ring Woglinde says

> Nur wer der Minne Macht versagt,
> nur wer der Liebe lust verjagt,
> nur der erzielt sich den Zauber,
> zum Reif zu zwingen das Gold.[154]

Her claim is not in the form of an if-then clause. Woglinde is not suggesting that if one renounces love then one will be granted the ability to fashion a ring from the gold but rather that *only* the person who has given up love has the power to fashion the ring. The nymphs then are not protecting the gold through strength but instead through guile: they are distractions. Only the man who can ignore their beauty can seize the gold. If we are taking Wagner to be presenting a psychology then it is this: humans are motivated by many things, among them are our drives for lust and for power, but being led by our lust will hamper our pursuit of power. Alerbich does not, on this interpretation, commit to renouncing love but instead finds something to replace it; "[i]f love is denied me a newer delight I may know!"[155] Whether we should trust his commitment to renouncing love is beside the point; Alberich's cursing love is the result of his motivations, not the cause of them. But there is still an important issue to be resolved here and Hardin's discussion of

[154] Richard Wagner, *The Rhinegold*, trans. H. and F. Corder, Grand Opera Librettos (Boston: O. Ditson Co., 1904), 20.
[155] Ibid., 21.

Alberich can be useful even when my disagreement about interpretation. When Hardin asks about Alberich's commitment, he wants to know why we should trust hairy imp. Hardin wants to know what assurances we have that one will do as one has claimed; "[o]ur task is to contrive real-world substitutes for such magical powers as Alberich faced to keep us in line."[156] Contriving of real-world substitutes that will constrain our actions or the actions of those who have committed to us leads us down the path of conceiving of trustworthiness as behaving in particular ways.

Verbal Commitments

Recognizing that promises are verbal commitments the use of Alberich as an example by Hardin raises a question. Alberich, by renouncing love, has effectively promised not to pursue the lovely Rhine-maidens. If Hardin is right in his discussion in *Trust and Trustworthiness* that promise is inappropriate as an analogue for trust, then what use is Alberich to his discussion in *Trustworthiness*? If promising requires a prior act while trusting does not then how are we to understand an extended discussion of the seemingly necessary failure of Alberich to keep to his commitment and thereby be trustworthy? Alberich's ability to make such a dramatic commitment is the focus of Hardin's discussion but we might ask to whom is Alberich committed? His commitment seems to be in the form of an exchange, renounce love in order to gain power, hence Hardin's question about whether we can trust Alberich. But, whether we can trust Alberich should not be affected by Hardin's conclusions about promising since, on Hardin's account, Alberich's commitment, as a promise, is not analogous to whether we should trust him. Because the promiser/promised relationship is disanalogous to the

[156] Hardin, "Trustworthiness," 42.

trusting/trustworthy relationship the value of Hardin's discussion is suspect. Ostensibly, the ability to make commitments and to be believed in regards to those commitments is at stake in both promising and trusting but if this is true then surely the analogy between promising and trusting is functional after all. Therefore, it must be that either Hardin's discussion of Wagner's classic serves no purpose because promising is not analogous to trusting or it confuses the distinction between accepting a promise and trusting because both rely on our ability to recognize the commitments of the promiser/trusted party. But, as I show in this section, making a promise is an act in a way that being trusted can never be.

There might be a way for us to escape the promise/trust conundrum however. Hardin admits that "trustworthiness is a motivation or set of motivations for acting" and that we are motivated by those commitments which we hold.[157] When he argues against the riskiness of trust Hardin does so by pointing out that trust is not an action and so his concern for the motivation in this instance is no great surprise. Because of this, however, we can avoid the problem of Alberich that Hardin presents; it is not the action of keeping one's promise that is the specific concern of trust but rather the motivation that leads to that action. Therefore, the action of Alberich failing to fulfill his promise is not the issue in our trust of him; actions are about entrusting not trusting. Rather, since trustworthiness is about motivation we can look to what motivates Alberich to hold to the promise he has made. Alberich's success or failure at fulfilling our trust is not as important as his motivation for acting, the motivation which led him to make the promise in the first place. As we will see later in this chapter Alberich's motivations are suspect and not of

[157] Ibid.: 31.

the sort that engender trust but it is not a problem that Alberich is incapable of knowing all the future situations in which he will be placed because Alberich, regardless of situation, could have the appropriate motivation for being trusted (although, in this case, he simply does not).

The Question of Alberich

Suppose, however, that we accept the discussion and criticism brought forth from *The Ring* that Hardin presents: accept that a promise and trusting are not analogous. His account still raises questions. He holds that Alberich cannot be trusted to hold to such a commitment because no one can predict how his future motivations will affect his later self. Hardin's concern is that Alberich, and indeed anyone else, cannot know that his future self will continue to hold the same beliefs that his current self does. Since we cannot know our future selves, Hardin claims that we cannot make commitments on the order of the promise that Alberich has made. But, if we cannot predict how future conditions will affect our motivations and the commitments which we hold then the problem is not with specific, seemingly difficult commitments like Alberich's but instead with all commitments. If I cannot predict how future events will effect my motivations then I cannot promise or be trusted with something as innocuous as arriving at an appointment tomorrow because I cannot know what things will transpire between then and now which might cause me to not want to be keep that appointment. But surely this claims far too much; we could *never* trust anyone if this criticism were correct. We could not even predict our actions mere moments in the future. This response would require us to treat our future self as some unknown stranger who may or may not support our

current interests and concerns. We would, in this scenario, be reduced to erecting all the same barriers against the defection of our future self that we feel we must now erect against the defection of complete strangers. We, if Hardin is substantially correct, would be forced to plan our future as an elaborate strategy; we would be required to commit intrigue against ourselves. That our future self may not share all of our commitments, which seems a reasonable point, when considered fully, is revealed to be far too problematic to account for our experience in making and sustaining commitments. We *do* make commitments and, when the time comes to fulfill them, while we may be inclined to ignore those commitments because they no longer match our life goals, often we do fulfill those commitments. That we have committed to them is a reason, above and beyond our current desires, to fulfill what we have said we will do.

The Brothers Karamazov

Our ability to trust others is, under Hardin's conception, wholly dependant on our ability to recognize the commitments of others and the motivations which they hold, we trust when we believe that another holds our interest within their interests. Granted that Alberich has made a particularly difficult commitment to some future motivation that should cause us to question the veracity of his promise but Hardin's position is still problematic; if our future actions are so variable that no prediction is ever possible then any commitment is in jeopardy, not just those of questionable initial motivation; it is possible that good, not just bad, motivations can change. Hardin's position suggests that we can never trust anyone, let alone Alberich. However, I contend that what is important in the distinction between promising and being trusted is the particular commitment; a

promise carries with it the commitment to fulfill the promise, which is to fulfill a particular action, while being trustworthy should carry the commitment of being concerned with the other's wellbeing, which need not require fulfilling any particular act.

Hardin uses the justification of promising as an example of how iterated prisoner's dilemmas can generate strong incentives without appealing to moral notions; "as is commonly true also of trust relationships, promising typically involves intentions on the parts of two people."[158] His purpose is to show that our continued interaction provides its own incentive for behavior, that maintaining a relationship gives reason for acting in a trustworthy manner. Karamazov's recognition of Trifonov's desire to continue their relationship would then provide Karamazov with the evidence necessary for trusting Trifonov.

In making his point Hardin uses *The Brother's Karamazov* in which Lieutenant Karamazov enters into a questionable arrangement with the merchant Trifonov. Karamazov hands over money meant to pay the troops under his command to Trifonov who, after using the money as capital for investments, returns the money with a bonus for Karamazov. This arrangement goes along swimmingly until Karamazov gets transferred at which point Trifonov claims to have no knowledge of the money or the arrangement. This puts Karamazov in a difficult situation; he cannot report the merchant because he has himself broken the law. In this example, so long as Karamazov continues to be stationed where he is, Trifonov has a reason to be trustworthy: their continued interaction. However, once Karamazov will no longer be in town Trifonov has no reason to continue the relationship.

[158] Hardin, "Trustworthiness," 19.

Hardin's discussion of *The Brother's Karamazov*, rather than helping his position, provides an example of exactly what is wrong with the 'encapsulated interests' account that he presents; once the incentive of future interaction dissolves there is no longer any reason for Trifonov to be trustworthy to Karamazov. Karamazov's interests are encapsulated within Trifonov's interests because it is in the merchant's interests to take Karamazov's interests seriously. But once it is no longer in Trifonov's interest to consider Karamazov the relationship immediately dissolves; their only tie was Trifonov's interests. If the trusting party has already invested in the relationship, as Karamazov did in Tolstoy's story, then the relationship dissolves via a breach of trust. The revelation of Trifonov's motivations, a concern for his own interests, makes it obvious that he was never the kind of person in whom our trust should be placed, which runs counter to the point Hardin is trying to make, which was that Trifonov was trustworthy when it was in his interests to consider Karamazov's interest. Obviously however, Trifonov was never worthy of trust because one who is worthy of trust would not violate that trust simply because she could gain a greater financial reward. Finding it in my interests to consider your interests is not enough to make me trustworthy. If I am considering your interests simply as a means for fulfilling my interests then I am acting strategically in regard to my interests, which is a motivation that can lead to betrayal as easily as it can lead to success.

Promise Requires Trust

There might, finally, be some analogue between promising and trusting after all. As Hardin pointed out, promising requires an overt first act of making a promise. But, under his description, trustworthiness has no comparable first act; I am able to form

opinions about your trustworthiness such that I can trust you without your knowledge or participation in the formation of that opinion. As we saw in Chapter 2, misplaced trust is not a failure on the part of the trusted individual rather it is a failure on the part of the one who is trusting; on Hardin's account the lack of a prior first act makes all breaches of trust cases of misplaced trust. Failing to fulfill a promise, on the other hand, is a failure on the part of the promiser and not on the one to whom the promise was made. This distinction separates the acts of trusting and accepting promises. However, contrary to Hardin I contend that being trusted in regards to some act or thing is not such that we enter into it unawares. Imagine that we have come to know each other fairly well and then one night, at a party, I find you sniffing glue under the stairs. I, in the vernacular, may have trusted that you were not the kind of person who would do that but my impression about what kind of person you are is not the kind of trust that Hardin is discussing. Trusting in your character does not match the three-part relation that is central to Hardin's thesis; that I trusted you not to be the kind of person who sniffed glue under the stairs fits the three-part relation in only the loosest of senses.[159] Admittedly, this lack of the necessary components for trust is one of the reasons that, on Hardin's account, there is no analogue between trusts and promises.

Imagine instead the case of leaving your children with a baby-sitter. It is understood that the babysitter's job is to protect and look-out for the children and as such any acceptance of a babysitting job should count as a tacit promise to do those acts. It might be prudent to make explicit the promise of baby-sitting but clearly the contractual nature of the relationship implies that a promise has been made, whether or not it is

[159] Hardin considers the trust relationship to be a three-part relation, A trusts B in relation so some thing C; a point that will be considered in detail in Chapter 5.

spoken. Since a promise cannot, as Hardin argues, add anything to your knowledge about the trustworthiness of the babysitter the existence of even a tacit promise in the relationship should not have any effect on the trust you have in the babysitter; the promise will not *make* you trust the babysitter. However, even with the existence of a tacit promise you *do* and, I argue, *must* still trust your children with the babysitter.[160] Additionally, the babysitter *is* making a claim about being trustworthy which goes beyond any particular promise that the sitter may, or may not, have made. Whatever reason the sitter may have for making himself available to care for the children of others one claim that he must, either implicitly or explicitly, make is to be the kind of person with whom others can safely leave their children. That is, the babysitter really is making a prior claim to being trustworthy by accepting the charge of taking care of another's child, which is not simply a promise; the sitter is claiming to have a certain kind of character to be responsible in the right kind of ways. This relationship has the requisite three-part relation for Hardin's position: you trust the babysitter with your children. This relationship also has the prior act inherent in making a promise. This case of trustworthiness then, at least as I have just described it, certainly seems to be an analogue to promise-keeping. Though it seems as if we are no closer to reaching a conclusion about the nature of trust it seems that we have strengthened at least one part of our understanding of it: whether or not trust is analogous to promise it is clear that the nature of trusting and trustworthiness, at least in the case of the babysitter, is not about particular actions but is instead about the emotions, feelings, or motivations of the people involved.

[160] The "must" used here will be discussed in Chapter 6.

Prior Acts

Hardin or others might argue though, if they were adamant about keeping promising separate from trust, that this example has the appearance of being analogous precisely because it is a case of promising. That is, when I contact you about sitting for my children you will make claims, provided that you want the job, about being able to take care of my children. You will promise to care for my children in return for compensation. There, Hardin might say, is the prior act, it is in the promising and not in relation to being trustworthy. If this were true then even the situation that I have described would not count as a prior claim of trustworthiness; it would be the prior claim of promising. However, Hardin admits that claims about being trustworthy, or even demands to be trusted, are not enough to cause others to accept us as trustworthy; "if I do not trust you, your demand that I trust you cannot be honored merely on the ground of your demand."[161] I will not accept your promise to dutifully care for my children unless I already trust you. That is, your promise to take care of my children adds nothing to my knowledge; the promise cannot convince me that you are able and willing to take care of my children. A promise might provide me with some kind of recourse in the event of your failure to keep it but the promise itself does not provide any additional assurance which will make me comfortable enough to trust you. Prior to accepting your promise about the disposition of something that is important to me, I must first believe that A) you are the kind of person who keeps their promises, B) who acts honestly, and C) who has genuine concern for me and my projects; that is, I must first believe that you are trustworthy. At the very least, I must believe that you are not *un*trustworthy; that is, I

[161] Russell Hardin, "Conceptions and Explanations of Trust," in *Trust in Society*, ed. Karen S. Cook, Russell Hardin, and Margaret Levi (New York: Russell Sage Foundation, 2001), 11.

must believe that you are not acting counter to my interests in order to relinquish control of those interests to you.

Considering the alternative will make this point clear. Imagine a father accepting the promise of a babysitter to take care of his child when that father had claimed that he did not trust, in fact distrusted, the babysitter. I find this example wholly implausible but think that it is worth pursuing in order to make that point. If the father truly accepts the promise then he should leave the babysitter at the house and never have any reservations or concerns. Admitting that the father has concerns or reservations admits that the father has not fully accepted the promise of the sitter. This should not be taken to be an insult – a child is a parent's most precious commodity, there is no wrong in being concerned for who is caring for your child. If the father is not comfortable with the babysitter, in fact, places a nanny-cam in the house in order to watch what is happening, then we have *prima facie* evidence that the father did not accept the promise of the babysitter but instead reluctantly entrusted the child into the care of the sitter and set up precautions against the sitter's defection. That is, when we maintain that we can accept the promise of a person that we distrust we continue to confuse actions with the beliefs that motivate them. If I expect that my neighbor is going to throw my mail in the trash and I still accept his claim to dutifully pick up my mail then I am holding contradictory beliefs: I believe that my neighbor will throw my mail away and I believe that my neighbor will keep my mail safe. If, on the other hand, I know that my neighbor does not want to start a block war, then, even though I otherwise find my neighbor to be repellent, I can allow my neighbor to pick up my mail without holding contradictory beliefs. We do not accept the promise of those we find untrustworthy no matter how vociferously they attest to being worthy of

our trust. Therefore, in order to accept a promise about something that is important to me I must believe that you are worthy of trust and, in the case of a babysitter, the prior act of claiming to be worthy of trust is the act of making yourself available to sit for a child. This means that trust may yet be analogous to promising but if so it is because promising is dependent on trust and not because the two are somehow equivalents.

The analogy between trusting and promising fails not because trusting is not an action but because trust comes prior to the accepting of promises. However, a consideration of the two concepts reveals that trust requires concern for the motivations of the other because it is our conception of their motivation that allows us to form expectations about which behaviors and actions they will fulfill. In the next section I take up the question of which motivation is the right kind of motivation for trust.

Good Will

I argue that promising requires action or the expectation of action in a way that trusting does not; when we promise it is always a promise to do something specific, promising is tied directly to an action. Being trustworthy, on the other hand, far from being tied to a specific action ought *not* be considered as being able to fulfill an action at all. I have been suggesting that it is the motivation for action and not the action itself that is important for trust. The preceding discussion about the disanalogy between trust and promising makes this point but there are other ways to get at it. As I suggested earlier, if action is the determinant of trust then all that is required in order to be trusting is to

believe that there is a sufficient probability of successfully completing that action.[162] It would be the possible success or failure of the action in question that would determine whether or not trust has been fulfilled. In Chapter 2 however it became clear that trust can be broken without any action taking place; misplacing our trust is a regular occurrence.

Further, having the confidence that a particular action will be successfully completed seems inadequate for generating trust; LaFollette points out that "[b]eing confident about how another will behave is insufficient for trust."[163] He is reminding us that it requires more than simply having our actions be predictable in order to lead to trust. I can be confident that you will abandon your obligations to me; I can be confident that you will pursue your self-interest; I can be confident of all sorts of things that have little or nothing to do with our interaction; and I can even have confidence in the judgments that I make about you and your actions, which is problematic because, as I show in Chapter 6, it is trust that allows us to be confident about the actions of others and not confidence about their actions that allows us to trust. Each of these confidence inspired beliefs should reveal to us that confidence does not, on its own, lead to trust. LaFollette further suggests, in discussing the difference between having expectations and trusting, that "[m]inimally what is missing is this: I don't trust them to take an interest in my interests. I don't expect that they will guard information shared in confidence; nor do I rely upon them for help or assistance."[164] Morton Deutsch also argues that trust requires

[162] It might seem that a better than 50% chance of success would be all that was required but, under certain circumstances 51% might not be enough; if 49% of my patients die while undergoing routine operations I am likely not to have very many patients willing to entrust their lives to me.
[163] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 117.
[164] Ibid.

more than confidence or predictability; "predictability is clearly not sufficient to characterize the everyday meaning of 'trust'"[165]

Predictability or our confidence in the successful completion of some action misses what is most important in judging the value of the action. But this point is not new; Immanuel Kant argues exactly this when he says "moral action depends…not on the realization of the object of the action, but merely on the principle of volition according to which, without regard to any objects of the faculty of desire, the action has been done."[166] Recognizing that it is the principle or, in this case, the motivation for acting which is important is not yet enough to provide us with a complete understanding of trust however. Not every motivation is going to be a motivation that we find acceptable; what kind of motivation is necessary for trust?

The Importance of Friendship

Considering friendship is a good place to look for clues about the nature of trusting now that we have eliminated a number of possible distractions. Whatever trust turns out to be it seems obvious that we trust friends more than we trust strangers or, alternately, it might be that we have as friends only those that we can trust (while non-friends are those that we cannot trust). Either way, it appears that trust and friendship have some link. Trudy Govier says of friends that "we have a sense of being liked and valued for ourselves."[167] Our true friends do not use us for some other end but instead are interested in who we are. She is not alone in considering friendship this way. Aristotle

[165] Deutsch, "Trust and Suspicion," 265.
[166] Kant, *Ethical Philosophy : The Complete Texts of Grounding for the Metaphysics of Morals and Metaphysical Principles of Virtue*, 12-13.
[167] Govier, *Dilemmas of Trust*, 31.

describes several types of friendships. In one type of friendship, friends of convenience, which is far from the ideal type of friendship, the 'friends' cease to be friends when the use found in the other ends; "when the cause of their being friends is removed, the friendship is dissolved too, on the assumption that the friendship aims at these [useful results]."[168] Govier continues discussing friendship by claiming that "we expect friends to be good for us, to do us good."[169] These two points are important; we expect friends to be concerned with us and not some other end and we expect friends to do good for us. Friends have the right kind of motivation in regard to us and not simply that they fulfill the right kind of actions. A friend that only cared for us in order to achieve some other end would not long be our friend while a friend that did bad for us would also not long be our friend.

It might be objected that we are often friends with those who do not further our interests, that we have friends that are a bad influence on us. This objection certainly has merit and undoubtedly we could each produce any number of examples of exactly this situation. But, we must be careful to separate those that have bad influence on us from those who *we* believe do us wrong. Plato points out that no one would actively corrupt their neighbors because to do so would bring with it the possibility of harming our self; "does the man exist who would rather be harmed than benefited by his associates?"[170] That is, it is easy to imagine having friends that unintentionally stymie our interests and it is easy to imagine friends that can be said, from an external perspective, to actively harm our interests but it would be strange for me to continue to call another a 'friend' when I

[168] Aristotle, *Nicomachean Ethics*, 211-12.
[169] Govier, *Dilemmas of Trust*, 32.
[170] Plato, G. M. A. Grube, and John M. Cooper, *The Trial and Death of Socrates*, 3rd ed. (Indianapolis, IN: Hackett Pub., 2000), 28.

believed that this person was actively doing me harm; at the very least, this opposes the drive for self-interest that is so commonly accepted as being part our decision process. Friends then, even considering the objection, are those who we can expect to, at the very least, not actively harm our interests.

Not actively harming our interests is not going to be enough however to account for trust or even friendship. A person could never have actually harmed our interests and yet could still behave in a manner that we dislike. LaFollette points out that "*how* information is shared – is an essential element of intimacy."[171] There are many ways that bad news can be shared. Someone who does not know us might be blunt and uncaring, simply revealing the information without regard to how it will affect us. But a friend, we expect, will be concerned with how the news will affect us and so will try to do what they can to minimize the damage. Mary Wollstonecraft says that "a being, with a capacity of reasoning, would not have failed to discover, as his faculties unfolded, that true happiness arose from the friendship and intimacy which can only be enjoyed by equals"[172] Friendship, as she is reminding us, requires that the two people have the right kind of motivations for interacting. Between unequals the motivations of each party will always be suspect but, as she points out, those with the capacity to reason will actually want to be on equal footing with those around them since that is the only way to true happiness. Aristotle also recognizes the problem posed by inequality; "it is also clear with kings, since far inferior people do not expect to be their friends; nor do worthless people expect

[171] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 110.
[172] Mary Wollstonecraft, *A Vindication of the Rights of Men*, Great Books in Philosophy (Amherst, N.Y.: Prometheus Books, 1996), 19.

to be friends with the best or the wisest."[173] Friends must be able to be confident about the motivations of the other in order to be equal and the right kind of motivation will be one in which the other is concerned primarily with the good of the friend.

The Importance of Incentives

Russell Hardin's conception of trust does not require this equality in motivations however. His position is misleading in its simplicity; "trust as encapsulated interest."[174] I trust you when I believe that you encapsulate my interests in your interests. That is, when I believe that you hold my interests in your interests, I trust you. Hardin's position is not that I trust you when you hold my interests *as* your interests or even that I trust you when we share interests. Instead, I trust another when I form expectations about her future behavior based on the belief that she has incentive to consider and, to some extent, work towards fulfilling my interests. It is in this sense that trust is a form of knowledge. This knowledge is neither absolute nor perfect; trust has limitations. When I believe that she holds my interests in her interests I do not believe that she holds all of my interests in her interest, only that she holds some of my interests in her interest; according to Hardin, when we trust others we do not trust them in regards to all things and at all times but instead only in regards to some things at some times. The belief that she holds my interests in her interest is not the same as believing that she will in all cases hold my interests in her interests. I, in trusting, must admit that sometimes, in some circumstances, other interests that she holds will override her interest in my interests. Trust is not

[173] Aristotle, *Nicomachean Ethics*, 221.
[174] Hardin, *Trust and Trustworthiness*, 1.

absolute. Neither is my belief that she holds my interests in her interest infallible; we can, and often do, make mistakes about whom to trust.

The recognition of the incentives that others have to behave in certain ways is essential to Hardin's conception of trust; "[t]he encapsulated-interest account of trust holds that the trusted encapsulates the interest of the truster and therefore has incentive to be trustworthy in fulfilling the truster's trust."[175] It is the motivation that the trusted party has to complete the expected task that justifies whether or not we trust them. To be trusting is "[i]n a sense…simply to be explained as merely the expectation that the person will most likely be trustworthy,"[176] although, as we have seen, forming an expectation about the behavior of another is not enough to account for trusting.

Hardin's encapsulated interest account places trust "in the family of such notions as knowledge, belief and the kind of judgment that might be called assessment."[177] This account makes our capacity to judge extremely important to trust relations; when I am trusting I must be able to judge the worthiness of the trustee and when I am trusted I must be able to judge how my interests are furthered by considering the interests of the truster. The judgment inherent in trust may give trust the appearance of being something that is free to be chosen or not. Hardin argues however that trust is not something that we can choose; trusting is something that comes to us, "it just is."[178] It comes to us in the same way that beliefs come to us; I do not choose to believe the sky is blue, the evidence presented to me either convinces me or it does not. Hardin sees describes his account of trust as cognitive because, according to him, it is "grounded in some sense of what is

[175] Ibid., 24.
[176] Ibid., 31.
[177] Ibid., 7.
[178] Ibid., 58.

true."[179] I cannot force myself to believe that another holds my interests in her interest. If I do not trust her I cannot convince myself to trust her simply because I want the benefits of what trusting her will get me; "my trust of you can be neither offered nor withheld."[180]

Hardin opposes an explanatory connection between trusting and trustworthiness; "[t]he explanation of your actually trusting in some context will be simply an epistemological problem, evidentiary matter – not a motivational problem."[181] The explanation of your being trustworthy, on the other hand, *is* a problem of motivation; what compels you to behave in a trustworthy manner. Because Hardin sees being trusting and being trustworthy as functionally different enterprises a single justification cannot suffice for the two. I am presenting an account that is slightly different. While I struggle to keep separate the conceptual understanding of trusting and trustworthiness I do not present them as functionally different enterprises. Hardin argues that only trustworthiness is a problem of motivation while I claim that trusting is also a problem of motivation: my belief in your goodwill is not simply cognitive – it is, at least in part, affective. How, exactly my account plays out is only beginning to appear but its fullest statement must wait until chapter 6.

On Hardin's account when we consider our own trusting we realize that what generates our trusting is a belief about how the potential trusted is going to behave with regards to some thing in certain circumstances; our trusting is a kind of knowledge. But when we consider our own trustworthiness we realize that what generates our trustworthiness is a motivation to act in a specific way with regards to some thing in

[179] Ibid., 7.
[180] Ibid., 58.
[181] Ibid., 31.

certain circumstances; our trustworthiness is a kind of motivation. Being trusting is about knowing while being trustworthy is about acting. What justifies something as knowledge is very different from what provides the background necessary for acting. For Hardin, a single justification cannot therefore explain both trustingness and trustworthiness. He says that we need to be wary of any account of trust that attempts to condense the explanation each into a single account.

Both trusting and trustworthiness are motivated by goodwill however. I have already argued at length that trustworthiness requires one to be motivated by goodwill but, the question remains, 'how is trusting motivated by goodwill?' The bulk of the argument for this position will have to wait until Chapter 6 but I can say at this point that our ability to trust requires that we have goodwill towards the trusted party. Since we cannot trust for specific ends, I cannot trust you because doing so will get me some desired end, then trusting is not an instrumental activity. Further, trusting requires that we have a positive interpretation of the motivations of the other; we have to think good things about the other. Finally, if allowing the other discretionary control over our interests were going to cause that person harm, by either putting the person in the position that they will necessarily fail or because fulfilling the trust will require the person to substantially harm their interests, then we ought not and likely will not be inclined to put that thing in the person's control. Our trusting another then requires us to have goodwill towards them; when we have ill will towards another we may allow them to control our interests but not because we want what is good for them. Both trusting and trustworthiness then require goodwill.

Hardin criticizes most discussions that treat trust as an expectation by pointing out that an expectation involves either knowledge or belief on the part of the truster. However, most discussions treat trust "as a matter of behavior," so it would be a mistake for those accounts to treat trust as an expectation. [182] We form expectations about how others will behave or act but Hardin argues that trusting itself is never an action; those actions that we associate with trusting are a result of trust but not the trust itself. He says that trust is a set of beliefs that we hold not some particular or even general activity that we engage in. We can entrust others for any number of reasons without having that instance of entrusting being the result of our trusting that other. For instance, based on my past experience I may have formed a number of expectations about various automobiles. I may have had a particularly good set of experiences with Volkswagen such that, all other things being equal, I would select a Volkswagen over other cars. I might explain this, if asked, by claiming that "Volkswagen's are reliable cars." We might, therefore, be inclined to say, in the vernacular, that "I trust [my] Volkswagen with my life" (since if it were unreliable it might break down at the most inopportune moment). But it would be a mistake to conflate my willingness to entrust my life with an ascription of motivations to an inanimate object. Since it is so obviously possible to relinquish control of my interests to even those that I actively distrust it cannot be that the act of entrusting is equivalent with trusting.

If, however, we treat trust as an activity or a behavior then we should not also consider it an expectation. Treating trust as an action is putting the cart before the horse, so to speak, because we often try to explain actions by reference to trust or expectations.

[182] Ibid., 56.

‘I trusted my accountant’ explains why you did not audit your books; the lack of action on your part was because you trusted your accountant instead of the lack of action on your part being an example of trusting. ‘I expected my ex to cheat on me’ explains why your hired a private detective; the action that you took was a result of your lack of trust instead of the action being the lack of trust. If trust were itself an action then it would fail to have explanatory power; reference to another action suffices to explain the consequences of the actions but fails to justify those actions. ‘Why did the 8 ball land in the pocket’ can be reasonably answered by pointing out ‘because it was hit by the cue ball.’ But the question ‘why did you give that stranger your credit card number’ is not answered by responding ‘because my muscles formed the words,’ which is an action; in order to explain why acting in instances like this we need a motivational explanation. Trusting is not about behavior but rather about motivation for acting. Further not all motivations are motivations that would let us trust others, only a certain motivation will do. Being motivated by the best interest of another is the only motivation that can reasonably be considered as a motivation worthy of trust; other motivations may provide reasons for acting but only goodwill is a reason for trusting.

Judgment

Having shown that it is uncertainty which leads to trust we were led to consider the importance of fulfilling actions to trust. But actions were revealed to be less important than the motivations which drove them because something built on uncertainty cannot be justified simply by appeal to success rates. Not all motivations could lead to trust however since many motivations run counter to our desires and interests. Only those

motivated by goodwill towards us are trustworthy. That is, we trust those that we believe are concerned with our well-being.

Being concerned with the best interests of another is not, by itself, enough to fully account for trust. A person might want what is best for another in all instances and yet lack the capacity to judge what is actually in the best interests of another. As Hugh LaFollette points out "[a]n important element of trust is trust in the moral sensitivity, insight, and judgment of our intimates. In this case June would trust May to be sensitive and kind enough to know when she should override the normally stringent constraints on sharing confidences."[183]

Annette Baier contends that there are a number of actions that may resemble trust, foremost being reliance, but are not themselves trust. One thing that differentiates trusting another from relying on another is that the truster believes that the other is concerned with her well-being. Because trusting requires that we turn over discretionary control to another we will not be willing to do that for any old set of motivations. Trust, on Baier's account, is a belief in the goodwill of another which opens us to a special kind of vulnerability. And her position makes sense. If I am to leave my children in the care of a babysitter then I will want that babysitter to be concerned for me and my children and not simply with getting paid. It sounds obvious but since the baby sitter will not get paid if my children are harmed, and certainly will not be called to sit again, it is clearly in the best interest of the babysitter to consider the well-being of the children. Further, getting paid provides the babysitter with incentive for holding my interests (the safety of my children) in mind. And yet, a baby sitter that is known to care for my children because

[183] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 118.

they are my children and not because the sitter is getting paid will always be chosen over the baby sitter who is concerned for the money alone.

Goodwill is what allows us to turn over that control and to do so in a way in which we are comfortable and not fretting. When we believe that the other wants what is best for us and not merely wants to find a way to please us to get some other end which they desire we are not just able to trust that other but are already trusting her. The alternative is to take the time to consider other avenues for which we can achieve the ends that we desire, to weigh the pros and cons and to determine that either the pros win or they do not. But, as Luhmann made clear, when the cost-benefit analysis turns out in favor of handing over control, trust is not required in the interaction. It is possible that trust might still be present even in cases where it is in my calculated interest to turn over control but the issue then becomes one of which force is the motivating force. Since I am concerned mainly with describing what trust is, I do not take this to be a significant problem for my account. What I am arguing will only be relevant in the cases where one is actually trusting; if you happen to be trusting and have good calculative reasons for entrusting then all the better for you.

Trusting then is a belief in the goodwill of the other. But, it seems as if a belief in your goodwill is not enough. The issue of competency rears its ugly head. A friend may have all the right intentions towards me but yet be a miserable heart surgeon; I would not, therefore, trust my friend to perform my heart surgery. However, as I show in the next section, my friend performing heart surgery is a poor judgment on his part. And, rather than taking judgment to be an issue separate from a belief in the goodwill it is wrapped up in the belief that we hold. That is, I claim that believing that the other has a goodwill

includes believing that she has good judgment as well; otherwise we think merely that they have good intentions towards us but lack the capacity to have a truly good will towards us.

Belief

Consider the point raised by Virginia Held "if we say we *trust* a critic's judgment, we suppose he *could* decide one way or another, and we trust he will make the right choice."[184] Part of trusting in the case of the critic is not simply that the critic fails to be nefarious but it includes that the critic makes sound decisions. Judgment is requisite for trust. LaFollette carries this idea further when he claims of those we trust that "[w]hat I trust is that she will be a certain kind of person, who, among other things, will not intentionally harm my interests. Thus, I trust she will not undertake a task she knows she cannot do."[185] Trusting, as these authors are arguing, is more than simply acting, even in a very specific way. In the final section of this chapter I focus on the issue of judgment and show that good judgment, rather than being a separate issue, is actually part of the assessment of another's trustworthiness. Having good judgment will include a judgment about one's own capacities such that a trusted individual can be trusted even in cases where capacity is of supreme importance, like heart surgery.

Bad Trust Is Not Trust After All

Baier claims that trust is not always a good thing. If trust were, contrary to Baier's position, similar to justice then we might object and claim that trust, by its nature, is always good. But Baier argues that the two are not the same; justice put to bad ends is not

[184] Held, "On the Meaning of Trust," 157.
[185] LaFollette, *Personal Relationships : Love, Identity, and Morality*, 116.

justice after all while trust seems to be just as useful when put to bad ends as when put to good ends. We would not say that any particular act of justice is wrong; a bad act is not a just act. Any particular case of trust, on the other hand, will still appear to be a case of trusting even when that trust is put to bad ends; the individual members of an assassination plot seem to be trusting each other because they must each give over discretionary power to the other members of the plot. Using Hardin's conception we can see how even nefarious plots can meet a conception of trust; Assassin X, when she expects Assassin Y to do his part, is believing that Y holds her interests within his interests. However, Baier challenges the intuition that trust is always a good thing. She claims that "if the enterprise is evil…then the trust that improves its workings will also be evil."[186] Her position leaves as unsettled the moral status of trust. She accepts Sissela Bok's claim that whatever matters to humans thrives in an environment of trust but Baier adds to Bok's account that among the things that matter to humans are things that ought not be encouraged to thrive; that is, some humans are engaged in morally deplorable activities and trust helps those activities thrive as well.[187] Trust can be put to bad ends just as easily as it can be put to good ends; trust can improve the interaction of individuals engaged in activities that are considered evil.[188]

Having the good of trust hinge on the good of the activity which the trust is aimed at seems to make the value of trust good only insofar as it serves some other end. If trust is only instrumentally valuable then there would be no independent reason for believing that trust is a good thing and this investigation would have found a natural end; the value

[186] Baier, *Moral Prejudices : Essays on Ethics*, 131.
[187] Bok, *Lying : Moral Choice in Public and Private Life*, 31.
[188] There is an expanded discussion of this issue in Chapter 5 in which I argue that these cases of "bad trust" are not cases of trust at all but rather only resemble trust.

and worth of trusting would be found in the end that it was serving. To say that trust is an instrumental good is to say that there are times when we should not trust because to do so would be wrong. Baier does indeed claim that some instances of trust are wrong. But, it seems obvious that not all cases of bad trust are examples of instrumentality; there are practical reasons for believing that trusting another might be wrong in some instances, like when a known liar is asking you to believe a particularly outrageous claim. The wrongness of trusting a liar is a mistake in one's judgment about whom to trust; that kind of trust is wrong because it is misplaced – it is an epistemic failure, not a moral one. Baier's claim that some cases of trusting are wrong is not simply a claim about the epistemic correctness of whom to trust. When we say that stealing is wrong we rarely mean it in a purely practical sense that the end for which we were aiming would not be served by stealing. Instead we mean that the action that is identified as stealing is wrong in the sense that it is incorrect for you to engage in it, the end is wrong. The claim that some instances of trusting are evil is a claim about the correctness of trust as it is expressed in regards to that relationship and not about the correctness of that trust in fulfilling some end or goal. That is, as a practical matter, it would be wrong to trust Alfred if he is a liar but, as a purely moral matter, it would still be right for you to trust Alfred to deliver food to starving orphans.[189] Baier's claim is that, as a practical matter, it would be right to trust Betty if she is trustworthy but, as a moral matter, it would still be wrong to trust Betty to poison the city's water supply. So, to say that trust is wrong, in the sense that Baier is claiming, is to say that the trust itself is sometimes bad not that trust can sometimes be misplaced. The terminology here is difficult; both cases of trust are

[189] However, you are arguably doing something morally wrong if you knowingly put the safety of those starving orphans into the care of an individual that is likely to fail, through negligence or malice.

cases of misplaced trust, one is misplaced because trust should not be used to further those ends while the other is misplaced because the end will not be furthered by trusting in this case. When we claim that the good of trust is dependant on the activity to which it is aimed we seem to say that the good of trust is instrumental to the good of the end but we need to be careful about describing Baier's position on trust as one that conceives of trust as instrumental; that would be to misunderstand her position. She is not trying to show that trusting is instrumental but rather to show that trusting is not universally good.

However, we might not be inclined to abandon trust as intrinsically valuable quite so easily. We have already seen in this chapter that it is the motivation for acting which is important to trust but here we have another reason: judgment, ours and that of others, is important to our assessment of trust. In this section I show that it is our capacity to make judgments with goodwill which is the key to trust; both trusting and trustworthiness require decisions to be made with goodwill.

Dependence on Goodwill

One of the most important aspects of Baier's position is that "[w]hen I trust another, I depend on her goodwill toward me."[190] Having goodwill requires more than simply two agents having the same intention.[191] To have goodwill for another is to be concerned about what is best for the individual in question, even, and perhaps especially, when what is best for that person runs contrary to what is best for you. A shared intention, unlike goodwill, only motivates one to act when the intention in question is

[190] Baier, *Moral Prejudices : Essays on Ethics*, 99.
[191] This point is exemplified by any film of the *Dirty Dozen* type. A pair or group of individuals can come together for a common cause (defeating evil) or a shared intention (individual survival) without first having, or even developing, goodwill towards those others involved.

affected by the action, either directly or indirectly; that is, if you and I share an interest in our business then you can expect that in regards to the business I will act in a way that furthers our shared interest but you can form no sound expectations on that basis about how I will act in regards to your interests. Shared interests do not get us to trust others it only gets us to allow them to control certain decisions. Goodwill will motivate even when the shared interest is irrelevant to the issue at hand.

Each day we face situations that seem to require us to trust those around us and which clearly require us to rely on others; we accept food from relative strangers, we ride in multi-ton machines traveling at high velocity in close proximity to other multi-ton machines traveling at equally high velocity, and we even allow men and women in white coats to operate on us. Onora O'Neill makes similar claims about reliance in her lectures on trust.[192] We engage in all of these potentially terminal activities for various reasons. Sometimes we have a good justification; the restaurant at which I am dining has had several good reviews by people that I know and respect, the chefs are known for their desire to serve good food well, or that the agents of health department are watching the restaurant for any sign of malfeasance. Sometimes, however, we fail to consider the potential danger inherent in the activity; I often pull onto the highway without considering the speed and danger involved in driving on public streets. At yet other times we simply have no alternative; if my appendix is about to burst then I must choose between the only doctor available or certain death. Even if I know that this doctor has only a 50% success rate with this operation I can still go forward with the operation. We have a tendency to claim that acting in any of these manners requires us to trust those

[192] Onora O'Neill, *A Question of Trust* (Cambridge ;; New York: Cambridge University Press, 2002).

others but engaging in even the most dangerous activity need not require me to trust at all; I do not have to believe that the restaurant, other drivers, or even a shoddy doctor are concerned with what is in my best interest. It is enough that I believe that they are each concerned with their own best interest and that their best interest happens to work out in my favor. And so, Baier makes it clear that "[w]e can still rely where we no longer trust."[193] But Baier's distinction here is not the distinction between trusting and entrusting, instead she is trying to show that what we allow others to do in our stead need not be motivated by trust and may even run contrary to our opinions and feelings about their trustworthiness.

Baier is explicit that we need not trust others in order to count on them when she says that "[o]nce we have ceased to trust our fellows, we may rely on their fear."[194] That we can rely on those that we actively distrust may seem counterintuitive but there are many ways to ensure the 'proper' behavior of our compatriots and fear is one of the greatest motivating forces. Governor Tarkin in Star Wars expresses exactly this point; "[f]ear will keep the local systems in line, fear of this battle station."[195] Fear is a motivation for acting that is insufficient for generating trust. Acting from fear is part of our psychological make-up and once mapped is reliably predictable. This is Baier's point when she says, "[w]e all depend on another's psychology in countless ways, but this is not yet to trust them."[196] To rely on some particular response to stimuli is not to trust the actions of the other and certainly fails under Baier's conception because the other's

[193] Baier, *Moral Prejudices : Essays on Ethics*, 98.
[194] Ibid., 99.
[195] George Lucas, "Star Wars," (United States: Twentieth Century-Fox Film Corporation : Lucasfilm Ltd., 1977).
[196] Baier, *Moral Prejudices : Essays on Ethics*, 99.

behavior does not have to be motivated by goodwill and does not have to be in my interest at all.

Being able to rely on the psychology of another can be useful for cooperative activities but is not necessary for trust; knowing that you are just as interested in having a stable supply of water as I am can be make our combined efforts to create a dam easier. Baier maintains that trust can be betrayed but depending on another's psychology cannot properly be said to be a betrayal. If I trust you to catch me when I fall then I can be let down by your failure to do so but if I expect, based on my knowledge of your psychology, that you will catch me when I fall then your failure to catch me is an indictment of my belief and not of your trustworthiness; that is, I should not properly be mad at you when I rely on your psychology but should rather be mad at myself, which is markedly different than when you have betrayed me.

In Chapter Two, I discussed the various ways that trust can be broken. Baier also talks about the relevance of breaches of trust to our understanding of the concept. Her discussion however reveals the importance of the discretion given to the trusted party; "the babysitter who decides that the nursery would be improved if painted purple and sets to work to transform it, will have acted, as a babysitter, in an untrustworthy way."[197] The babysitter, in the example presented by Baier, will have taken far more liberty than what can be reasonably said to have been left in his care; he was left in charge of the care of the children and not in the decoration of the home. Conversely, one can do less than what being trusted has given us the discretion to do; a babysitter that fails to escort the children out of the house when the house is suddenly found to be infested with a strange and

[197] Ibid., 101.

disturbing type of vermin, perhaps a cloister of skinks, will not have exercised exactly that faculty which is the reason for his being trusted with the children; our capacity to judge is crucial to our being trusted. But it is only when we have trusted the individual that we are "let down" by the failure to behave as expected. If you had no other choice but to leave your children with a sitter and the sitter decided to paint your home then, as annoyed as you might be, the disappointment would be with being forced rely on the sitter and not with the trust you had (because you did not have any trust in the sitter). However, if you believed that the sitter had goodwill towards you and, upon your return, you found your home redecorated the disappointment would be with the sitter; you learn either that the sitter is not concerned with your best interests or you learn that the sitter does not know what your best interests are. These are obviously very different ways of being let down, the former being a betrayal and the latter being simply a failure, but in either case what changes is your relationship with the sitter, which is very different than the case in which you were forced to rely on the sitter. The separation between reliance and trust combined with the giving over of discretionary powers shows us the importance of goodwill and judgment to trust.

Accounting for Trust

Baier's definition of trust is meant to account for some of the subtleties of trusting, "to trust is to give discretionary powers to the trusted, to let the trusted decide how, on a given matter, one's welfare is best advanced."[198] Trusting is to let another determine, at least for a little while, how to care for something that is important to you. It would be strange to willingly put your interests in the care of someone that you do *not*

[198] Ibid., 136.

believe will be concerned with the fact that they are *your* interests. When I believe that you are concerned with my well-being and the well-being of those things that I care about I will be willing to put my interests into your care; the belief I have in your goodwill is a belief that you will further my ends. A shared-interests account attempts to explain why we give over discretionary power as well but that account, from Baier's perspective, ultimately fails. Hardin makes it clear that we can never be perfectly certain of how another will behave therefore giving over control based on the belief that the other shares our intention with regards to some thing is to invite disaster; the belief itself admits that the other has his own hopes, wants, dreams, and desires which he will attempt to further and, it happens, in this case our interests are the same. The other is acting on *his* interests when he furthers my interests and so when my interests come into conflict with his interests I must expect him to pursue his interests. But a goodwill account allows me to at least hope, if not expect, that the other will attempt to pursue interests that favor me; because the other is genuinely concerned about my well-being he will *want* to pursue things that are in my interests, sometimes even when those interests conflict with his own. Although, as Baier will make clear, concern for my interests must always be tempered; I would not want my friends to ruin themselves simply to fulfill a simple interest of mine.

Misusing Discretionary Powers

Baier states that "the belief that [another's] will is good is itself a good, not merely instrumentally but in itself."[199] To trust another is, on Baier's account, to believe in the goodwill the other has towards oneself. Sometimes, however, the good will behind

[199] Ibid., 132.

an action is not enough to consider that action good. Baier gives us the example of being alone in the library stacks: a woman alone in the library stacks is roughly shoved.[200] The stranger may have the good intention to help you become more aware of the dangers found in the library, like a collapsing stack of books. That individual might even have your best interest in mind by trying to help you avoid future harm. But, if that individual grabs you and shakes you in an attempt to frighten you into realizing how vulnerable you are then that individual does not appear to have done a good thing because the individual has shown poor judgment in how to achieve what is in your interests. In fact, on Baier's account the library do-gooder has broken, rather than fulfilled, your trust because he has misused his discretionary powers; he has gone beyond the scope of what you, as a stranger, have allowed him. We are vulnerable when we allow others to care for our interests; it necessarily means that our interests are at risk.

Baier maintains, contrary to Hardin, that trust is risky. Furthermore, she claims that trust brings a specific kind of vulnerability; "[t]he special vulnerability which trust involves is vulnerability to not yet noticed harm."[201] Baier's position here is, in part, concerned with the harms that we open ourselves up to in our normal lives; we eat food prepared by others, we drive on highways, we walk down darkened streets. We, when we think about it, have much that is left in the care of others, much that is not immediately and openly recognized as something that we have entrusted to others. Instead, most of the cases of entrusting in our daily lives are cases, cases that are often considered to be cases

[200] Ibid., 102.
[201] Ibid., 104.

of trusting, in which we have unconsciously accepted that others will look out for our interests.[202]

The vulnerability to unseen harms, which is the vulnerability to the myriad of situations that the trusted party will confront that cannot be predicted, is considered by Baier to be a "special sort of vulnerability."[203] It comes from the discretionary power of the individual who is trusted. By turning over some discretionary power you open yourself not merely to the harm associated with the breach of that trust but you open yourself up to the possibility of that breach as well, which might seem oddly redundant; the possibility of harm is the concern after all. However, I can involve myself in actions that have a necessary harm, like boxing or running a marathon, as well as actions that have a probability of harm, like gambling or skydiving. We become vulnerable to the harm itself when we trust and to the probability of the breach of trust which leads to that harm. But, as Baier points out when she references Hume, we cannot separate the possibility of harm from the chance of good; by trusting others we open ourselves to goods that we could not achieve on our own. The harm that we are in danger of receiving when we trust is a direct result of the autonomy of other individuals; in giving over discretionary powers we allow others to make decisions, decisions which could or could not go in our favor. The decisions of others involve concerns that we knew nothing about when what we hold dear was entrusted to them; in trusting my babysitter I may anticipate that he will invite some friends over, or that my child might receive a burn, but I cannot anticipate all the things that the sitter may do against my interests. We turn over control

[202] Baier allows for these to be cases of trust. We can, on her account, have a general belief in the goodwill of others such that they do not want to poison or kill us; that others are concerned enough about our interests such that they do not try to thwart these goals.
[203] Baier, *Moral Prejudices : Essays on Ethics*, 104.

of those things we care about precisely because we cannot predict all the things that may transpire. If all the possibilities of all the things that could happen were accounted for then we could leave precise instructions about how to respond; we might still have to worry that the sitter was concerned with our interests but we would no longer have to leave that concern to his discretion. But because we cannot know what may come we must leave our children in the care of someone who has the discretion to act as they see fit; we trust that person when we believe that he will use his discretion in a way that furthers our interests because they are our interests. The risk that comes from trusting helps us to understand why trusting is different than relying; I can rely on a sitter to fulfill my specified wants but I cannot rely on him to use his discretion as I would use my discretion. Asking the sitter to use his discretion exactly as I would use mine, even leaving the imagined list of exhaustive instructions actually damages trust because it take from the sitter one of trust's key elements: that the trusted party is responsible for making judgments. When we combine the notion of accepted vulnerability with a belief in the good will of another we differentiate trust from reliance.

Trust without Purpose

One point that is not clear strictly from the discussion so far is that Baier feels that trust, "as an intentional mental phenomenon, need not be purposive."[204] Baier's position here is that even when we become aware that we are trusting we need not have some end in mind by continuing to trust; our trusting is not goal oriented. Further consideration of Baier's position shows that she is concerned with trusting as something different than entrusting; "If we try to distinguish different forms of trust by the different valued goods

[204] Ibid., 100.

we confidently allow another to have some control over, we are following Locke in analyzing trusting on the model of *en*trusting."[205] Clearly, there is at least some distinction between trusting and entrusting if Baier is willing to claim that there is a model of entrusting which can be appealed to. This position leads her to point out that there is more than the initial question about "who" in relation to trusting but also the question of "what" in relation to the object of that trusting. In fact, Baier suggests investigating trust as a "three-part predicate (A trusts B with valued thing C)"; a position which coincides with the claims of Russell Hardin, a position that is discussed in the next chapter.[206] We may, given Baier's discussion of reliance, be willing to entrust, even when we do not trust.

Discretion

Trust, as this chapter has shown, requires judgment and goodwill for both trusting and trustworthiness. Sound judgment is necessary for trustworthiness because in order to truly exercise a goodwill for another judgment is necessary to know what is in his best interests. But judgment is also necessary for trusting as well; the judgment to recognize whose will is good and who is merely pantomiming goodwill in order to curry favor. Goodwill, rather than being subordinate to judgment, is necessary for recognizing the judgment of the other. Sound judgment without goodwill will not lead one to be trustworthy; simply recognizing what is in my best interests will not always be enough to motivate you to pursue that interest in my stead. But goodwill is also required in order to hold the belief that the other is concerned for my best interests in the first place; since our

[205] Ibid., 101.
[206] Ibid.

motivations are inevitably obscured the only way for you to interpret me as being concerned for your interests is for you to think positively about me. Further, since handing over discretionary control can actually cause me harm, to be trusting requires that you act in my interests as well; trusting requires goodwill. The need for goodwill was revealed when behaviors and actions were found to be insufficient for generating trust; our motivations are what make us trustworthy not the actions we engage in.

In the following chapter we consider a problem of a different sort. Does trust consist in a two-part or a three-part relation? By far the majority of authors claim that trust requires a subject, and object, and a direct object. I, on the other hand, argue that, given what has been revealed so far, trust is really only a two-part relation; all of this concern for a three-part relation is mistaking entrusting for trusting.

Chapter Five: Trust as a Two-Part Relationship

The discussion of trust has been dominated by the language of economics. My previous arguments have shown that standard conceptions of trust, conceptions that result from the dominance of the economic perspective, are inadequate to explain the experience of trust. I begin this chapter by showing that it is the economic perspective that leads us away from a viable alternative explanation. This is necessary because the economic perspective, one in which costs are weighed against benefits, precludes seeing trusting as consisting only in two parts, A trusts B. Once the two-part trust relationship is established I consider the effect this has on our conception.

David Messick and Roderick Kramer defend a conception of trust that treats it as a decision in their article "Trust as a Form of Shallow Morality." After giving this definition they claim that;

> Some economic features of this definition need to be highlighted. Trusting is exploitable and potentially costly. In all cases, if the target of your trust is not trustworthy, you will suffer. If, on the other hand, the target is trustworthy, both parties are better off.[207]

Their position assumes rather than defends the economic perspective, which is that the economic features of trust need to be identified, rather than identifying the trust features of economics. This is a bit off-putting because it takes as a given that trust is the kind of

[207] David M. Messick and Roderick Moreland Kramer, "Trust as a Form of Shallow Morality," in *Trust in Society*, ed. Karen S. Cook, Russell Hardin, and Margaret Levi (New York: Russell Sage Foundation, 2001`), 92.

thing for which economics is or can be explanatory. The language of their discussion about trust is couched entirely in the language of economics and this goes beyond even those terms used simply in claiming that trust has economic features. For them, trust is, like a natural resource, exploitable; it can be taken advantage of and used up. Trust is, also, potentially costly; it can end up putting us into a kind of debt – financial, emotional, or physical. Speaking in terms of costs is so accepted and so common that it goes unnoticed and yet speaking in these terms is the most obvious example of just how much the language of economics has come to influence our understanding about trust. Trust is often described as being an investment or as being a form of social capital; for example, Victor Nee and Jimy Sanders intertwine these two descriptions when they say of the trust that is evident in ethnic-based relationships that for many immigrants "this form of social capital is often more important than human and financial capital."[208] Nee and Sanders, by conceiving of trust as a form of social capital, explicitly treat trust as economic in nature but so too do Messick and Kramer without themselves adopting that position.

Iris Marion Young, in her article "Displacing the Distributive Paradigm," is concerned with the nature of justice. She recommends that we move beyond the language of economics and consider the concept of justice from another perspective, from the perspective of oppression and domination.[209] She worries that focusing on the distribution of goods, which is a decidedly economic perspective, obscures a number of important cases of injustice. Discussions about trust are often thought to escape the distributive paradigm that Young finds so problematic because trustworthiness, rather than being

[208] Victor Nee and Jimy Sanders, "Trust in Ethnic Ties: Social Capital and Immigrants," in *Trust in Society*, ed. Karen S. Cook, Russell Hardin, and Margaret Levi (New York: Russell Sage Foundation, 2001), 374.
[209] Iris Marion Young, "Displacing the Distributive Paradigm," in *Ethics in Practice : An Anthology*, ed. Hugh LaFollette (Malden, MA: Blackwell Pub., 2007).

concerned with some distribution of rights, begins with care and concern for the well-being of another. On Annette Baier's account being trustworthy is to be concerned with the well-being of the other, to care about him.[210] I argued for trust as a belief in the good will of another in chapter 4.

One might argue, however, that a strict Kantian could, indeed, be perfectly trustworthy without caring at all about the well-being of any particular person. This objection seems misguided for two reasons. First, one would fail to be a strict Kantian if she did not care about the well-being of others; one of the formulations of Kant's categorical imperative is "[a]ct in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means."[211] A strict Kantian would, therefore, violate the categorical imperative if, in dutifully babysitting my child, she was not also concerned with my well-being as an autonomous rational agent. Second, and for my account this is more important, the cold calculation of duty can be achieved by machines as well as people; people can be reliably predictable without being trustworthy. For an individual to be trustworthy she must have the right kind of motivation and a brute reliance on duty will not achieve the interpersonal relationship that is the focus of this investigation. The second response differs from the first in that it does not rely on Kant; if it turns out that a Kantian could be dutiful and treat another as an end without being concerned for that person's well-being my second response would still remain – no Kantians would be worthy of trust.

---

[210] Baier, *Moral Prejudices : Essays on Ethics*.
[211] Immanuel Kant, *Grounding for the Metaphysics of Morals*, trans. James W. Ellington, 3rd ed. (Indianapolis: Hackett Pub. Co., 1993).

Even if we accept the importance of good will to trust, however, we will not have fully accounted for trust. If the language used to describe and explain trust remains the same, if we continue to talk about trust in economic terms, then we will never move beyond our current focus on distribution of goods in society, a focus that dictates that trust be considered as part of a cost-benefit analysis. So, even with the addition of good will, the economic model forces us to consider whether a belief in the good will of another outweighs the possibility of defection by that other. This view is deeply problematic for any account of trust, it presents a chasm that seems impossible to cross.

The position that trust involves a cost-benefit analysis, by starting out with the assumption that it has economic features, falls easily into game-style models of trust (which is, incidentally, exactly where Messick and Kramer are leading) because judgments of appropriate behavior within the models is made by reference to costs and benefits. The prisoner's dilemma, for example, stipulates that the trusted party can make his situation better by defecting while both parties will be made better off by cooperating – this is simply the definition of the dilemma. The idea behind the game is that there are trade-offs associated with trusting that are tabulated and calculated when considering whom to trust. But, what reason is there to accept this as an accurate explanation of trust? The work which I have presented in previous chapters has challenged many of the ideas underlying the cost-benefit schema such that it should already be in question. The present chapter will continue in this vein by looking at the structure of the trust relationship in order to show that it is inappropriate to consider trust economically.

Trust surely puts one in the position of being vulnerable but the fact that we are vulnerable does not require an economic interpretation. The issue is not whether there are

costs and benefits to trusting but whether a consideration of those costs is part of the process of coming to trust. Certainly, the costs and the benefits are important in deciding whether or not to *entrust* but entrusting has already been revealed to be separate from trusting. The costs and benefits of any act of entrusting are not relevant to the process of trusting which may, or may not, lead to he act. I take it as obvious, if often overlooked, that "[c]ooperation can, but need not result from trust," which means that the reasons I have for undertaking an action, conscious or otherwise, are not themselves reasons for trusting.[212] On the contrary, I argue that trusting is a reason, but not the only reason, for undertaking an act of entrusting. Trusting, or distrusting, helps to weigh the costs and the benefits of entrusting.

The economic perspective, if it is to be successful, must consider both the costs of trusting and being trustworthy as well as the benefits of trusting and being trustworthy (not merely the costs of one and the benefits of the other). Messick and Kramer, in elucidating their definition, claim that one will necessarily be harmed by trusting the untrustworthy; as they state in the earlier block quotation "[i]n all cases, if the target of your trust is untrustworthy, you will suffer."[213] It is worth noting that Messick and Kramer are not suggesting that, generally speaking, we will be harmed by trusting the untrustworthy but are instead explicitly stating that **in all cases** you will be harmed by trusting the untrustworthy. They also claim that if you trust the trustworthy then you will necessarily benefit. At the same time, they suggest that the trustworthy person benefits for having been trusted;"[i]f…the target is trustworthy, both parties are better off."[214]

[212] Claire A. Hill and Erin A. O'Hara, *A Cognitive Theory of Trust* (SSRN, 2005), 6.
[213] Messick and Kramer, "Trust as a Form of Shallow Morality," 92.
[214] Ibid.

None of these claims is true or sustainable in light of experimental evidence and our common experience.[215] Untrustworthy people are not untrustworthy in each and every encounter. The Prisoner's Dilemma succeeds in showing the failure of choosing to always defect; in order to be successful as a strategy it requires a partner that will always cooperate, which is incredibly unlikely for a known defector to find.[216] Messick and Kramer's position, by holding to an "in all cases" account, takes the action of a breach of trust to be the mark of untrustworthiness; taking the act of breaking a trust to be untrustworthiness is the only way that trusting the untrustworthy can *always* be worse for the trusting party, but we have already seen that it is the motivation behind the action and not the action itself which is important to determining trustworthiness; it is not the breach that is important but the motivation, or lack thereof, behind the breach.

In chapters three and four I argued that the Prisoner's Dilemma and other similar games are incapable of differentiating the motivations for trust. Messick and Kramer's position, therefore, should already be in doubt. In this chapter, however, I want to suggest something slightly different, something that returns to the distinctions made in chapter two: trusting an untrustworthy person does **not** always lead to suffering and trusting a trustworthy person does **not** always lead both parties to be better off. That is, a person's status as being trustworthy or untrustworthy does not necessitate success or failure in any particular instance of being trusted. The stipulation of a game like the Prisoner's Dilemma, knowingly or unknowingly, assumes that trusting is an action. It does this because it forces us to consider trusting as entrusting. Further, the stipulation assumes

[215] One of the lessons that we learn from Robert Axelrod's study of cooperative behavior is not just that behaving altruistically can be a survival trait but that there are long term benefits for forgiving (trusting) those who have previously turned against us.
[216] For a more extensive argument making this point see Chapters 4 and 3.

that trust is a three-part relation, where A trusts B in regards to some thing C: prisoner A cooperates with prisoner B in regards to confessing to the police by not snitching. For example, Jack Barbalet describes a situation of trust as on in which "either A cooperates with B to achieve C, or A cannot have C."[217] Michael Carolan says, while discussing the effect of recent cases of Mad Cow disease, that "if you do not trust that person's knowledge on this subject to be true…then you will likely not perceive beef within the United States as representing any particular risk to your health," which we can understand as a case you trusting person X in regards to safety of eating American beef.[218] Paul Falkner gives us another example of the three-part conception of trust relationships; in discussing the acceptance of testimony he says that "an audience's reason can be simply that the audience trusts the speaker for the truth."[219] For Faulkner the ABC relationship is the (A) audience trusts the (B) speaker with regards to the (C) truth. Kevin Gibson, in talking about what we can learn from the Prisoner's Dilemma, suggests that "if one has an agreement with the other prisoner to keep quiet and trusts her to do so, it is prudent (at least in a single play-game) to betray her."[220]Gibson's claim is that an agreement between prisoner's requires (A) one prisoner to trust the (B) other prisoner with regards to (C) keeping quiet. Madeleine Hayenhjelm describes trusting as a "belief about the trusted person's care for X."[221] Pamela Hieronymi suggests that it is appropriate to say that you do *not* trust a person "to the extent that you lack confidence in

---

[217] Jack Barbalet, "Trust and Uncertainty: The Emotional Basis of Rationality," in *SCARR Network Trust Conference* (London: 2005), 5.
[218] Michael S. Carolan, "Risk, Trust and 'the Beyond' of the Environment: A Brief Look at the Recent Case of Mad Cow Disease in the United States," *Environmental Values* 15, no. 2 (2006): 235-36.
[219] Paul Faulkner, "On Telling and Trusting," *Mind* 116, no. 464 (2007): 876.
[220] Kevin Gibson, "Games Students Play: Incorporating the Prisoner's Dilemma in Teaching Business Ethics," *Journal of Business Ethics* 48, no. 1 (2003): 59.
[221] Madeleine Hayenhjelm, "Trusting and Taking Risks : A Philosophical Inquiry" (Royal Institute of Technology, 2007), 22.

a person's trustworthiness in some matter," which is to say that (A) you do not trust (B) in regards to (C) some matter.[222] Claire Hill and Erin O'Hara claim that people are trusting to different degrees in different circumstances and they use roadside fruit stands in order to make their point; "[a] jar of jelly can be easily swiped, but the cash box typically has a very narrow slit that prevents the money from being taken."[223] In their example, (A) the farmer trusts (B) the public in regards to (C) the jelly but not in regards to (C) the cash box. In all of these cases trust relationships are defined as consisting in three parts. Even Cristiano Castelfranchi and Rino Falcone, who criticizes the prisoner's dilemma, treats trust as consisting in a three-part relationship; "one trusts another only relatively to a goal."[224] Michael Bacharach's position will help reconnect this with my claim about PD; he reminds us that it "is common to regard the cooperative choice by the first mover in a 'staggered' Prisoner's Dilemma as a case of trusting"[225] Therefore, the trust of Prisoner A in the Dilemma is exemplified by the act of cooperating with Prisoner B by not confessing, which is a case of Prisoner A entrusting the length of her sentence to Prisoner B. But, trusting is not the same as entrusting and, I argue, trust does not require a third part. The base stipulation of these games therefore fails to capture the nature of trust and our relationship to others.

This chapter is devoted to showing that trust is a two-part rather than a three part relation; it involves a subject and an object, *A trusts B*, but does not require the object of a prepositional clause, *A trusts B with C*; that is, trust relationships are not defined by the

---

[222] Pamela Hieronymi, "The Reasons of Trust," *Australasian Journal of Philosophy* 86, no. 2 (2008): 214.
[223] Hill and O'Hara, *A Cognitive Theory of Trust*, 4.
[224] Cristiano Castelfranchi and Rino Falcone, "Social Trust: A Cognitive Approach," in *Trust and Deception in Virtual Societies*, ed. Cristiano Castelfranchi and Yao-Hua Tan (Dordrecht ; Boston: Kluwer Academic Publishers, 2001), 2.
[225] Michael Bacharach, Gerardo Guerra, and Daniel John Zizzo, "The Self-Fulfilling Property of Trust: An Experimental Study," *Theory and Decision* 63, no. 4 (2007): 352.

actions that would constitute a breach of that relationship. Trust as a three-part relation is a position held by nearly all other authors on trust, Robert Solomon and Fernando Flores being an obvious exception where even Annette Baier, a strong proponent of trust as goodwill, follows suit. In order to defend trust as a two-part relation it is important to first understand the arguments for trust as a three-part relation. I begin, therefore, by looking at Russell Hardin's account of trust. I follow this with a criticism of the three-part understanding of trust which also begins to present a positive account of a two-part relation. I then turn to Solomon and Flores in order to help show the benefits of understanding trust in this way.

Hardin

The crux of Hardin's opposition to conflating a justification of trusting with a justification of trustworthiness is that "[t]he explanation of your actually trusting in some context will be simply an epistemological problem, evidentiary matter – not a motivational problem."[226] The explanation of your being trustworthy, on the other hand, *is* a problem of motivation; what compels you to behave in a trustworthy manner. Because Hardin sees being trusting and being trustworthy as functionally different enterprises a single justification cannot suffice for the two. But, more generally, we can see that this sort of claim is true even if we do not yet accept Hardin's position on what trust is.

Hardin claims that "trust is a three-part relation that restricts any claim of trust to particular parties and to particular matters."[227] This claim directly relates to the problems

[226] Hardin, *Trust and Trustworthiness*, 31.
[227] Ibid., 7.

Hardin sees in the work of others, namely the use of trust as two-part relation between either the truster and the trustee or the truster and the object that is trusted. Hardin differentiates two-part trust relations by separating the object of trust and the object which the trust is about; I may trust some individual with any object or I may trust some object with any individual. The essential difference between these two aspects is that one describes a very high threshold for trust while the other describes a very low threshold. Hardin's criticism of two-part trust relations will be covered in greater detail later in this chapter but, put simply, he feels that two-part relations fail to account for the complexity of trusting. For Hardin, I cannot know if Chunk is concerned for my interests if I do not consider what my interests are in relation to. That is, only a three-part relation takes into account how trust varies with the context of the relationship. I can readily trust my employees to pick up checks from our customers when my employees recognize that their salary is dependent on the income received by those checks, but I cannot so easily trust them to speak well of me when I am not around.

The recognition of the incentives that others have to behave in certain ways is essential to Hardin's conception of trust; "[t]he encapsulated-interest account of trust holds that the trusted encapsulates the interest of the truster and therefore has incentive to be trustworthy in fulfilling the truster's trust."[228] It is the motivation that the trusted party has to complete the expected task that justifies whether or not we trust them. In this section I discuss Hardin's conception of the trust relationship in order to present a full-fledged account of the three-part relationship and the challenge it poses to the two-part trust relationship.

[228] Ibid., 24.

A Matter of Belief

According to Hardin "I trust you because I think it is in your interest to take my interest in the relevant matter seriously."[229] Merely taking my interests seriously is not obviously enough to generate a strong incentive to be trusting; you can seriously consider my interest without necessarily having those interests be part of the final decision criteria. So, Hardin revises and clarifies his position by claiming that "I trust you because your interest encapsulates mine."[230] Trusting is therefore, on Hardin's account, to believe about another that my interests are contained within her interests; it is to believe that another will take your interests into account while considering her interests about some particular issue. It is not that all of my interests are encapsulated in her interests but that, in regards to some finite set of interests, she takes my interests to be relevant in her decision-making process. We can draw a number of conclusions and implications from his definition.

We can see that, on this account, trust is a matter of belief. He claims that when I am trusting of you it is because I have particular belief about you; "[y]ou trust someone if you believe it will be in her interest to be trustworthy."[231] Hardin's account, therefore, is expressly focused on what I take you to be concerned with, what things you take to be within your interests. For him trust is something that I believe, which, Hardin argues, makes his account epistemological. This conception of trust conceives of trusting as a kind of subjective state because it is not concerned with an objective view of your trustworthiness, it is about whether I *believe* that you are concerned with my interests. To

[229] Ibid., 1.
[230] Ibid., 3.
[231] Ibid., 13.

trust someone is believe that some fact about that person is true; the fact in question is that she will hold your interests in her interests. On Hardin's account then, when I am trusting of you I believe that you hold my interests within your interests. Interestingly, conceiving of trust as a piece of knowledge has the effect of denying that the trust is an action. We have knowledge about actions but that knowledge is not itself an action. When I am running, I may know that I am running but the knowledge that I am running is not the running. Knowing that *you* are running does not mean that *I* am running. Further, knowing how to run is also not running. That I am running may indicate that I know how to run but we must be careful not to mistake knowledge of the action and the action itself for being the same thing. Trusting you is to know something about you, not to do something about you.

We can also see that from Hardin's perspective trusting is justified by being trustworthy, trusting is *not* justified independently of being trustworthy. He argues that it is because I believe that your interests encapsulate my interests that I trust you. My belief that you have incentive to consider my interests is what gives me reason for believing that you *will* consider my interest. There is a delicate distinction to be made here. Trusting is not justified by any justifications for trustworthiness; we cannot appeal to a moral justification for trustworthiness in our explanation of trustingness because that will conflate different explanations. Instead the justification of my trust in you is that I believe you are or will be trustworthy, that you do or will have reason for considering my interests. For Hardin, then, the existence of the conditions for trustworthiness is a prerequisite for any trusting.

On Hardin's account, trusting follows from expectation; "my trust of you must be grounded in expectations that are particular to you."[232] When I trust you it is because I have formed the expectation, or a set of related expectations, that you will take my interests into account in regards to some relevant matter, which means that I can trust you to behave or respond in a manner that is compatible with my interests. For Hardin trusting is not simply expecting you to complete some particular action because, for him, expectations that I have about you involve your motivations and commitments. The expectation that I have formed about you specifically relates to your incentives to behave in a way that furthers my interests.

Hardin's account also leads to the idea that trusting is a capacity that we have to judge the commitments of others; my belief that you will hold my interests in your interests is a judgment about what things you hold in your interests. Hardin argues from this that trusting is not a disposition; having a disposition is to have a tendency to behave a certain way under a set of conditions but trust, as knowledge, is about what we can know about others. Trusting is to *know* something about the motivations of another but knowing is not a disposition. Hardin admits that we might say that there is a disposition to trust "as though to say I have a disposition to know certain kinds of things"[233] but dispositions do not obviously involve what we can know.

It follows from Hardin's account that trust must be learned. This might seem obvious – all knowledge must be learned and so trust, because it is a form of knowledge, must also be learned. His account here is two fold. We must learn that someone is worthy of trust but there is another sense in which trust must be learned. As an ability, trusting is

[232] Ibid.
[233] Ibid., 37.

not something that is born in us We are born with the capacity to run but not the ability, we are born with the capacity to speak but not the ability, and we are born with the capacity to trust but not the ability; we must learn to trust others. What we learn is not simply a matter of whom to trust, who is worthy of trust, but to be trusting at all. Because we must learn to be trusting some individuals will learn to be more or less trusting than others; trusting will come more easily for some and with more difficulty for others. We all have the capacity to trust but we develop different abilities to trust, which is why it seems as if there is there are dispositions to trust and be trustworthy.

We also discover from Hardin's account that trust is not something that we choose; we either trust or not but we do not choose it. We do not look at the evidence in front of us and then choose to know or not know what the evidence teaches us. Our ability to trust determines what we can realize about others and what it takes for us to trust them. We do not make the conscious decision to trust others based on what we know of them; one of the things that we know is that we can or cannot trust them.

Explaining an Epistemological Account

We have, at this point a good idea of what Hardin believes trust to be but the explanation of an epistemological account needs further clarification. We know, for instance that Hardin claims that "[t]rusting is merely a bit of knowledge."[234] We know that he means that trust is a specific kind of belief that we have about the commitments of others. But it is not yet clear how this account succeeds as an epistemological account.

[234] Russell Hardin, "Conceptions and Explanations of Trust," in *Trust in Society*, ed. Karen S. Cook, Russell Hardin, and Margaret Levi (New York: Russell Sage Foundation, 2005), 10.

Hardin calls a "street-level epistemology."[235] He is not concerned with justified true belief but rather what it takes to come to have a belief; it is not a matter of the content of the belief but instead of subjective worth of that belief. This street-level or folk epistemology is an economic account; he explicitly compares the investment cost of trusting to the value of the knowledge. This sort of perspective lends itself to the learning account of trust described above; that is, to trust someone is to have come, through experience, to have a belief about the motivations of that individual. But Hardin's account is also about our capacity to trust, a capacity that is affected by our past experience. Some individuals who have had numerous prior bad experiences will find the costs of trusting to be much higher and so will not easily develop the kind of knowledge about others that those who have had prior good experiences will find it easy to come to.

Hardin's encapsulated interest account of trust is meant both as a definitional account and as an explanatory account. He eliminates certain conceptions of trust thereby providing a basis for a positive account of trusting. He also defines the scope of things for which we can consider something to be trusting and so presents a definition that can be used to identify trust. He further explains how our trust influences our behavior by connecting the beliefs that we have to incentives, both our own and the incentives of others.

Hardin maintains that trust is a three-part relation consisting in one who trusts, one who is trusted, and the thing with which the trust is concerned. Hardin claims that many other accounts present trust as a two-part relation consisting in one who trusts and either one who is trusted or the thing with which the trust is concerned. The account

[235] Hardin, *Trust and Trustworthiness*, 115.

presented by Annette Baier, with its appeal to good will, clearly seems to be one of the accounts that Hardin has in mind. However Baier, like Hardin, explicitly admits to considering the trust relationship as one that consists in three parts when she asks "what good is it that they are in a position to take from you, or to injure?"[236] It might seem therefore that the distinction between Hardin's account and the accounts of others is insignificant but the formulation of the trust relationship has a tremendous effect on our interpretation of trust. Hardin points out that there are two ways to formulate the two-part relation: truster/trustee and truster/trusted thing. The truster/trustee two-part conception is one in which one person is willing to trust another in regards to any thing. The truster/trusted thing two-part relationship is one in which one person is willing to trust some thing to any person. Hardin challenges both formulations of the two-part relationship and although they share much in common their differences are important.

Trust as a two-part relation consisting only in the truster and the trustee is not so much an untenable view as a contentious one. Hardin does not deny the possibility of an open-ended, two-part trust relationship but claims that it is rare at best, which would make a two-part relation not the common experience and therefore not particularly interesting to a discussion of trust. More importantly, he argues that it is often "smuggled into discussions."[237] He maintains that the vernacular usage of trust, when used to describe a relation lacking in the object of trust, simply omits the third piece of the relation; when I say 'I trust her' I mean to say that 'I trust her with θ'. Thus, the two-part trust construction only loosely describes a two-part relation when in reality the third portion is left unspoken. That is, we are mistaken when we claim to generically trust

[236] Baier, *Moral Prejudices : Essays on Ethics*, 101.
[237] Hardin, *Trust and Trustworthiness*, 60.

person X; instead we trust person X with regard to some thing θ and we simply neglect to mention that thing.

Hardin's position is that "we trust only certain others with respect to some things, and maybe an even more inclusive set of people with respect to somewhat less demanding things."[238] He admits that it may be possible in very close relationships to trust another in regards to all things, but he feels that in the vast majority of cases this claim is elliptical and that the third part of the relation is simply omitted. That this type of trust is so uncommon is reason enough to be wary of any account that attempts to use this kind of two-part relation as a basis for explaining trust; it will not account for the majority of all cases of trusting. Hardin's necessity for a third part of the relation need not be overly demanding however, the third part of his three-part relation does not have to be a specific thing; when I trust you with θ that need not be some particular object or action, although it can be. Rather the third part can refer to a set of things, like those things that do not threaten my life or those things that refer to single engine airplanes; the third part of the relation can be a very large set or it can be fairly small.

Hardin's criticism of the truster/trustee relation is that it claims too much; to say that I trust you, under this two-part construction, is to claim that I trust you in regards to all things. Hardin argues that to honestly say 'I trust you' means that 'I trust you with everything.' While this might be a romantic conception of trust, it is surely too strong. One reason for thinking that it is too strong is that, if correct, we could not say that we trusted someone if we failed to trust them in regards to even one thing. That is, if I did not trust you to carry me over Niagara Falls on a tight-rope then I could not say that I

[238] Ibid.

trusted you at all, even tough I might trust you with regards to everything else. Another reason Hardin argues that this claim overreaches is that in order to have two-part trust we must exceed that which with we can justifiably trust of another. We cannot have the experience or knowledge necessary to justify trusting another with such an open set of future conditions; it would mean that we would have to blindly allow another to care for our interests. According to Hardin's conception of trust, we could not consider this as trust because we could not expect another to maintain concern for our interest in the face of a near infinite set of possible alternatives. It is simply impossible to suppose that you should hold my interests in your interests under all conceivable future conditions; it would be impossible for your interests to ever trump my interests under those conditions. This would entail that no interests were encapsulated because there would be no space around my interests for your interests. This relationship would more accurately be described as a sharing of interests. Instead of trusting you to care for my interests I would believe that you and I had the same interests; your care for these interests would be a concern for your interests not a concern for my interests.

Unlimited trust in another is not the only form of a two-part trust relation. Hardin calls the willingness to trust some θ to any person X "social trust." Social trust is to trust some object to the care of an undeclared individual. This is similar to the open-ended trust of "I trust you" but here the variable is in a different position. Again, Hardin's criticism is that this kind of trust admits too much. However, unlike the criticism of truster and trustee, this rebuttal is concerned with the extent to which we are willing to allow others to care for our particular interests. If this position has merit then we would have to be willing to trust any person in regards to some thing θ, which will seem strange.

To be able to trust some particular thing θ to the care of any person X requires that we believe that every person will hold our interests in their interests or, at least, that any person would consider our interests in regards to this thing. This account moves trust away from something for which there can be justification and moves it closer to something for which there can be no justification, something more akin to blind faith. Blind faith, however, is not trust.

On Hardin's account a two-part trust relation fails to count as trusting at all. Neither the two-part relation between truster and trustee nor the two-part relation between truster and trusted thing presents a coherent picture of the trust relationship. Both two-part relations are far too open-ended and far too demanding. Hardin concludes therefore that the two-part trust relation is a conception of trust that should be avoided.

A Dispositional Account

Hardin's account, with its focus on what we know about others, can appear to be very similar to a dispositional account. The amount and kind of evidence that we are willing to accept is affected by our past experiences. Some people have had a number of positive experiences in regards to trusting and so have a lower evidentiary threshold. Hardin describes this as a "learning account." Although this account is similar to a dispositional one he denies that dispositions can accommodate the variety of trust relationships that we experience. He uses *The Ring of the Nibelung*, which I discussed in chapter 4, as his primary example of how dispositions fail to make sense of

trustworthiness. Hopping on the band wagon we might say that dispositions, like character traits, inhabit "philosophically shaky ground."[239]

In Wagner's opera the character Alberich forswears love in order to gain great power. What is of interest to Hardin in this example is the possibility that an individual could make such a dramatic claim and that he could hold to it. Hardin's position is that, given how we understand dispositions, our motivations are not that reliably stable. The problem with so-called 'bald dispositions' is in being able to follow through on past commitments that lack the same present motivations.[240] Competing motivations can, in the present, pull us away from those commitments that we previously had held; Hardin argues that "when the time comes to live up to our assertions, we may face forcefully contrary motivations."[241] Once our interests change, it should be no surprise that what we apply our effort to completing will change also. We cannot rely on our dispositions in order to behave a certain way because our dispositions can be overridden by present concerns. Simply put, Hardin's position is that a disposition to act cannot make sense of all the ways that our motivations and incentives might change at some future point. Our dispositions, at best, are able to describe how we are behaving now, not how we will behave in the future. For Hardin this idea is simplified, "we cannot tie our future and present motivations together to yield a single 'net' motivation" which is what he argues a disposition must be.[242]

---

[239] John M. Doris, *Lack of Character : Personality and Moral Behavior* (Cambridge, U.K. ; New York: Cambridge University Press, 2002), 16.
[240] Hardin uses 'bald disposition' to mean a commitment that is made purely by fiat of the will without another motivating force.
[241] Hardin, *Trust and Trustworthiness*, 36.
[242] Ibid.

Hardin's conception of a disposition is an "ability to make commitments for future actions."[243] His discussion is about the commitments that we can make. This conception of a disposition seems strange however and not fully in line with what we might otherwise think a disposition to be; a disposition seems to be the tendency we have to behave a certain way and not the capacity we have to commit ourselves to behaving a certain way. In fact, if we reinsert "disposition" back into Doris' conception we get a very different conception than the one presented by Hardin: *if a person possesses a* disposition*, that person will exhibit* disposition-*relevant behavior in* disposition-*relevant eliciting conditions.*[244] Future conditions may be such that they should count as relevant eliciting conditions for relevant behavior but the ability to make commitments for future actions is not itself a description of a disposition under those conditions.

Hardin's concern however is with our ability to impose a disposition, seemingly from out of nowhere, on ourselves. His point is well taken. We, like Alberich, would be hard pressed to, because of a present concern, adopt some pattern of behaviors that would be recognized as a disposition. This would require us to, in an instant, adopt as part of our motivational structure a concern that is only presented to us in that instant; our present motivation cannot be sustained in the face of all possible future concerns. Maintaining a pattern of behaviors, one that would be recognized as a disposition, in the face of contrary future concerns would be difficult, if not impossible, were we not already able to keep commitments of that sort; "[o]ur strongest commitments are often merely those that that are clearly backed by our interests."[245] Alberich forswore love in order to gain the

[243] Ibid., 32.
[244] Doris, *Lack of Character : Personality and Moral Behavior*, 16.
[245] Hardin, "Trustworthiness," 27.

Rhinegold but by what power is now withheld from loving? Hardin posits that perhaps there was some magical power that stopped Alberich from loving and suggests that we should "contrive real-world substitutes for such magical powers."[246] Keeping commitments and our motive for doing so is Hardin's concern, which therefore leads him to argue that there can be no disposition to trust.

Hardin does however go on to present an account of dispositions that is about something other than our ability to make commitments. He admits that his conception of how we learn to trust resembles a dispositional account of trusting.[247] His learning account of trust denies that there are people who "just trust." Through positive and negative reinforcement we learn how to act and we adopt various kinds of behavioral tendencies. Being trustworthy does not seem to Hardin to be the kind of action that can really harm us; "one can be disposed to trustworthiness without considerable risk."[248] The reason that we can appear to be disposed to be trustworthy is that the positive effect of engaging in cooperative relationships reinforces the willingness to cooperate; behaving in a trustworthy manner is not likely to cause us harm and, further, is likely to bring some benefit, which will encourage acting in that manner. Being trusting on the other hand, does not seem to be sustainable as a disposition because our willingness to trust others can get us into a significant amount of trouble; "if I am always trusting, I will be the gullible target of increasingly many people."[249] Being taken advantage of will quickly teach an individual that trusting is a behavior that should be avoided. So, the disposition to trust, if it exists in the first place, should quickly fade.

---

[246] Ibid.: 28.
[247] ———, *Trust and Trustworthiness*, 36.
[248] Ibid., 37.
[249] Ibid.

Hardin's conception of the disposition to trust can be exemplified in his discussion of the relationship between Karamazov and the merchant Trifonov in *The Brothers Karamazov*. In this relationship Karamazov is placing a great deal of trust in Trifonov. If we suppose that Karamazov had a disposition to behave this way then he would be inclined to trust the merchant even when he lacked sufficient evidence to suppose that the merchant was worthy of trust. A disposition to be trusting can cause a great deal of harm in cases like these, especially if the proclivities of someone like Lieutenant Karamazov are widely known. Under these circumstances, he would be readily taken advantage of by charlatans and confidence men. His tendency to trust others would lead to his trust being let down more often than it is fulfilled. Karamazov was so damaged by the interaction with Trifonov that had he known the full extent of the repercussions prior to his involvement he would not have participated. He should very quickly then learn to no longer be disposed to trust.

A disposition to be trustworthy on the other hand carries little chance of the kind of harm associated with trusting. A disposition to be trustworthy, even when widely known, only brings about more successful interactions with others. The merchant Trifonov benefited from his cooperation with the lieutenant colonel but the merchant also benefited from his breach of trust. Trifonov would have participated in the relationship with Karamazov even knowing what the full costs would be. Trifonov was rewarded for being trustworthy as well as being untrustworthy. Karamazov was rewarded initially for trusting but was eventually punished for it as well. A disposition to be trustworthy may be possible to maintain in light of experience but a disposition to be trusting should cause

considerable negative reinforcement. Because of this Hardin does not think that a moral disposition to be trusting is feasible.

No Good Reason to Trust

As we saw in a previous chapter, to trust without sufficient knowledge of the individual or her motivations is to make a mistake; there would be no good reason to justify the trust. Having 'no good reason' is a problem for trust but blind trust is a trouble spot for any discussion of trust and so presents no particular problem for Hardin's account. Faith, on the other hand does not have this kind of problem. On one account, faith requires us to act contrary to reason; Richard Taylor describes faith as "the very thing that would come to stand diametrically opposed to faith."[250] On another account, faith requires us to exceed the boundaries for which we have justification. Kierkegaard positively describes this overreaching as "the paradox which cannot be mediated."[251] W.V.O. Quine, somewhat less positively, says of Kierkegaard's position that "one can't believe a thing if one sees that it is impossible."[252] Faith, then, is somewhat different from trust. But the similarity between faith and trust remains unmistakable; acting on trust and acting on faith put us in the position of relying on another for something that is important to us. If faith is acting when we have no, or cannot have any, knowledge of desire, motivation or willingness to act of the other then insofar as we are trusting when we lack complete assurance of the behavior of the other, trust seems to be the same as faith. But even though the two concepts seem very similar they are not quite the same. Faith,

---

[250] Richard Taylor, *Ethics, Faith, and Reason* (Englewood Cliffs, N.J.: Prentice-Hall, 1985), 79.
[251] Søren Kierkegaard, *Fear and Trembling*, trans. Alastair Hannay, Penguin Classics (New York: Penguin Books, 1985), 95.
[252] W. V. Quine and J. S. Ullian, *The Web of Belief*, 2d ed. (New York: Random House, 1978), 60.

especially in its religious context, is something about which can have no knowledge or understanding of the motivations of the actor. It is possible to have knowledge of the motivations of people but it is impossible to have knowledge of the motivations of God. To take something on faith is to admit a complete lack of justification. Trust on the other hand is something that only comes about with knowledge and justification. To say that you trust someone with whom you are not acquainted and with whom you have no knowledge is to misuse the word trust. I cannot trust Chester without having some reason or justification for believing that Chester is worthy of trust. If Beverly trusts Chester and I trust Beverly then I have a reason for trusting Chester, namely my confidence in Beverly's judgment of character. That is, having a belief of this kind is to no longer lack knowledge about Chester. So, even though faith and trust are similar, presenting an account of faith is not to have presented an account of trust.

We might not want to throw the two-part relation out the window however. By conceiving of trust in a case by case basis, we provide a space for entrusting and the three-part trust relation. Trust is more than a response to the question 'what should I do now?'

Three-Part

Trust is often considered to be a good, something that can be exchanged or a kind of social capital. Hardin considers this kind of conception to be misguided as well. He believes that this view is really concerned with trustworthiness and not trusting. Although this criticism is somewhat familiar we should not therefore take it lightly. Trust, as a commodity, would be worthless without there first being some incentive to be

trustworthy. His second criticism of the economic perspective is that it would create free-riders but free-riding on trust makes little sense. When we free-ride we allow others to do the work of investment while we share in the profits. To free-ride on the investment of those who trust would mean that we were not trusting but that gains the free-rider nothing. We can free-ride on trustworthiness but not on trusting.

Hardin presents Dasgupta's account of trust, which, he says, treats trustworthiness as "an incentive-to-be-trustworthy theory."[253] If trust is economic then the demand for trustworthiness should generate trustworthiness in the same way that any demand provides incentive to increase supply. Hardin maintains however that under this account our trusting would reflect our confidence in the systems of enforcement. We would, he believes, "economize on [trust]."[254] We would over and under invest in trust depending on market forces. When the market was right, like when we had good opportunity to acquire trust at low cost we would stock up on it. But when the market was bad, like when we were part of a collective group whose status was not fully within our control, we would limit our investment. All of this leads Hardin to conclude that the commodity at issue is not trustworthiness but rather one's *reputation* for being trustworthy.

Hardin's objection to trust as social capital is slightly more complex. He criticizes discussions of this type as mistaking social relationships for trusting; there is incentive to be trustworthy to those with whom we are in social relationships but it is the social network which provides the foundation necessary for that incentive. If trust were a form of capital then we would be able to use it as buffer, a kind of entrance fee to gain access to new relationships. But trust does not accrue in this way. With a truly new relationship

[253] Hardin, *Trust and Trustworthiness*, 82.
[254] Ibid.

you risk cooperation with the individual for the possibility of benefit to yourself but generally you limit your interaction with that individual precisely because there are no buffers. If the initial interaction is positive then it reinforces the tendency to cooperate which can eventually lead to a relationship of trust. But the trust relationships I have developed with others have no effect on the initial risk I take in cooperating with a new individual unless the new individual is known by those that I already have trust relationships with. That is, the 'capital' that I have generated in my trust relationships does not transfer to relationships beyond my social network. Therefore, it is not some object-like trust that has been kept in a safety deposit box which allows me to interact with those in my extended social network rather it is the network itself that allows contact and interaction. Interactions beyond my social network start with a balance of zero. Trust is not a form of social capital. Hardin's position, however, while rightly criticizing other positions for presenting trust as an economy is itself doing just that. We can see this by considering two of Hardin's mistakes: 2) Hardin conflates cases of trusting with cases of entrusting and 2) in his Karamozov and Alerbich examples he conflates trusting and trustworthiness.

Serving Trust

That dishonesty can serve the end of trust seems, perhaps, to be counter-intuitive but when we consider the role that so-called 'white lies' play in maintaining the amicability of most relationships we see that small amounts of untruth can actually help maintain trust between individuals by allowing individuals to avoid having to question the beliefs, motivations, or commitments of the other. One might argue that we would all

be better served by being completely honest with each other, that these little deceits do actual harm to true trust. However, I point out how dishonesty can serve to encourage trust not to argue that lies are good for trust, rather I do it in order to suggest that tying trustworthiness to trustingness via encapsulated interests makes a breach of trust not only possible but much more plausible; the necessity of being trusting, if it forces trustworthiness into a subservient role, encourages untrustworthy behavior.

That a lie can actually encourage trust is evident in one of Hardin's examples. In *The Brothers Karamazov* Dmitry Karamazov tells us of a lieutenant colonel who is led to believe that his dealings with the merchant Trifonov will be aboveboard. He believes the merchant to be worthy of trust because the merchant has acted in a manner to encourage that belief. But the merchant is not truly concerned with being truthful or trustworthy. Trifonov is interested only in the end that appearing trustworthy can get him - having access to a large sum of money. Trifonov acts in a trustworthy manner but is not actually worthy of trust, which is revealed when he fails to return the money that the colonel left in his care. A lie about the merchant's trustworthiness served the end of having the colonel trust Trifonov. Hardin's position is that Trifonov failed to be trustworthy when he stopped holding the colonel's interests in his interests when Tirfonov ceased to see the value in maintaining a relationship with the lieutenant and refused to return the money that Karamazov had left in his care. The lieutenant colonel trusted Trifonov throughout because he believed that Trifonov held his interests as his own and it seems that he was correct in that belief. All his previous dealings with the merchant were very successful. In each of the previous encounters the lieutenant colonel had good reason to believe that

Trifonov would want to maintain and continue their relationship. It was only at the end that Trifonov could no longer be trusted.

On Hardin's account Lieutenant Karamazov trusts the merchant Trifonov because Trifinov holds Karamazov's interests in his own. The lieutenant is justified in his trusting of the merchant because the merchant does indeed find it in his interest to consider the interests of the lieutenant. Where the lieutenant is wrong however, is in not taking greater precautions against the inevitable result: Trifonov turning on the lieutenant. Hardin's account makes sense of this relationship because encapsulated interests explain how the merchant can be trustworthy at one point and untrustworthy at another. But Hardin makes a critical mistake in considering the merchant to be trustworthy. The merchant was never trustworthy, he merely appeared so. Trifonov reveals his untrustworthiness when he refuses to return the money to Karamazov. Hardin's mistake comes prior to this revelation; he overlooks that the arguments against linking acting on trust with trusting also argue against linking acting as if one is trustworthy and actually being trustworthy. One may behave in a trustworthy manner in order to achieve a set of desired ends without actually being worthy of trust. It is not simply that it turns out that Trifonov is untrustworthy that is at issue. Instead it is the structure of the analysis that is at fault. Hardin shows that it is inappropriate to link acting on trust for trusting itself, to mistake entrusting for trusting. We can act in a manner that appears as if we trust another without having to actually trust the other. That is, I can give over control of my financial assets to you even though I do not believe that you hold my interest in your interests. It might be unwise for me to do this but it is neither impossible nor even uncommon. Yet, even without a disparity between how we act and how we feel there is still reason to believe

that we should not confuse acting on trust with trusting. The act of true trust comes about as a result of actually trusting the other but that act is still not the trusting. We have already see that trusting is more than the acts which come from it, the same is true for trustworthiness.

Trust, when considered as a encapsulated interests, requires a three-part relation but trust when considered as the good will of the other need not be a three-part relation. A belief in the good will of another can be a three-part relation; I may believe that in relation to some thing X you have a good will towards me but a three-part relation is not necessary for trust. For the three-part relation I imagine a couple, Jack and Jeanne, who are divorced. Jeanne no longer trusts Jack to have her best interests in mind; that is, she does not think that in relation to things in general that Jack will look out for her interests *because* they are her interests. However, Jeanne might still believe that, where their children are concerned, Jack wants what is best for her. Jeanne may trust Jack with their children but not with much else. Jeanne can believe that Jack's will is good when he is considering their children; this is surely a three-part relation, Jeanne can entrust her children to Jack. I imagine another couple, Kyle and Kelly who trust each other but unlike Jack and Jeanne, have no limitations on the conditions of their trust; neither Kyle nor Kelly has a thing to which their trust is in relation. Hardin suggested that this two-part trust relation is unrealistic or even impossible because it seems to demand that Kyle and Kelly must trust the other absolutely with regard to all things. That is, on the criticism of the two-part trust relation if it turns out that there is even one thing with regard to which Kyle will not trust Kelly then he cannot claim to trust her at all. Further, this position puts a heavy burden on the individual who trusts; it forces him to consider

all the possible interests that another may have. But these difficulties only appear when we consider trust as a form of knowledge, which requires the perspective of a three-part relation. The criticism, however, is misguided when discussing trust as a belief in the good will. In fact, it appears that the three-part trust relation is a limited form of trust that exists only as a pale relative of a robust, two-part trust.

We saw in the previous section that a three-part conception of trust led to entrusting but a two-part conception, in which I believe that you are concerned with my interests because they are my interests, allows me to belief that you are concerned with my interests regardless of the situation or object, which is necessary if were are to escape a conception of trust that treats it as entrusting. Ned might be a wonderful auto mechanic and friend of Nelly; Nelly can believe that when she brings her car to Ned that he will not only fix it but that he will not rip her off in the process. The objection to a two-part trust relation suggests that because Ned, as an auto mechanic, has no experience or capacity to operate on the human heart it would be foolish of Nelly to trust him to perform her operation. But this criticism mistakes the distinction between trusting and acting on trust that Hardin emphasizes; it treats trusting as Nelly engaging in the activity of letting Ned operate on her, which is entrusting. But, if Nelly really believes that Ned has a good will towards her then she is able to leave Ned in charge of her operation because Ned, being trustworthy, will choose to *not* operate on Nelly. A real concern for Nelly's best interests will lead Ned to want the best possible heart surgeon to perform the operation Further, since goodwill of the type I describe demands good judgment Ned should recognize that it is not in Nelly's best interest to have him, an auto mechanic, operate on her heart. Nelly's belief in Ned's good will towards her is the belief that he will do what he can to

help her when he can help her and that when he cannot help her he will find those who can. To suggest that Nelly cannot trust Ned in regards to her surgery is to suggest that he cannot trust her to make a decision regarding her surgery but the decision-making capacity is exactly what trust as goodwill is concerned with.

Under this conception Kyle can trust Kelly even in situations where Kelly lacks the relevant capacity to succeed at the end which best fulfills Kyle's interests. In fact, Kyle can believe that Kelly is acting with good will towards him even when she proceeds with the successful completion of an activity which Kyle is against. Kyle might, for example, hate surprises.[255] Kelly, on the other hand, might think that a surprise party is a wonderful example of someone's care and concern.[256] If Kyle does not believe that Kelly has a good will towards him then he will view the surprise party as evidence that Kelly will eventually lie and deceive him. Because Kyle lacks a belief in Kelly's good will towards him he will take her actions to be motivated by some nefarious reason. But if Kyle does believe that Kelly has goodwill towards him then he can overcome his own dislike of surprise parties and view the party as further evidence that Kelly really does have a good will towards him. A belief in the good will of the other allows actions that are, in some objective sense, against the interest of the individual to help facilitate trust.[257] Kyle's belief that Kelly has a good will towards him is a belief that Kelly is concerned with Kyle's interests because they are his interests. If Kyle ever starts to think that Kelly is concerned only with his interests in regards to some finite set of actions then

[255] Surprise parties require deception and a willingness to mislead
[256] Surprise parties require a great deal of planning and a great deal of care in organization
[257] If the trust ever fails these perspectival positions are reinterpreted as evidence that the good will was never there, which explains, in part, why rebuilding trust is so difficult; it requires the foundation of understanding to be reset which is a hermeneutical process.

Kyle will cease believing the Kelly has a good will toward him and will instead believe that she holds his interests in her interests because furthering his interests furthers her own. That is, Hardin's position that trust is encapsulated interest forces Kyle to view Kelly as only being concerned with his interests as a means of furthering her own.

We can see the problem with Hardin's encapsulated interest account by considering a popular story more recent than Hardin's example of *The Brothers Karamazov*: *Star Wars*. Near the end of the original film Luke Skywalker accuses Han Solo of "taking care of himself," which is "doing what he is good at."[258] Luke is visibly upset at Han who is in the process of packing up the reward that he was promised for saving Princess Leia so that he can leave. This is happening immediately prior to the start of the final battle. Han was enticed into this adventure by the promise of getting paid; first he was to be paid for transporting Luke and Obi-wan then he was promised monetary reward for saving the Princess. Luke made the second promise to Han, he even suggested it himself. Luke clearly knows that Han is motivated by the prospect of getting paid. He further knows that given the right price Han is willing to be concerned about whatever interest Luke cares about. Luke also knows that they do not share the same belief system; Han says of Luke's interest in The Force that "hokey religions and ancient weapons are no match for a good blaster at your side."[259] Knowing all these things about Han, why does Luke get upset when Han acts to fulfill his own interests? The answer is that Luke trusted Han. Even the recognition that Han has an overriding personal concern, a bounty hanging over his head, does not mitigate his leaving at this time of need. If Hardin's conception of trust is accurate then Luke and the audience watching him is

[258] Lucas, "Star Wars."
[259] Ibid.

making a mistake by being angry at Han. But being upset at Han's departure helps us to realize that Hardin is wrong about the nature of trust. We expected more of Han than simply acting in a mercenary way. We thought, because of his actions, that Han had come to care about Luke and his well-being, that they were friends. We believed that Han was someone who we could trust. His concern for the bounty hanging over his head trumping the interest of others during a time of crisis shows that Han is not worthy of trust; "[t]he ultimate measure of a man is not where he stands in moments of comfort and convenience, but where he stands at times of challenge and controversy."[260]

The revelation that turns us against Han is not that he would leave once he got paid, this is exactly what both he and Luke agreed to, it is that by leaving Han's motivations are exposed as not being care and concern for Luke. Instead Han is revealed to be motivated by his own self-interest and any concern for the interests of others is merely because it is in his own interests to consider the interests of others. Han is eventually redeemed when he returns, at an opportune moment, which reveals that he in fact is motivated by his concern for Luke (because he has come to help without the promise of any reward). This shows what is wrong with Hardin's account of trust; Han fulfills the encapsulated interest account of trust even when he leaves. Because it forces us to view others as acting only instrumentally on our behalf Hardin's position encourages distrust between individuals because.

We can now see that trust as good will and the two-part trust relation that it allows answers Annette Baier's challenge about the good of trust. Baier suggests that "if the enterprise is evil, a produce of poisons, then the trust that improves its workings will also

[260] Martin Luther King, *Strength to Love*, 1st ed. (New York,: Harper & Row, 1963).

be evil."[261] She sees this as a challenge to the idea that trust is always a good thing. What we come to realize by considering a two-part trust relationship that depends on good will is that evil enterprises are not motivated by good will and so are not cases of trust. Baier's conundrum, trust that leads to bad, is simply a mistake. Certainly, Tom and Tabitha, life-long criminals, can depend on each other and Tom can even want to promote some end that Tabitha is striving for but insofar as that end is not in Tabitha's best interest then Tom is not acting in a manner worthy of trust. Tom would not be showing concern for Tabitha's well-being.

Because trusting another requires us to believe that the other not only wants what is best for us but also that they exercise good judgment about those things that are important to us we cannot conclude that trust can lead to bad ends; if it willfully produces bad ends it simply is not trust that is driving it. Baier argues, as do I, that trusting is to give over discretionary control, to allow another to make decisions on our behalf. The trusted individual should then make decisions that will benefit us, not simply decisions that match our desires.

Consider the example of the babysitter. If I leave my children in the care of a babysitter and then, later, call home to find no answer I may become worried and upset. If I ask my babysitter why he failed to answer the phone and he responds "because we were outside" I might even get angry. "I told you that the children were not to go out" is a response that I might have. But, if the babysitter took the children outside because the house had mysteriously filled with gas and he was trying to save my children, then he made a good judgment about how best to look after my interests. If I had purposely

[261] Baier, *Moral Prejudices : Essays on Ethics*, 131.

broken the gas main in order to kill my children, the babysitter would have thwarted my desire but he would not have behaved in an untrustworthy manner. Even if I had told the babysitter of my plans and had promised to reward him handsomely he would not have behaved in an untrustworthy manner by saving my children because, ultimately, it is not in my best interests to be responsible for the death of my children no matter how much I might desire it.

My argument here follows Plato when, in *Apology*, Socrates asks whether a man "would rather be harmed than benefited by his associates."[262] We do not want our associates to harm us and doing evil, even in our name and following our desires, certainly causes us harm. Giving over discretionary control for the completion of some specific act, as opposed to because of the belief in the good will of the other, is a case of entrusting and not a case of trusting. The solution therefore to Baier's problem is that giving over discretionary control in order to fulfill some evil end is entrusting and entrusting that is put to bad ends, even though it appears to be driven by trust, is not trust. People are often wrong about what things are in their best interests; what I want and what is best for me are not always the same thing. Trusting is about recognizing those who have good judgment about what is best for me, not simply those who can get me what I want. This is why we can trust our friend who refuses to buy us drugs and not trust, yet still hang out with, the friend who gets us whatever drug we request. The latter might be a "good guy" or even someone that you can turn to in a pinch but he is looking out for your desires and not necessarily your best interests.

[262] Plato, Grube, and Cooper, *The Trial and Death of Socrates*, 29.

There might be some resistance to this idea, especially if we consider certain assassination plots; did the members of Operation Walküre, the plot to assassinate Adolph Hitler, trust each other? They very well might have. I am even inclined to say that they must have given the dangers that they faced. But, taking the life of another by assassination can only be considered a good act when we consider the greater good. That is, it is a decidedly consequentialist position to suggest that a plot to kill someone is morally correct, even when the person in question is Hitler. But, more importantly, a plot, whatever its end, is necessarily focused on some particular end and we cannot trust in order to fulfill an end; our trust may allow us to fulfill an end but it cannot come for the purpose of generating that end. That is, insofar as I participate in a plot I am not trusting the members of that plot. My trusting those people may allow me to begin to plot with them, in fact it seems to require it, but I am not trusting the completion of the plot to any of them, I am *entrusting* the completion of the plot to them. It is not the activity of the plot that is the issue but whether the individual in question will look after our best interests even, and perhaps especially, when we are wrong or ignorant of our best interests.

Three-part Expectation

Trusting is often seen as having formed expectations about the behavior of another but this conception is mistaken; it combines that which is separate. That is, this conception considers trusting and expectation to be the same where the experiences are separate. My contention is that trust allows us to form expectations about others.

A conception of trust that envisions a three-part relation leads us to conceive of trusting as expectation. If A trusts B in regard to some thing C then A has formed expectations about how B will behave in regard to some thing C. The forming of expectations, in this example, is because of the regard for the thing C and not because of the "trust" between A and B. When Hardin talks about this three-part relation he makes it clear that A has formed a set of beliefs about how B will behave when given discretion over C, that B will consider it as part of her interest to consider the interest of A. The interests of A are fulfilled by a limited set of behaviors that A recognizes as being within his interests. In Hardin's example, Karamazov trusted the merchant because he believed that the merchant considered it in his interests to make the Colonel money. One of Karamazov's expectations was that the merchant would return the money when it was needed. Karamazov had, not general expectations, but specific expectations about the behavior of the merchant.

When, taking the three-part relation as accurate, Mark trusts Jane with his boat we believe that Mark has formed a set of expectations about how Jane will treat his boat. We believe, for example that Mark expects Jane to return the boat in good condition or that Mark expects Jane not to take off for Mexico in his boat. The three-part relation sees these expectations not as stemming from Mark's trust in Jane or even as evidence of Mark's trust in Jane but instead sees the expectations as Mark's trust in Jane. Just as it denies the two-part trust relation, the three-part relation does not allow for generalized expectations. Hardin argues that were a two-part relation to exist then we would have to either 1) trust some individual with any and all things or 2) trust some thing to all individuals, this relationship is, on his conception, simply too broad to account for

trusting. Likewise, an open and general trust of an individual, even in regards to some specific thing, would require a willingness to accept any treatment of that thing. Hardin's three-part trust relation must therefore, in actuality, be a four-part relation, where the fourth part is the expectations formed by the truster about the trustee in relation to the thing trusted; [provide an example].

Expectations, however, are not what it is to trust but rather are a specific kind of by-product of trusting; we form expectations based on the trust we have. It is not 'I trust you because I have formed certain expectations about you.' It is instead 'I have formed certain expectations because I trust you.' Under the two-part, good will account my expectations can be open and general. In fact, though I *may* form specific expectations, I can only do that because I have generalized expectations about your motivations.

Unexpected Behavior

Even when we take trust to have a specific target (A trusts B with regards to C), describing it as trusting *for* that particular thing is misleading. If I trust you to open the door when I arrive, I may have formed a specific expectation about how you will behave, but the expectation is not the same as the trusting, rather the expectation is a result of the trusting. This is why unexpected behavior is not limited strictly to failure of trust but can also be a betrayal. In the beginning of *Raiders of the Lost Ark* Indiana Jones finds himself standing on the far side of a pit with the temple collapsing around him.[263] His companion has made it across the pit in advance and demands a precious relic in exchange for the rope Indiana needs to cross the pit. Indiana's companion ends up leaving him stranded on the far side of the pit but it is not the *act* of leaving him stranded which is the betrayal.

[263] Steven Spielberg, "Raiders of the Lost Ark," (United States: Paramount Pictures, 1981).

Indiana had done more than simply form some expectations about how his jungle guide would behave, he had to: asking an individual to risk life and limb for a small amount of gold, when the object of the search is a something worth a tremendous amount of gold, is inviting that individual to turn on you. That is, if you form an expectation about the behavior of another based on the idea that they will pursue monetary reward then failing to recognize that they will, in fact, pursue the greater monetary reward is a failure on your part not on the part of the other. We do not blame others for failing to fulfill our expectations *except* when those expectations were generated by some specific beliefs about the motivations of that other. When we are betrayed it is not the expectation that is hurt but rather our belief about the motivation of the other but because the expectation is the immediate source of the hurt, we mistake the expectation as being that to which the betrayal is attached. Instead it is the trust, which led to the expectation, which is betrayed by the failure to match the conditions of the expectations.

The jungle guide failed Indy's expectations when he demanded the exchange of the idol for the rope but more importantly he betrayed Indy's trust in that demand. The act revealed that Indy's beliefs about the guide's motivations were wrong (but the act itself is not the trust or the trustworthiness). The guide showed that he was more concerned with the idol than with Indy in the moment that he put Indy's life at risk in order to gain possession of the relic. The guide failed Indy's expectations precisely because Indy had come to believe that the guide was concerned for Indiana's wants and dreams. The guide revealed that the concern he had previously shown for Indiana's interests was really just a concern for his own interests. The guide really had encapsulated Indy's interests within his own interests and when the guide decided to take

the idol it was simply that his interests were trumping Indy's. This situation is exactly as Hardin describes trust but in doing so he losses the ability to explain why the guide's actions are wrong. On Hardin's account the guide has merely done what is expected of him; he is not untrustworthy. This makes the betrayal Indiana's fault and not the fault of the guide.

Two-Part

In "Trust and its Vulnerabilities" Baier is concerned with the ways in which trust can be broken. She does this, in part, as a way of showing that trusting is open-ended. Although Baier describes trust as a belief in the good will of the other, the trusting that she is concerned with is still something that we do; it is a kind of activity. Bentley Glass's account of trust, with its insistence on the necessity of trusting, and the subsequent importance of being trustworthy depends on treating trust as an activity; trusting is not an attitude for him it is an action. Luhmann's account, though it admits that we cannot trust for some specific end, still treats trusting as something in which we engage; trusting is a reduction of complexity that allows us to act. Hardin, while criticizing others for confusing trusting with entrusting, still confuses trust with an action. Each of these accounts, though unique, treats trust as an activity. Robert Solomon and Fernando Flores have a slightly different approach to the issue however; they present an account that describes trusting as a way of being.[264] The Solomon-Flores account considers trust from the perspective of authenticity. Their account of authentic trust is concerned with the manner in which we interact with others and how that comes from and creates an authentic life.

[264] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*.

Solomon and Flores are arguing for more trust in business, where 'trusting' is understood as accepting the promises or commitments of others without all the usual demands for contract that are normally required. That is, trusting, in a business context, is allowing the trusted party the flexibility to fulfill the obligations of the interaction without formalizing those obligations into a contractual rigidity. Even contracts that allow for some flexibility specify, in advance, the conditions of that flexibility; the rigidity of contracts is opposed to the flexibility of friendship. Trusting allows the trusted party the opportunity to respond to the fluctuating demands of business but it also opens the trusting party to the possibility of being harmed. The possibility of a positive interaction comes with the associated possibility of a negative interaction; the possibility of success and benefit is intertwined with the possibility of failure and harm. Under other business models trust is discouraged because either the administrative structure demands absolute submission to authority or because trusting makes the business vulnerable.[265] However, Solomon and Flores argue that trusting should be encouraged because in addition to the vulnerability associated with trusting, trust opens avenues of success which are otherwise closed. Without trust businesses must rely on contracts to enforce the exchange of goods and services but contractual obligations, by delineating the extent of the relationship, limit the ability of either party to be fluid. The standard view is that this rigidity provides protection to both parties but under Solomon and Flores' account it actually stifles a business' ability to match changes in the market. In addition, this fixation on contract separates partners rather than bringing them closer together. Although their account gives the impression of being motivated by something like an

[265] Ibid., 21.

economic model – they are concerned with the benefits of being trusting – their account is not, itself, an economic account. As I show below, Solomon and Flores do not argue for particular instances of trusting based on the benefits that can accrue to the trusting person. Nor do they argue that trusting, as a rule, will produce more benefits than costs. Instead, they argue for a view of human interaction that takes meaningful relationships to be built on trust. They attempt to convince those in the business community of the benefits of trust but the benefits gained from trusting are not part of the justification of trust. Solomon and Flores consider three types of trusting in an attempt to make clear why being trusting is something that should be encouraged in American businesses: Authentic, Blind, and Simple.

Simple Trust

Simple trust is the beginning of authentic trust. As we shall see later, the requirements of authentic trust are such that we cannot start out authentically trusting; we must first engage in other forms of trust. Basic trust is trust "without thought or reflection."[266] Basic trust then is our unthinking acceptance of others; that they will either do as they have claimed or will be concerned for our interests; it is a kind of primordial state of being. Basic trust is a position beyond ignorance, it is innocence. It is the trust of a child that has yet to learn that his parents are fallible.

The connection between basic trust and simple trust is often blurred. Solomon and Flores describe simple trust as trust that is "devoid of any sense of the possibility of distrust."[267] This description of simple trust is such that it seems undifferentiated from

[266] Ibid., 60.
[267] Ibid.

basic trust. It certainly appears that being 'devoid of any sense of the possibility of distrust' is to be unreflective but simple trust is also "trust that is taken for granted, that has gone unchallenged and untested, trust that is undisturbed."[268] This brings us closer to an understanding of the distinction. Simple trust is about particular cases of trust while basic trust, like the conception of authentic trust that we are working towards, is a way of being. Simple trust is trust that does not recognize the particular possibilities of the betrayal. Solomon and Flores continue, "[i]t is an attitude of assumption, trust by default, not a decision by way of deliberation and ethical and evidential consideration."[269] Simple trust is the trust that we have in our eyesight until we begin to notice it failing. Simple trust is ignorant; that is, simple trust does not consider the possibility that the other might be unworthy of trusting. To have simple trust is to behave as if the question of trust need not be asked.

It might seem that this kind of trusting, simple trust, is the ideal form of trust because so much of what we often stipulate as trust is unquestioning, an innocent attitude towards the possibility of betrayal. To recognize that another can, and possible might, betray you appears to embrace an attitude of distrust. Many people, especially those in the business world who Solomon and Flores are addressing, believe that to question the motives or capacities of others is just what it is to distrust them. To recognize the possibility of betrayal is, under alternative conceptions, what it means to distrust. But under in the Solomon-Flores conception of trust the recognition of the possibility of betrayal is far from being distrust, it is a necessary condition for authentic trusting.

[268] Ibid., 61.
[269] Ibid.

Simple trust is neither the model for nor the ideal of trust. The ideal trust, for Solomon and Flores, is authentic trust.

Blind Trust

Blind trust is trust that denies the possibility of betrayal. When one trusts blindly "one denies the obvious evidence."[270] The conceptions of blind trust and simple trust, as presented by Solomon and Flores, are similar but they are not the same; it is the hallmark of simple trust that "one does not even conceive of the alternative."[271] That is, when we do not recognize that betrayal is possible we trust simply. Simple trust can lead to substantial harm from betrayal but its innocence provides it a certain measure of protection.[272] Blind trust, on the other hand, is much more detrimental. Blind trust recognizes the evidence for the untrustworthiness of the other but denies that evidence, not because of other, more convincing, evidence, but rather by being willfully self-deceptive; "One does not ask, or, asking, does not listen."[273] This definition might appear confusing; 'not asking' is how we recognize simple trust and appealing to 'not asking' in the description of blind trust obscures the distinction between the two concepts. We must realize that there is something different about the 'not asking' associated with simple trust and the 'not asking' associated with blind trust. For Solomon and Flores the 'not asking' of simple trust is motivated by ignorance about the possibility of betrayal; you do not ask because you do not realize that there is anything worth asking about. The 'not

[270] Ibid., 64.
[271] Ibid.
[272] This position can perhaps be made more clear by suggesting that when I am ignorant of the possibility of betrayal I can be harmed by the betrayal but it is no indictment of my decision making capacity. Rather, I can assure myself that it was only my ignorance that led me to my current situation not some flaw in my capacity to judge trustworthiness.
[273] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*, 64.

asking' of blind trust is motivated by a desire to not know the answer; the ignored wife is often an example of blind trust. She suspects her husband of cheating but never asks because so long as she does not confront the cheating, she does not have to face her husband's infidelity.[274] But not only does blind trust deny real betrayals it denies potential betrayals as well. The ignored wife trusts blindly even when her husband remains faithful; willfully overlooking the ways in which another may betray you leads to blind trust.

Simple trust is to not ask because one does not ever think to ask, one does not even imagine that there is anything to ask about. Blind trust, in contrast, is willfully self-deceptive.[275] Blind trust is different from simple trust because blind trust denies the evidence against trusting while simple trust is ignorant of the evidence. One way to think about the distinction between these two conceptions is to consider the end which the trust is aimed; blind trust maintains a trusting attitude for some other end, for something other than the trust itself; we would not deny evidence against trusting if our concern were for the trust. It seems then that the only motivation for trusting blindly is what that trusting relationship can get you. But it can only do so when it is artificially propped up. Blind trust subordinates trust to some other end or desire. We can see this by returning to Solomon and Flores' example of an ignored wife. She may be presented with all the evidence that her husband has begun to stray and yet she continues on as if nothing has changed. Some other concern, like maintaining the family, provides the justification for trusting her husband. This is in contrast to maintaining the relationship because her

[274] Ibid., 67. This is an example that they provides though, in this instance, I am using it in a slightly different context.
[275] Ibid., 65.

husband is actively working to rebuild the trust they once shared. Blind trust, when the betrayal is finally admitted, destroys trust.

Blind trust denies the possibility of betrayal. One need not have been betrayed in the past to blindly trust others. Saying 'he would never betray me' is to deny the possibility of that betrayal; it is to ignore all the ways that he could betray and all the reasons why he might. The conception of blind trust presented by Solomon and Flores begins to show us what authentic trust is going to be. We can learn from blind trust what authentic trust is not; authentic trust does not deny that others may betray you, it accepts it.

Authentic Trust

Authentic trust is the kind of trust for which Solomon and Flores are arguing. It is the mature recognition of the freedom of others and the ideal form of trusting. This is trust that knows that the other might defect; it is trust that is open to the possibility of betrayal. Simple trust is ignorant of betrayal while blind trust denies betrayal but authentic trust accepts the possibility and even the reality of betrayal; "[a]uthentic trust is a judicious combination of trust and distrust, superior to blind trust, which is foolish precisely because it bars distrust from consideration."[276] When we are blind to the possibility of betrayal we have a diminished capacity to express our agency; willfully ignoring all the ways a partner may deceive eliminates those concerns from the process of deliberation. Authentic trust is trust that sees the evidence against trusting, recognizes its worth, and still trusts.

[276] Ibid., 46.

However, authentic trust cannot be merely recognizing the possibility of betrayal and trusting anyway; to recognize the possibility of betrayal and to ignore that possibility in favor of trusting is to deny the evidence. This would mistake blind trust for authentic trust. The evidence against trusting, or the possibility of betrayal, must make a difference to the trust without either being ignored or being denied. Solomon and Flores are making a very fine-grained distinction between authentic trust and its alternatives; innocence, ignorance, and denial.

The recognition of the possibility of betrayal must make a difference to the trusting but "the weight of the evidence [is] at best a secondary concern, because trusting changes both the person trusted and the person who trusts."[277] Solomon and Flores see trusting as a reciprocal relationship. When I trust you the trust changes the way I interact in the relationship; it alters the available options and modifies the available responses. When I trust you the trust changes you as well; being trusted changes the responsibilities of the trusted party. And, being trusted itself provides a reason for being trustworthy. A trust relationship is dynamic, continuous, and, if authentic, life altering.

To trust authentically is not merely to recognize that the other may betray you and to hope that they do not; "[a]uthentic trust…is self-confident rather than simply optimistic."[278] Self-confidence, on the Solomon and Flores account, is a hallmark of authentic trust. The recognition of self and how a relationship can strain, stress, and harm that self is important to what it is to be self-confident and therefore important to trusting. Having confidence in your own capacities and understandings is what gives you the ability to enter into relationships without first erecting barriers to protect you from the

[277] Ibid., 95.
[278] Ibid., 92.

possibility of betrayal, barriers that limit you as much as they limit others. Knowing where you are vulnerable and being comfortable with that vulnerability is essential to authentic trust. Authentic trust is not focused solely on the other instead "[i]ts focus is on one's own responsibilities in trusting."[279]

Contentious Conclusions

In many respects the Solomon-Flores account is most similar to Baier's position; authentic trust has many of the same characteristics as Baier's account of the belief the goodwill another has towards you. Each of their accounts focuses on the individual who trusts and how that individual relates to and perceives the world around them. The Solomon-Flores account is open-ended like Baier's but these accounts have different justifications for this; authentic trust is open-ended because it creates new opportunities. The differences between these two accounts are of more significance though. On the Solomon-Flores account "[t]rusting is a choice, a decision, and authentic trusting takes into primary account the way the relationship will change as result of that choice."[280] Baier's position is that trusting is not a choice that we can make in so straightforward a manner and so the changes that come about in relationships because of trust are strictly within our control.

The authentic account of trust has similarities with the other accounts as well. Solomon and Flores make a distinction between trust and confidence much like Luhmann does. Luhmann's distinction is that we trust others only when we have more to lose than to gain but the authentic account holds that trust is in some way reciprocal while

[279] Ibid.
[280] Ibid., 95.

confidence is not. That is, we can only trust those things which have to capacity to be concerned about the trust we have in them. Inanimate objects cannot be trusted; "trust, properly understood, is a function of human interaction."[281] Beyond simply being limited to animate objects the concern for reciprocity limits trust relationships to those in which the trusted party can, conceivably, be concerned about the trust that is placed in them. I cannot therefore claim that I trust some random person to do some action X since that person could have no knowledge that I have formed this expectation about them. But, I can claim to trust the pilot of the airplane I am riding in because I can reasonably assume that the pilot is aware that the passengers have put a great deal in her care and that fact should change the way she conceives of the act of keeping the plane in the air. Trust on this account is not merely regularity; "trust is not the same as predictability, and without unpredictability, there is no role for trust."[282] If we were all regular and predictable we would have no need to trust others; regular and predictable things are not affected by the perception of their regularity and predictability. A regular thing is regular whether or not you take notice or even depend on its regularity.[283] Trusting then, on the Solomon-Flores account, is not an outgrowth of predictability but instead on unpredictability; we must be unpredictable in order to be trusted. This is a somewhat contentious conclusion to be sure; it means that not only is unpredictability necessary for trusting but, because authentic trusting is a good thing, unpredictability is necessary for living a full and authentic life.

---

[281] Ibid., 70.
[282] Ibid.
[283] Cf. Annette Baier 99. Baier makes a similar distinction when she discusses the ways in which we are let down.

Solomon and Flores, like others, consider the competence one has in relation to some task to be important for trusting but they do not hold that an individual's competence is a necessary or sufficient condition for trusting.[284] They believe that competence varies in its importance to any instance of trusting but that it is an issue involved in all cases of trusting. That is, we should not trust our gardener to perform open heart surgery because he lacks the proper competence to complete the task and this task is such that its failure has profound and irreversible effects. But, this position does not demand that we should trust our doctor to perform open heart surgery simply because he has displayed a competence at doing so. To consider the competence of the individual to be sufficient condition for trusting is to mistake trust for reliability; using trust in this limited three-part relation is to see trust as this kind of mechanistic predictability. The Solomon-Flores account is not presenting this kind of competence dependent conception of trust; "competence as such is the object of reliance, not trust."[285]

Another aspect of trust that the authentic account shares with others is that it considers trust to be a risk. Where is differs from those others is the degree to which it embraces that uncertainty; "[t]rust involves risk, and authentic trust involves knowingly going into the unknown – together."[286] This glorification of the unknown is what best exemplifies the Solomon-Flores account, it presents trusting as a way of being and not merely an action that we sometimes engage in.

The Solomon and Flores account of trust successfully denies the need for a three-part relation in which there is the truster, the trustee, and the object of the trust. Given the

[284] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*, 84.
[285] Ibid.
[286] Ibid., 128.

arguments of the preceding chapters and both my arguments as well as theirs then the conceiving of trust as consisting in a two-part relation should not so far fetched as other authors have suggested. Further, given the hurdles that a three-part trust relation faces, especially its connection to entrusting, we should be suspect of the veracity of that conception. My presentation of the conceptual understanding of trust is therefore complete. Trusting and trustworthiness are conceptually connected through our belief in the good will the other has towards us; to be trustworthy is to have good will towards the other and to be trusting is to believe another to have good will towards you. Having good will is not in regards to some other thing, either object or person, but is instead a general statement about the motivations the other has; specifically that one is motivated by a genuine concern for another's well being. What remains, the subject of Chapter Six, is to show the effect that trust has on our understanding and interpretation of the behavior of others.

Chapter Six: Trust as an Interpretive Framework

The goal of this chapter is to show that trust, in the words of Trudy Govier, "affects the way we interpret what others say and do."[287] I argue that the trust we have for another determines which of the possible interpretations we will consider when we make a judgment about how the other has behaved or a prediction about how another will behave. For any action there are far too many possible interpretations of that action for a single individual to possibly canvas them all. We must reduce the number of those interpretations to a manageable number in order to function in the world. My position is reminiscent of the view of trust suggested by Niklas Luhmann when he claimed that "[i]n trusting, one engages in action as though there were only certain possibilities in the future."[288] Though reference to engaging in action is perhaps misleading; it is not, on Luhmann's account, that trusting is to engage in particular actions but that when we are trusting we entertain only certain possibilities – trusting is the reduction of the complexity of possibilities. I argue that when we are trusting of others, we limit the possible interpretations to the set of positive interpretations, even, if the trust is strong enough, to only the very positive interpretations. When we are distrusting we limit the possible interpretations to the negative set. This gives us a guide for forming expectations about how others will behave; if I distrust Bob then I will form the expectation that he would, if I allowed him, take advantage of me. The trust that I have in Bob (in this case

[287] Trudy Govier, "Trust, Distrust, and Feminist Theory," *Hypatia* 7, no. 1 (1992): 17.
[288] Luhmann, *Trust and Power : Two Works*, 20.

the belief that he has an ill will towards me) is not an action; it is a description of how I perceive Bob.

Trust affects the way we interpret the actions of others, therefore, it is necessary for making moral judgments. It is not enough to have learned that someone has done an action, we must see this action through the lens of our understanding, which is to interpret the action, in order to judge the rightness or wrongness of that action; if Bob reveals to Charles something that you have kept secret, your judgment about Bob choosing to do that will depend on what you think of Bob's motivations towards you. Did Bob do this to help you or to hurt you? Did Bob, perhaps, do this to help himself? Even if Bob claims that he did it for your benefit, whether you accept his claim or not will depend on how you perceive of him. This point follows from, although perhaps not obviously, from the arguments made in previous chapters. It is useful, therefore, to reconsider the arguments made so far with an eye towards making this last position clear. In this chapter I make four claims: 1) trust is not an action, as evidenced by the various ways that trusting and trustworthiness fail to be causally linked; 2) trust is risky, and therefore is instead about motivation, because it is necessarily uncertain, which follows from the claim that trust is not an action; 3) trustworthiness is to have a goodwill towards another and trusting is to believe that the other has a goodwill towards you, which is derived from an investigation of possible motivations; 4) trust is a two-part, rather than a three-part relationship, which follows from the recognition that trust is concerned with good will; and finally 5) trust is an interpretive framework that is necessary for making judgments about others.

Introduction

In October of 1962 the nations of the world collectively held their breath while the United States and the Soviet Union stood on the edge of a precipice. Relations between the two nations, already strained, became taut as a violin string over the Soviet Union's involvement in the creation of missile launch sites on the island nation of Cuba, sites capable of launching nuclear missiles, of unleashing "weapons of sudden mass destruction."[289] The United States, fearful of the presence of powerful weapons so close to American soil, attempted to stop the creation of those sites by a naval quarantine of the island. The stalemate created by the U.S. blockade of Cuba had the potential to catapult the world into a war that, because of the extent of the proliferation of nuclear weapons, would have been absolute. One false move or misstep could have resulted in the reduction of every major American city to a heap of slag; evidence that surfaced after the crisis indicates that we only narrowly escaped that fate. While nuclear war would have been horrifying for America, Europe, and the Soviet Union the effects of this assault would not have been limited only to those nations who were the direct target of missiles. The fall-out of nuclear Armageddon would circle the globe raining death on every nation; the film *On the Beach*, which gives form to our fear of mutually assured destruction, presents exactly this scenario.[290] If I am being overly melodramatic regarding this conflict, which seems impossible to do given the panicked response of the American populace (bomb shelters being built in Key West, Florida is but one example), it is only

[289] John Fitzgerald Kennedy, "Radio and Television Report to the American People on the Soviet Arms Buildup in Cuba," http://www.jfklibrary.org/Historical+Resources/Archives/Reference+Desk/Speeches/JFK/003POF03CubaCrisis10221962.htm.

[290] Stanley Kramer, "On the Beach," (The United States of America: Lomitas Productions. Released by United Artists, 1959).

because I want to emphasize the stakes inherent in the interaction between these two world powers. Robert Kennedy described the crisis as one that "brought the world to the abyss of nuclear destruction and the end of mankind."[291] There was, quite simply, no more obviously important political event of the 20th century and, therefore, no more obviously important series of decisions and judgments.[292]

The Cuban Missile Crisis stands as a Cold War testament to the importance of trust in our daily lives and, even more directly, to the importance of trust on the grand scale of nations. It was trust, either through trusting or distrusting, that led to the conflict and its resolution, trust between Americans and Russians and trust between Kennedy and Khrushchev. Generally, as it is understood in relation to something like the Cuban Missile Crisis, trust is presented in the form of a Prisoner's Dilemma, a thought experiment that is well rehearsed in this dissertation. Still, it bears further consideration. Considering trust as the Prisoner's Dilemma, President Kennedy and the United States were faced with a decision about how to proceed given the words and actions of Premier Khrushchev and the Soviet Union: do we act as if they are going to cooperate with us or do we act as if they will pursue only on their own self-interest? As I argued in Chapter 3, the Prisoner's Dilemma was first conceived as a way to understand the dynamic and potentially volatile interaction between the agents "just as nuclear proliferation and arms races became serious concerns."[293] The Prisoner's Dilemma and nuclear proliferation

---

[291] Robert F. Kennedy, *Thirteen Days : A Memoir of the Cuban Missile Crisis* (New York: W.W. Norton, 1969), 19.

[292] There are events that are as important (the first powered aircraft flight and landing on the moon) and there are events that have as profound an effect on our thinking (the Holocaust) and there are even events that change the way we view the world (the bombing of Pearl Harbor and the assassination of John F. Kennedy) but none of these events required the kind of judgment and decision-making response that is important to this conversation.

[293] William Poundstone, *Prisoner's Dilemma*, 1st Anchor Books ed. (New York: Anchor Books, 1993), 9.

came of age together and were mutually reinforcing. Since the Cuban Missile Crisis is one of the excesses of nuclear proliferation it is not a stretch to view it as an example of the Prisoner's Dilemma and the Crisis certainly would have been seen as an example of the Prisoner's Dilemma for the American's involved. We see in this example what Andrew Kydd might call a security dilemma, the heart of which "is mistrust, a fear that the other side is malevolently inclined and bound to exploit one's cooperation rather than reciprocate it."[294]

In this chapter, however, I want to challenge the relationship between the Prisoner's Dilemma and the Cuban Missile Crisis. I argue that trust is to be found in how we understand others, how we interpret their actions based on what we believe their motivations to be, and not in how we act towards those others. I present an alternative view of the Crisis that maintains the importance of trust to the situation while eschewing all those features which have been shown in the preceding chapters to be problematic for a conception of trust. This chapter consists primarily in a rehearsal of the arguments made in previous chapters. However, here the focus is altered slightly in order to show that the arguments support the reciprocal-interpretive model of trust that I am suggesting. I begin, therefore, by returning to the argument of Chapter 2. I show that trusting and trustworthiness are not causally connected in order to show that trust is not directly linked to the actions in which we engage; actions are, in this sense, irrelevant to trust. This is followed by a return to the argument of Chapter 3 in which I claim that trust is risky. However, in revisiting this point I show that the risk inherent in trust is an uncertainty about not only the motivations of others but the meanings of their actions as well. Since

[294] Andrew Kydd, "Trust, Reassurance, and Cooperation," *International Organization* 54 (2000): 325.

actions are not the mark of trust and since there is uncertainty about motivations we must interpret the behaviors of others. This leads to the argument from Chapter 4, which put forth the claim that trusting was a belief in the goodwill of another. Here we see that merely having an affect on our interpretation of events is not differentiate trust from other ways of forming expectations; my sister has a tendency to give very negative interpretations of things when her blood-sugar is low but this is not a sign of distrust on her part. Trusting, as different from other influences on interpretation, must be concerned with a particular kind of interpretation, an interpretation about the motivations towards you of another. When we believe that another has goodwill towards us our interpretation of her behavior skews to the set of positive interpretations while believing that she has ill will would skew my interpretation to the negative set. Since the role of trust is as an interpretive framework, it cannot be the case that trust is limited to a three-part relationship but is instead a much more open-ended two-part relationship, which is the argument of Chapter 5. The interpretive model, having been established, I then show that the interpretation itself is subject to revision and interpretation. Trust is not a static belief that we have but rather is in constant flux. This allows for change, which, I think, gives us hope for the future; however much the world might be full of deceit and distrust today, we can change.

The Cuban Missile Crisis, from the perspective of game theory, is understood as an example of Chicken or Brinkmanship, a game in which each player tries to get the other player to flinch first. In this style of game "both players want to do the *opposite* of whatever the other is going do."[295] Obviously, the stand-off that was the Cuban Missile

[295] Poundstone, *Prisoner's Dilemma*, 199.

Crisis can be seen this way; each side wanted the other side to back down first. But, it can also be understood as an example of the Prisoner's Dilemma, in which each side only wants to cooperate but only *if the other will also cooperate*. Whichever style of game we take to be paradigmatic of the Crisis, the concern for either game is on action. Do we defect or do we cooperate? Will they defect or will they cooperate? Should we choose to cooperate knowing that our cooperation could lead to our annihilation? The questions that are asked, and therefore the answers that are formulated, invariably highlight and even demand an action oriented perspective of the trust relationship. From the standpoint of the Prisoner's Dilemma, trusting and trustworthiness are both seen as the act of cooperating while distrusting and untrustworthiness are seen as defecting. I have already argued, in Chapter 3, that actions are not and should not be taken to be equivalent to beliefs and further that the Prisoner's Dilemma fails to distinguish between trusting behavior and trustworthy behavior (or distrusting behavior an untrustworthy behavior). That is, I have shown that trusting is something distinct from the actions that appear in conjunction with it. Trusting, following the position of Annette Baier, is to believe that another has good will towards you. [296] Trusting is to believe of another that she is concerned for your interests because they are your interests and not simply because doing so will further her interests. I have claimed that trusting is not limited to specific instances or particular cases but is instead a generalized view of the motivations of the other. The position that I present in this chapter will follow, if not expressly mimic, the position put forth by Niklas Luhmann when he suggests that trust is a reduction of

[296] Baier, *Moral Prejudices : Essays on Ethics*.

complexity.[297] Trust limits the totality of possible interpretations of behavior to a manageable set, which then allows us to make judgments about future actions.

Interpretation

The Cuban Missile Crisis is not unique in the history of U.S. foreign relations. There are many instances that can be seen as examples of Prisoner's Dilemma style decision-making. For instance, in November of 1983 the members of NATO were conducting a series of military exercises across Europe called ABLE ARCHER – 83 the point of which was to practice allied response to the threat of a full-scale nuclear assault.[298] The nations involved, striving for realism, actually went so far as to have various heads of state participate in the exercise. Seen externally, by an opposing power like the Soviet Union, the combined actions of NATO could be evidence of preparations for a nuclear first strike. Peter Pry tells us that the "Soviets were leery of NATO military exercises, because Soviet military doctrine warned that an enemy might use training to conceal preparations for an actual attack."[299] The ABLE ARCHER exercises followed shortly after the United States had finished install installing the Pershing II missile sites in Europe, which, once the missiles arrive, greatly increased the attack capabilities of the West.

I must digress briefly about the missile sites as there may be some confusion regarding the installations but this digression will help clarify my position. The missile **sites** were complete at the time of the ABLE ARCHER exercises but the missiles

[297] Luhmann, *Trust and Power : Two Works*, 13.
[298] Peter Vincent Pry, *War Scare : Russia and America on the Nuclear Brink* (Westport, Conn.: Praeger, 1999).
[299] Ibid., 38.

themselves had not yet arrived. This means that NATO forces were effectively playing with plastic guns; even if they had wanted to fire the new Pershing II's, they could not. The Soviets, however, believed that the missiles had already been installed. Pry says that "ever since the Cuban Missile Crisis in 1962, each side credited the other with more strategic stealth than it really possessed."[300] So, from the perspective of the Soviets, the sites were fully operational, which means that all of their decisions were made on the assumption that NATO had the new missiles. The misunderstanding about the missile sites is a revealing example of the position that I am arguing for in this chapter: trust sets the limits of viable interpretations for the actions of those around us. The distrust, the belief in ill – rather than good – will, of the Soviets for Western forces was such that they expected NATO to be acting deceptively about the installation of the new missiles. It was the Soviet interpretation of the events of ABLE ARCHER that made November of 1983 so dangerous, which shows the importance of the position that I am presently outlining; trust effects how we understand the actions of others.

The West, however, did not immediately recognize how the exercises would be viewed. The installation of new missile systems, as a matter of simply military protocol, demanded a test run so that the various people instrumental in the deployment of those systems could become familiar with their use. The apparent increase in offensive capabilities that preceded the exercises made the actions of NATO forces even more suspect to the Soviets. And, in addition, the exercises in question began only three weeks after the United States invaded Grenada in order to oust the communist regime that had

[300] Ibid., 34.

recently taken over the small island.[301] The invasion of Grenada was seen by Americans as a moment of liberation; this view is presented in the movie *Heartbreak Ridge*, in which a platoon under the command of Clint Eastwood seems to single-handedly eliminate the communist forces that have overrun the island.[302] But, the invasion of Grenada was seen by communists as part of the aggressively imperialistic agenda of the United States as led by Ronald Reagan, a kind of trial run of combat effectiveness made even more pressing by the planned war games. Each side attached a different interpretation of the events leading up to the ABLE ARCHER because of the way they viewed each other's motivations. However, considered more traditionally, in the fall of 1983 the Soviets were presented with something that can easily be seen to be an example of the Prisoner's Dilemma.

Imagine the United States and the Soviet Union as criminals that were brought in for questioning. The Soviets could see but not hear what the Americans were doing in an adjacent room. The Americans were presenting themselves as if they were preparing to defect on their partner, by giving all the appearances of preparing for a nuclear first strike (troop build-up and deployment, increased secret communications, altered schedules and movements of leaders). The Soviets had to decide whether they should defect or continue to cooperate. That is, the Soviets needed to decide whether they should launch their own first strike and thereby avoid the worst possible outcome of being annihilated – cooperating while their partner defected. Seen in this way, the ABLE ARCHER incident

[301] It is worth noting that the U.S. did this, apparently, without the consent of Queen Elizabeth who is, technically, the head of Grenada's government. This precipitated a series of top-secret, encoded communications between Great Britain and the United States: exactly what the Soviets anticipated would happen prior to an attack.
[302] Clint Eastwood, "Heartbreak Ridge," (United States: Warner Bros., 1986).

is about action rather than being about belief; the focus here is on the behaviors of the other and on what behaviors the agent should engage in. From this perspective, the Soviets were not concerned with the motivations or intentions of the Americans they were only concerned with what they should do. Or, at least, given the usual understanding that is how the story is normally told.

The alternative that I am presenting can be found, in kernel, in Pry's recognition that the Soviets had to *interpret* the behavior of the Americans; he suggests that "Soviet alarm over communications anomalies in ABLE ARCHER – 83 was reinforced by a too-literal interpretation of the exercise scenario."[303] That is, all assurances to the side, the Soviets had to make a judgment about what the activities of the Americans *meant* and the only way to do that is to have ascribed motivations and intentions to the Americans that would allow concerned parties, i.e. the Soviets, to describe what the Americans were doing. It was not enough to merely look at the activities of the Americans in order to understand what they were doing; there was no objective way to view this situation. A behavioristic approach to action, an approach that holds that all we can know about people is how they behave, would have to conclude that the activities of NATO were evidence of preparations for a first-strike nuclear assault because all the actions of NATO were exactly that – preparations for a nuclear assault; NATO was, in fact, practicing for a full-scale war. It is perhaps cliché to point out that if it looks like a duck and quacks like a duck then it is a duck but to the Soviets the Americans were looking and quacking like a nation that was preparing to launch a nuclear assault.

[303] Pry, *War Scare : Russia and America on the Nuclear Brink*, 39.

Understanding what, exactly, the Americans and their allies were up to, however, required more than simply looking at their behavior or listening to their words. It required a belief about NATO's motivations and about their relationship to the Soviet Union. If the Soviets believed, as many did, that Americans wanted to annihilate the Soviet Union, that we were aggressive and imperialistic, that we saw the Soviets as a threat that needed to be destroyed, then it is easy to see how they would interpret our actions during the ABLE ARCHER exercises to be an indication of our intent. Their belief about us would be reinforced by the actions in which we were engaged. If, on the other hand, the Soviets believed, as some did, that the Americans, though threatened by the Soviet Union, had no desire to start World War III, that we desired long and fruitful lives for ourselves and our neighbors, and that although we wished them gone we had no desire to harm the Soviets, then our actions could be taken to accord with our description of them (as training exercises). Actions must be viewed through the lens of our beliefs about the motivations of others in order to form any interpretation of what those actions mean.

No Connection Means That Trust Is Not About Actions

The interpretive model that I am suggesting is far from common as a way of conceiving our relations with others. The behavioristic, action-oriented, Prisoner's Dilemma style conception continues even today. We can see this action-oriented model in our diplomatic relationship with Iran. We, the United States, are fearful of the worst possible outcome (Iran developing nuclear weapons) which leads us to be unwilling to 'trust' them (accept their sovereign independence and right to self-rule) and so have considered defecting (invade) before they have a chance to defect on us (launch missiles).

More recent current events can even be seen in this light; the stock-market crash at the end of 2008 and the corresponding government bail-out/rescue had to confront the specter of the Prisoner's Dilemma. President Bush, Speaker Pelosi, an army of senators and representatives, as well as legions of analysts and commentators argued that we cannot 'trust' that the economy will rebound, we must defect and abandon free-market principles in order to save ourselves from the worst possible scenario. Those who opposed government intervention nevertheless used the same terminology and the same methodology in responding to the problem. The market had, they claimed, been artificially propped up by socialist principles. We can, they suggested, either continue endorsing these ineffective and un-American principles by cooperating with law-makers and Wall Street executives or we can defect by refusing to cooperate, let the businesses that are over-leveraged fail. The perspective of cooperation and defection continues to dominate discussions of social interaction but it inappropriately focuses our attention on actions, which, as we have seen, are not what it means to trust. The lack of a necessary causal connection between trusting and trustworthiness means that the actions that are associated with trust are not, themselves, trust. Therefore, the actions of the United States and the Soviet Union during the Cuban Missile Crisis or during the ABLE ARCHER incident cannot have been examples of trusting or distrusting. Neither should we understand our relationship with Iran or how to respond to our faltering economy in these terms. Instead, it is because we have such a poor view of Iran's motivations towards us (and because they have such an incredibly poor view of our motivations towards them) that we are unwilling to accept their claims about the development of nuclear technology; we interpret their actions in the worst possible way. And, our view of markets or

government involvement in markets influences the interpretation that we have of the actions of markets and governments. This explains why the same data continues to be understood differently by those who seek a road to recovery: they interpret each action differently.

Trusting and trustworthiness are often taken to be two sides of the same coin. Their connection seems to be beyond doubt; besides the obvious similarities in spelling, trusting and trustworthiness, in some sense, imply each other. It seems improbable to most of us that there could be trusting without trustworthiness or that there could be trustworthiness without trusting. Bentley Glass holds to this very ideal in his article "The Ethical Basis of Science."[304] He argues that there can be no trusting without trustworthiness and, additionally, because we need to be trusting there is an obligation to be trustworthy. Even Russell Hardin suggests that the "best device for creating trust is to establish and support trustworthiness."[305] Glass' focus, the issue that motivates his argument, is a concern for accepting the testimony of others. Trust continues to be of concern to those who are interested in testimony. Paul Faulkner examines the importance of trust to testimony in his 2007 paper and even suggests a position very similar to that of Glass; "we accept what others tell us because we depend on them for information."[306] But Faulkner, unlike Glass, believes that an important aspect of trusting is "that it need not be evidence based."[307] Elizabeth Fricker also considers the importance of trust to testimony when she describes the paradigm case of testimony as knowledge "spread from speaker

[304] Glass, "The Ethical Basis of Science."
[305] Hardin, "Trustworthiness," 29.
[306] Faulkner, "On Telling and Trusting," 885.
[307] Ibid.

to trusting hearer."[308] There remains a strong belief in a necessary connection between trusting and trustworthiness to this day.

Given the continued concern for the relationship of trusting to trustworthiness one might expect the issue to be settled yet even a cursory investigation reveals how radically separate the two concepts really are. Two major problems immediately come to mind when considering the connection. First, trusting or distrusting another does not entail that the other is either trustworthy or untrustworthy; untrustworthy people can be trusted and even trustworthy people can fail to behave in a trustworthy manner. Second, trustworthiness does not entail being trusted; those who are worthy of our trust often fail to be trusted. And, additionally, it turns out that untrustworthy people often manage to fulfill trust. If being trustworthy demand particular actions then positions like that of Glass would make sense. This issue is an old and long-standing one; Plato addressed a similar problem early in *Republic*. Socrates suggests that when "a sane mane lends weapons to a friend and then asks for them back when he is out of his mind, the friend shouldn't return them."[309] If trusting a friend is to believe that the friend will act in particular and specific ways then the friend ought to return the weapons lent regardless of the condition of the lender. That is, a necessary connection between trusting and trustworthiness would entail that the action taken, and not the motivation for the action, was paramount to trust. However, Plato's account indicates that even in ancient Greece people recognized that a friend could be doing what is best for you by not keeping his promise, or rather, that to be trusted is to be trusted to make a good judgment and not to

[308] Elizabeth Fricker, "Second-Hand Knowledge," *Philosophy and phenomenological research.* 73, no. 3 (2006): 593.
[309] Plato, G. M. A. Grube, and C. D. C. Reeve, *Republic* (Indianapolis: Hackett Pub. Co., 1992), 6.

fulfill some pre-specified action; actions are definitive of trust. On this account then there is nothing in particular about either trusting or trustworthiness which guarantees the existence of the other. Denying the importance of particular actions opens the door for an account of trust that focuses on motivation.

In Chapter Two I used Bentley Glass as the exemplar of a position that failed to recognize the importance of keeping trusting and trustworthiness separate. He overlooked the weight and the implications of his position's insistence on a necessary connection between trusting and trustworthiness. By claiming that the necessity of being trusting provides the justification for being trustworthy he ties the two concepts together in a way that defies the obvious objections stated previously. Glass, however, is by no means the only example available of this kind of oversight. We can see now, after having reviewed a number of other positions, how important a necessary causal connection is to most understandings of a trust relationship. Russell Hardin's position, for instance, by appealing to encapsulated interests, holds that my trustworthiness is generated by the recognition that my interests can be furthered by considering your interests.[310] We have seen further that the Prisoner's Dilemma, which Hardin specifically appeals to, makes no distinction between trustworthy and trusting behavior and, because it views the acts of cooperating and defecting to be equally trusting and trustworthiness or distrusting and untrustworthiness, it ignores the clear distinctions between trusting, trustworthiness, distrusting, and untrustworthiness.

Returning to a consideration of the Cuban Missile Crisis we can see the importance of our conception of trust to understanding what happened. If trust is

[310] Hardin, "Conceptions and Explanations of Trust," 19.

understood as actions then the only way to understand the events of 1962 is by directly considering the actions in which each side was engaged. The Soviet Union had defected from cooperative non-interference in the sphere of American influence by covertly attempting to install missile launch sites capable of hitting American soil. The action of building those sites would answer the question 'are the Soviets trustworthy' with a resounding 'No,' which forced the United States to defect as well since it would be a sucker's payoff to allow the Soviets to defect while continuing to cooperate. Further, the American response of a naval blockade was also a defection. By isolating Cuba, America was preparing for a defensive defection; they were trying to avoid the worst possible outcome by making sure that the Soviet defection was met with a similar defection. But, by isolating Cuba, the United States showed that they were, in fact, prepared to invade Cuba, which is what the Cubans and the Soviet Union feared in the first place. The actions taken by each nation, if we accept the validity of the Prisoner's Dilemma, are supposed to be examples of our (dis)trust, which reveal that neither side trusted the other. Toshio Yamagishi makes this point when he claims that "[f]rom A's perspective, her defection is necessitated by B's perceived untrustworthiness, but from B's perspective, A's defection is a proof of her own untrustworthiness."[311] The interpretive account that I am suggesting makes better sense of this scenario because instead of appeal to the actions of others as examples of trust it looks to how we understand those actions. My position considers actions to be, at best, indications of trusting or distrusting and not themselves examples of trusting and distrusting.

[311] Toshio Yamagishi et al., "Separating Trust from Cooperation in a Dynamic Relationship: Prisoner's Dilemma with Variable Dependence," *Rationality and Society* 17, no. 3 (2005): 279.

Because the Soviet Union distrusted the United States, they saw American actions in the worst possible light; they took America to be plotting against them. The Cubans also tended to interpret American actions in negative ways, which was only exasperated by the failed Bay of Pigs invasion, which reinforced this negative assessment and encouraged the Cubans to prepare for an eventual full-scale invasion. Accepting the installation of missile sites may appear to be a defection in a game-style analysis but it fails to be a defection of trust because the Cubans already failed to trust the Americans; the true defection is beginning to distrust or beginning to be untrustworthy. Cubans were well past the point of beginning to distrust when the missile sites began being built.

It can be argued further that building those sites cannot be an example of distrust because the reverse could not also true; failing to build missile sites would not be a case of trusting, it would not count as cooperation. Cuba could choose **not** to build sites for nuclear missiles for any number of reasons, not the least of which is distrust. In choosing not to build missile sites they might simply have been afraid of nuclear weapons. Or, the Cubans might have believed that the Americans sought an excuse to actively deny Cuban sovereignty and that American spies were everywhere in Cuba such that no attempt to build a missile site would go unnoticed. Should the Soviets be involved the Cubans could then assume that the Americans would appeal to the Monroe Doctrine, as they had in the past, and attempt to invade Cuba. Given this scenario, distrust of American involvement in Cuba could lead to refusing Russian missile sites as opposed to the real world example of distrust of American involvement leading to Castro accepting Soviet missiles. If distrust of the Americans can explain both choosing to build and choosing not to build missile sites then it seems that, as an explanation for what happened, it is ambivalent. Not

building the sites is indeterminate in regards to trust but so too is building them. Just as an action, by itself, in the Prisoner's Dilemma fails to be clearly either trusting or distrusting so too are actions in the world; the act of building the missile sites does not clearly show, on its own, whether the Cubans trusted or distrusted the Americans. This reminds us of the argument from Chapter 2: there is no necessary connection between being trusting and trustworthy (or distrusting and untrustworthy).

Revelations about the lack of a necessary connection are, at best, merely reminders of the facts of life to most people; few are surprised to learn that being trusting of another is no guarantee that the trusted person will fulfill that trust; we are only ever surprised to find out that we have, in fact, been betrayed in any particular instance. But this recognition is by no means a claim that nothing important can be learned from considering the relationship.

First, we can see that there is, indeed, a connection between trusting and trustworthiness. My trusting of another may not entail that the other actually *is* worthy of my trust but it does necessarily mean that I *believe* that the other is worthy of my trust. Trusting, then, simply is to believe that the other is trustworthy. This separates the concept of trust from something like concept of justice. Although I can be just without anyone believing me to be so and I can be trustworthy without anyone necessarily believing me to be so I cannot be *justed* in the way that I can be trusted. That trusting is to believe another is worthy of trust is an important component to our understanding of trust. Further, this point mirrors Hardin's claim that the "assessment of your

trustworthiness in a particular context is simply my trust of you" without adopting his problematic connection.[312]

Second, the lack of a causal connection between trusting and trustworthiness shows that a trust relationship cannot be predicated on action. That is, since there is no necessary causal connection between trusting and trustworthiness it is a mistake to assume that my trusting of you is in any strong sense connected to the actions in which you engage. My failure to fulfill your trust in me by performing some particular act does not demand that I be considered untrustworthy. Even if, as some might suggest, your trusting in me is a kind of prediction or anticipation of me it cannot be a prediction about my behavior because then we would not be able to make distinctions like the ones made earlier; that is, we would not be able to claim that an untrustworthy person can, at times, behave in a trustworthy manner. In addition, a person who has never failed or betrayed our trust can still be distrusted, which would be nonsensical if trusting were strictly about the behaviors of others; trust as behavior should demand that one who has never failed a trust should be trusted. One's status as trustworthy or untrustworthy, therefore, is not merely an aggregate or even a summation of the behaviors that she has engaged in, it must be something else.

Trust is not about action. James Faulkner, even though he argues for trust as a kind of interpretive model like the one I suggest, still holds tightly to the idea that trusting is an action. In referring to the action of an employer hiring one who has been previously convicted of theft, he says "she simply chooses to trust."[313] The trust of Faulkner's position is not the trust of a belief about the motivations of another but is, rather, the

[312] Hardin, *Trust and Trustworthiness*, 10.
[313] Faulkner, "On Telling and Trusting," 885.

action of one individual in relation to the expected actions of another individual. Separating trusting from trustworthiness, by showing that trust is not about the actions which we engage in, is important for recognizing that trust is about motivation, which leads to the conclusion that trust is interpretive.

Trust Is Not About Objective Certainty, It Is Risky

The actions of human beings are not knowable independently of the meaning that we ascribe to them. Instead, trust shapes our understanding of the actions of others. Bernd Lahno claims that "because trust somehow determines how we think, it cannot be understood as the immediate result of rational consideration."[314] Trust emerges from the uncertainty that we have about people's motivations and it leads to uncertainty about their actions. This inherent uncertainty means that trust is risky. This claim is, like many of the points that I argue for, contentious but its contentiousness is caused by an equivocation on the term 'risk'. Risk can mean harm or it can mean uncertainty. When Hardin argues that trust is not risky he bases his position on the idea that only things that we can choose can be risky; "I do not, in an immediate instance, choose to trust, I do not take a risk."[315] Since trust is, for him, knowledge or a kind of belief it cannot be the subject of choice and it cannot cause harm, only the actions that we undertake can harm us. If Hardin were correct and I wanted to easily make my claim that trust is risky then it seems that I should argue that trust is, indeed, an action in order to show this. Perhaps unfortunately, I am adamant, like Hardin, that trust is not an action so that route is unavailable to me. However, I differ from Hardin by considering a fundamental principle

[314] Bernd Lahno, "On the Emotional Character of Trust," *Ethical Theory and Moral Practice* 4, no. 2 (2001): 177.
[315] Hardin, "Conceptions and Explanations of Trust," 12.

of trust to be that it is uncertain. Trusting necessarily involves uncertainty because certain knowledge of the behavior or motivations of another would actually preclude trusting. Onora O'Neill argues that when "we have guarantees or proofs, placing trust is redundant."[316] Annette Baier also points out that certainty makes trust redundant; she says that "the special vulnerability which trust involves is vulnerability to not yet noticed harm."[317] Were we able to be certain of all the harms that might confront us we would be able to erect barriers to those harms and so trust would be unnecessary. Niklas Luhman's entire account of trust is premised on an attempt to reduce the complexity of our uncertain future; "[t]he problem of trust therefore consists in the fact that the future contains far more possibilities than could ever be realized in the present."[318] Alya Guseva and Akos Rona-Tas support the idea that uncertain is necessary component of an account of trust when they describe trust as "positive expectations in the face of uncertainty emerging from social relations."[319] Uncertainty is ever-present, so too is trust.

The certainty that a bridge will fall or remain standing when a specific amount of weight is placed on it is analogous to the problem of certainty in regards to trusting others. However, we must be careful in our analogies else we admit too much; inanimate objects can never be, on my account, trustworthy. If I knew with certainty that the bridge would fail or survive my weight then I would not need to 'trust' in it. A certain bridge collapse would lead me to avoid crossing the bridge while a certain bridge stability would lead me to cross unquestioned as if I were walking on solid ground. Likewise, if I knew

---

[316] O'Neill, *A Question of Trust*, 6.
[317] Baier, *Moral Prejudices : Essays on Ethics*, 104.
[318] Luhmann, *Trust and Power : Two Works*, 13.
[319] Alya Guseva and Akos Rona-Tas, "Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared," *American Sociological Review* 66 (2001): 627.

with certainty that you would fail or succeed in some instances then the thing that we recognize as trust would be irrelevant to our interaction. If it were certain that you would fail in regards to the trust that I placed in you then my trust in you would be irrelevant. Absolute certainty destroys, rather than enhances, our ability to trust. Linda Molm *et al.* reinforces this point when they argue in their paper that truly binding agreements negate the need for trust;"[i]f, in addition, the exchange is secured with conditions that make the agreement binding…the actors face no risk that the other party will not honor the terms of the agreement."[320] It is important to note that although my example leans heavily on the success of failure of the trusting, which is attached to the action, the certainty in regards to trusting is actually in reference to your motivations for acting and not the acting itself.

Uncertainty is important to the interpretive account of trust. Because there is uncertainty about the meaning of actions and about the motivations of another the evidence that confronts us in a given situation can be understood in multiple ways. If it were certain that the bridge would remain standing then the creaking noises it made as I placed my weight upon it would be nothing but a sign of the dryness of the wood, which would not affect my crossing. On the other hand, if it were certain that the bridge would fail then the creaking would be evidence that failure was eminent, which would be a reminder to turn back. If I were uncertain about the stability of the bridge then the creaking of the wood could be an indication of several things, determining between them would require an assessment on my part of what the creaking actually meant.

[320] Linda D. Molm, Nobuyuki Takahashi, and Gretchen Peterson, "Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition," *American Journal of Sociology* 105 (2000): 1401.

Uncertainty about your motivations allows the meaning of your actions to be open to interpretation. Your decision to defect in the Prisoner's Dilemma might be a sign that you distrust me or it might be a sign that you are not worthy of my trust. Or, to complicate the matter further, you might think that defection is actually helping me. By defecting, you give me good reason to defect against you. You might believe that this is the best thing for me; perhaps we have been friends since childhood and you have come to understand that you have been a weight that is dragging me down. If I did not have to spend so much time and effort getting you out of trouble I could have the successful life that you think I deserve. You might therefore, act for all the world as if you are about to defect against me in order to get me to sever our relationship but this act would not be an act of distrust on your part but instead a great sacrifice for friendship; it would be your attempt to be worthy of my trust.

Uncertainty about motives is not, however, the end of a discussion of the risk inherent in trust. Harm can be caused by our beliefs; our beliefs open and close doors, they control which avenues of action are available by either enhancing or limiting our opportunities. And, our beliefs, when consistently reinforced, shape the way we view ourselves and the world. A child that has been abused has great difficulty forming close relationships with others because the child believes either that others are inherently unworthy of trust or that he is not worthy of others being trustworthy or trusting to him.

If I believed that it was certain that the bridge would collapse then I could not consider crossing the bridge as a viable route for my progress, my beliefs about the bridge would drastically affect what choices were available to me. In *Romancing the Stone*, Kathleen Turner's character attempts to cross a bridge that Michael Douglas'

character had already eliminated as a possible means of escape.[321] Her uncertainty about the strength of the bridge allowed her to proceed where his certainty did not. However, her uncertainty about the motivations of his character limited the choices available to her when she first got into trouble; she did not fully confide in him until much later in the movie. Uncertainty is a double edged sword, it is necessary for some beliefs while it precludes others but its lack leaves no room for interpretation and therefore no room for trust. In short, we need uncertainty in order to live in the world in any way other than simple reflexive machines. The ubiquitous-ness of uncertainty leads to trust being a primary mode of understanding; uncertainty leads us to interpret the behaviors of others, which can only be done through the lens of our trust.

Robert Solomon and Fernando Flores remind us that "trust is not the same as predictability, and without unpredictability, there is no role for trust."[322] Just as there is no way to be certain about the motives of an agent in the incredibly simple scenario of two prisoner's being interrogated separately (in which there are only two possible choices – cooperate or defect) so too is there no way to be certain about the motives of an agent in the real world (in which there are many possible choices) let alone an entity like the United States Government (in which there is a near infinite combination of possible choices). The uncertainty of an agent's motives has already been shown.

The Soviets, given a certain set of beliefs, were provided ample evidence to lead them to suspect that the United States was preparing to invade. In the ABLE ARCHER scenario there was a great deal of uncertainty about the motivations of others, with little

[321] Robert Zemeckis, "Romancing the Stone," (United States: Twentieth Century-Fox Film Corporation, 1984).
[322] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*, 70.

recognition of that uncertainty. Where there was recognition of the uncertainty it was believed that we should err on the side of caution, which, in this context, was a defensive stance – defection in the Prisoner's Dilemma. Our beliefs preclude us from holding certain, alternative beliefs; they blind us to other options. President Reagan did not imagine, prior to receiving security briefings about the Soviet response to our war games, that the Soviets honestly believed we wanted to destroy them. He thought it obvious to all that America was not going to aggressively invade another nation but his interpretation of the political climate nearly started World War III. To claim that it is only our actions that can harm us grossly underestimates just how much our actions, and our response to the actions of others, are shaped by our beliefs.

How we interpret the behaviors and motivations of others can have a profound effect on our lives. Not only can the actions that follow from our beliefs harm us but so too can those beliefs. Risk is important to understanding trust because our uncertainty in the face of the liberty of others drastically limits the final configuration of our conception of trust.

Trust Is Not Expectation

Trusting is not merely a kind of expectation about how another will behave, however; "[e]xpectation is the state (outcome) at the end of the process."[323] If trust were simply an expectation then we would see trust as resulting from a kind of deliberation, either conscious or unconscious. And, although my position is that trust is self-informing, it is a mistake to imagine trust as the result of deliberation. As we have already seen, trust

[323] Guido Möllering, "The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension," *Sociology* 35, no. 2 (2001): 403.

is neither an action nor is it directly attributable to actions. Treating trust as an expectation, as an outcome, is to treat trust as being the same as the actions that we take. This is, perhaps, arguing too strongly. If we understand expectation in the generalized sense, like 'I expect great things of Jane' then expectation seems compatible with the account of trust that I am presenting. However, talk of expectations generally ends up being about specific expectations, like 'I expect Jane to be here at noon.' In this latter sense, trust and expectation are separate even though we often say colloquially that 'I trust Jane to be here at noon.' Used in this way, trust is a statement of an expectation about Jane's behavior and not of a belief about Jane's motivations.

We have seen that trust is not about actions and that because it is uncertain it requires us to interpret the motivations of others but the way that trust directs our interpretation is particular. Baier points out that we can rely on others' "dependable habits, or only their dependably exhibited fear, anger, or other motives compatible with ill will toward one, or on motives not directed toward one at all."[324] Her claim is that even in cases where we cannot trust another we can still form expectations about his behavior. That is, we can interpret the behavior of another, often reliably, based on a set of cues that is wholly separate and distinct from trust. Even though we hold tightly to the idea that we have the ability to freely choose our actions humans are surprisingly predictable. Baier is pointing out that many of the activities in which we seem to engage that seem to require trust can actually be covered by other predictive capabilities. I can, for the most part, depend on other drivers not to swerve between lanes randomly because I can rely on their fear of getting into a wreck or their fear of a ticket or, more simply,

[324] Baier, *Moral Prejudices : Essays on Ethics*, 98-99.

because it is their habit to stay in one lane. If trust were merely an expectation then it would be indistinguishable from any of these other ways of forming expectations. Unfortunately, when she claims that we "can still rely where we no longer trust" Baier confuses our actions with our motivations for action, which is a distinction that she herself makes and a confusion that we would do well to avoid.[325]

Baier argues that "[w]hen I trust another I depend on her goodwill toward me."[326] Depending on goodwill may coincide with the expectation that some other will behave in a certain way but the expectation cannot be a case of trusting. I have been arguing that it is the trust that we have in another that leads to the forming of expectations, which is to say that trust is not the outcome of the process but rather the beginning. But as a beginning, trust must direct the formation in some fairly specific way or else, even though it does direct expectations, it will continue to be indistinguishable from other processes. Believing that another has goodwill towards me causes me to form expectations about the behavior of another that are positive, at least as regards our relationship. Goodwill is important to human interaction. Baier is not alone in thinking this; famously, Immanuel Kant argues that "good will is good not because of what it effects or accomplishes, nor because of its fitness to attain some proposed end; it is good only through its willing."[327] Daryl Koehn suggests that "[s]howing good will, not getting advantage, is the mark of the true friend."[328] And, he quotes Motojiro Kawashima as saying "the actual value of social obligations depends upon the good will and favor of the

[325] Ibid., 98.
[326] Ibid., 99.
[327] Kant, *Grounding for the Metaphysics of Morals*, 7.
[328] Daryl Koehn, "What Can Eastern Philosophy Teach Us About Business Ethics?," *Journal of Business Ethics* 19, no. 1 (1999): 75.

obligated person."[329] Robert Solomon and Fernando Flores tell us that "[w]ithout caring, what is called trust is no more than prediction and reliance."[330] Trust directs our interpretation of the behaviors of others by focusing our concern on their motivations in reference to goodwill, specifically in reference towards their goodwill for us.

John Wisdom, who had possibly the best name in all philosophy, presents an argument about the effect of a belief in God, which he tries to separate from expectation in much the same way that I am trying to do with trust. He claims that what has been understood from any given scene "isn't merely a matter of 'the facts.'"[331] Wisdom tells a story of two people returning to a garden that has been untended for a significant amount of time but, much to their surprise, some of the plants in the garden are surviving quite well. In his example it is assumed that the plants that are doing well require tending and so to find them not merely alive but thriving after substantial neglect is a mystery. In Wisdom's example, one of the companions suggests that an invisible gardener has been tending to the plants while the other denies this possibility. He argues that the two companions could proceed with a very serious investigation of the garden to determine whether a gardener was involved or not. They could even agree about all of the facts of the matter; that the garden had been abandoned, that gardens deteriorate at a certain rate of speed, and that plants do not tend to themselves in ways that resemble human tending. The companions can agree on everything they see in the garden and yet, he argues, "the

[329] Ibid.: 76.
[330] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*, 105.
[331] J. Wisdom, "Gods," *Proceedings of the Aristotelian Society* 45 (1944): 190-91.

difference is in a sense no one as to the facts."[332] The two companions are interpreting the facts of the garden differently based on how they feel.

For my part, the interpretive scheme of trust is based on how we feel about the person who is acting but I agree with Wisdom that this interpretation is not "a matter of collecting from several inconclusive items of information an expectation as to something further."[333] Trusting is not an expectation about how another will act it is a belief in the goodwill of the other that informs and shapes our understanding of her past actions and our expectations of her future actions.

Trust Is Not About the Object of Trust

There is one final point to be made about the interpretive model of trust: the interpretation is not dependent on the purported object of that trust. Bernd Lahno provides a useful example here; he imagines a mother who has left her daughter in the care of a neighbor and then asks us to compare that situation "with that of a burglar, who has observed the situation and who now grasps the opportunity to rob the neighbor's home undisturbed."[334] Lahno points out that both the mother and the burglar can see this situation in a way that brings them benefit. The mother gains the benefit of having someone look after her child, allowing her to go off and do some other activities without feeling the concern for her child's well-being. The burglar gains the benefit of having a person in the house who is unfamiliar with the setting and one who is, in essence, incapable of stopping him from robbing the place. Lahno's point is that the motivation that is attached to both the burglar and the mother is very different in this scenario; both

[332] Ibid.: 195.
[333] Ibid.: 194.
[334] Lahno, "On the Emotional Character of Trust," 178.

the mother and the burglar weigh the costs and the benefits of the baby-sitter taking care of the child and both find it to be favorable. The benefit to be gained to each party is therefore obvious but, as Solomon and Flores point out, "trust is never the result of some cost-benefit calculation."[335] From the outset this example seems problematic but for my purposes the important aspect of this example concerns the object of trust and not the apparent cost-benefit calculation that Lahno appears to believe drives trusting. Lahno argues that the "burglar does not believe in the goodwill of the neighbor towards himself" and yet he argues that the burglar is able to trust in this situation.[336] This discussion comes is after a direct reference to Baier, whose account Lahno is criticizing.

"Based on her knowledge about the relationship between her neighbor and her daughter, she can be fairly certain that the neighbor will take care of the girl."[337] Lahno's argument, in this case, is extremely problematic given the position that he has sketched so far; he suggests that the mother may not personally have much interest or care for the neighbor but that the mother can recognize that the neighbor cares for her daughter such that the mother can trust the neighbor; his suggestion is that it is not care and concern for you that allows one to trust but instead care and concern for your interests. Lahnoe claims that you can trust another when that other shares your interest. He takes the focus on goodwill to be undirected. That is, he takes it to apply to any object that we might have concern for. Baier's account of goodwill is one in which the goodwill is directed at the self; I believe that you have goodwill towards *me*. But Lahno is expanding it to mean that I can believe that you have goodwill towards my interests. This is a seemingly minor

[335] Koehn, "What Can Eastern Philosophy Teach Us About Business Ethics?," 74.
[336] Lahno, "On the Emotional Character of Trust," 179.
[337] Ibid.: 180.

change; to have goodwill towards me is to be concerned for my interests, which seems to include having goodwill towards my interests but this surely does not follow. I can want the best for you without wanting the best for your drug habit and I can want the best for your daughter without being at all concerned for your interests.

Lahno explicates Wisdom's account of faith, which he uses as an analogy for trust, but he seems to miss the point of such an account. The mother cannot trust the neighbor unless she [the mother] believes that the neighbor actually has goodwill towards her. On Wisdom's account, recognizing that the neighbor cares for the daughter is not an objective fact to be known by the mother but is, instead, something that depends on the way the mother interprets the behaviors of the neighbor. On my account of trust this happens in exactly the way that Wisdom describes the effect of faith; the facts of the situation are viewed differently depending on how we interpret the scene. If we were to try to describe the situation in a neutral way we would say things like 'the neighbor always pays attention to the daughter,' or 'the neighbor expresses interest in the daughter,' or even 'the neighbor always gives presents to the daughter.' But, as these examples show, without a positive view of the nature of the neighbor's interest, those statements can tell a very different story. Imagine instead a mother who is suspicious of the neighbor. From that perspective, the concern and interest shown by the neighbor for the child could support a belief that the neighbor has reprehensible, even horrendous, aims for the child. Therefore, Lahno's conception that the mother can trust the neighbor because the mother believes that the neighbor has good intentions towards the child is implausible. Believing that the neighbor is concerned with the daughter's interest is not enough to generate trust because the daughter's interests might include running away or

raiding the liquor cabinet. Under those conditions the mother would instead believe that the neighbor's intentions towards the daughter are such that it would hurt the mother's interest.

Amusingly, the movie *Uncle Buck* can be a vivid example of this.[338] Buck, played by John Candy, is a slovenly, out of work, cigar smoking, bachelor who becomes guardian of his brother's children when their parents must abruptly leave town. The children's mother is terrified of leaving the children in Buck's care; at several points in the story she bursts into tears at the thought of Buck in charge. But, she knows that Buck cares about the children. This might seem to indicate that the capacity to succeed at the task at hand is important because it might indicate that she worries about Buck's capacity to be an adult and not his concern for the children. However, I argued in Chapter 5 that success is not an immediate concern for trust. What I intend it to show with this example is that the mother's belief that Buck cares and has goodwill towards the children is not enough to make her trust him. Buck's sister-in-law, Cindy, knows that he is concerned for her children. Further, he has made all the appropriate verbal promises to suggest that he is aware of his obligations. And yet, Cindy fails to trust Buck – this is the crux of my position about trust. There does not seem to be any way around this problem: actually trusting another requires you to hold the belief that the other has goodwill towards your interests because they are your interests and not for some other set of reasons, you cannot will yourself to trust him. Other reasons were not enough to make Cindy trust Buck and it will not be enough to make the mother in Lahno's example trust her neighbor. But, both

[338] John Hughes, "Uncle Buck," (United States: Universal Pictures, 1989).

mothers can, and in fact one does, *entrust* the well-being of her children even under the circumstances in which she does not fully trust.

Entrusting does have an object. It always has an object. I must entrust *something* to another, otherwise there would be no act of entrusting. But trusting does not have this feature. Cindy entrusted her children to Buck but she did not trust Buck. Since trusting is a belief in the goodwill of another its presence or absence is independent of any object that might be entrusted to that person. If trust had a third part then it would not be interpretive, the interpretation has already taken place when the objects that I will entrust to another begin to be delineated. Trust, were it to consist in three parts, would be the result of deliberation not the start of it.

Reciprocal

Once we accept that trust allows us to interpret the behaviors of others we are confronted with trust's reflexivity. Trust is, itself, informed by our interpretations of the behaviors of others. Since the behaviors of others are an indication of their motivations those actions influence whether or not we believe that they have good will towards us. In more practical terminology, Mary's actions are evidence for whether or not she is concerned for my well-being. This is complicated by the interpretive role of trust because even though Mary's actions influence my assessment of Mary, it is my trust of Mary that allows me to interpret her actions in the first place. Because I am trusting of Mary, I interpret her actions as being supportive of my interests and those actions reinforce my belief that Mary is concerned with my interests. This will seem problematic to some because it describes a trust relationship which is, for the most part, circular, one lacking

in an obvious foundation or starting point. But what is important to understand about my position is that we all already stand in a relationship of trust to each other. Being human, *qua* human we have a base of interaction with others, which is to say that we being with some interpretive structure for understanding the behavior of others. Discussing the ethics of Watsuji Tetsuro, Daryl Koehn says "[t]rust exists because we are all always already related to each other in a variety of ways."[339] Returning to our examples will help to clarify this idea.

Earlier we considered how a belief that another has good will towards you limits your interpretation of his behavior to a positive set. Barney and Barbara, married for 10 years with several children, have settled into a fairly routine and simple life. Then, one day, Barney starts coming home form work late. When asked about his late arrival he always has an excuse. If Barbara trusts Barney than she accepts his excuses and, perhaps, will not even be concerned when he fails to come home without calling. Further, Barney might even come home with lipstick on his collar, smelling of women's perfume, and looking disheveled without raising the suspicion of one who trusts him. Barbara's trust of Barney will direct her understanding of his behavior towards those interpretations that support the belief that Barney is worthy of trust. But this positive interpretation cannot be maintained indefinitely in the face of externally suspicious behavior, nor can it be maintained in the face of particularly egregious violations of trust. Should Barbara ever find Barney in bed with another woman, especially at a time when he has claimed that he would be busy at work, the available interpretations is already limited in such a way that

[339] Koehn, "What Can Eastern Philosophy Teach Us About Business Ethics?," 74.

a positive spin would be difficult to produce. That is, some acts are such blatant violations of trust that trusting cannot be maintained when confronted by them.

Barney's philandering will cause Barbara to no longer trust him. This is an example of how the trust we have for another can be influenced by the actions of that other. This is where many will see a problem with my position. It appears, from the way that it has been presented, that the act of infidelity is bad in and of itself, that it is an objectively bad thing, which contradicts my earlier claims about the need to interpret all actions. This would be true if I were arguing that Barney's indiscretion was to blame but I am not. Instead, it is that Barney's indiscretion violates the trust relationship he shares with Barbara that makes his action so deplorable. Barney's behavior, by itself, might not have been bad; they might have an understanding about Barney's trysts such that they are acceptable under certain conditions. This would mean that Barney's failure to be worthy of trust in the example came not from the fact that he slept around but instead from the fact that he claimed he would be busy at work. In this case, it is the lie that broke the trust. Those who think that I have this all wrong will probably want to point out that this appears to only back the question up; instead of being about the act of sleeping around it is about the act of lying. But again, the breach of trust is constituted by the relationship of trust; what counts as a violation of trust will depend on the nature of our relationship and not on any particular act: the lie is only a breach of trust if it is embedded in a trust relationship already. Trust is, therefore, informed by the actions that it interprets.

When Barbara catches Barney in his lie this challenges the interpretive scheme that she had been using. Not only will she look at his future actions differently, she will also re-evaluate his past actions. She had, perhaps unquestioningly, accepted his

explanations for coming home late in the past but now, armed with a new interpretive scheme, she begins to doubt those instances. If she finds more evidence that he has lied, she will go further than doubting, she will begin to view those instances as cases in which he was also out with the other woman. What had been trust becomes distrust and the distrust leads to finding further evidence to distrust. Trust is self-informing.

Consider instead the Cuban Missile Crisis. Kennedy was uncertain yet positive about the motivations of Khrushchev while his military advisors were uncertain and significantly less positive about Soviet motivations; Robert Kennedy tells us of a moment when one "of the Joint Chiefs of Staff…argued that we could use nuclear weapons, on the basis that our adversaries would use theirs against us in an attack."[340] Distrust of the Soviet Union and Premiere Khrushchev in particular was leading President Kennedy's advisors to suggest some of the worst possible motivations and therefore to predict some of the worst possible behaviors. Kennedy had to be open to the possibility that the Soviets were motivated by something other than malice in order to proceed without attacking. Had he believed that the Soviets had, in fact, intended us harm, he would have had no choice but to attack.[341] We can postulate that Kennedy, rather than relying on any direct knowledge of Khrushchev's motivations, was instead coming to his trust on the more basic assumption that humans do not generally intend to harm other humans. The Prisoner's Dilemma contingent will suggest that Kennedy's response is more easily explained by pointing out that given the stakes of the scenario any defection would be

[340] Kennedy, *Thirteen Days : A Memoir of the Cuban Missile Crisis*, 38.
[341] I must admit that it is entirely possible that could believe that they intended us harm and yet choose not to attack. If the rumors are true, this is essentially what FDR did prior to the start of World War II. But this point just further reinforces the separation between trust and the acts that follow from it. I am content, therefore, to let this anomaly remain.

disastrous. The horror of nuclear annihilation is summed up nicely by Joshua, the computer from *War Games*, when he says "Silly game. The only way to win is not to play."[342] However, suggesting that Kennedy was choosing to act based on the Prisoner's Dilemma forgets that he would still have to interpret the behavior of the Soviet's in order to employ that strategy. The interpretation remains regardless of whether the Prisoner's Dilemma is used as a tool or not and the interpretation must be reevaluated based on new evidence. Kennedy had a particular interpretation and he was rewarded for that interpretation when the Soviets withdrew.

The way in which I am describing the reflexivity of trust is similar to the learning account of trust presented by Russell Hardin. He argues that our previous experience colors how we will react to any present concern; "those who start life badly are disadvantaged by the continuing loss of welfare in forgone opportunities from low capacity for trust."[343] His position highlights the importance of having positive experiences early in life. If I am confronted at a young age by interactions in which trust is broken then my experience will teach me not to trust as often. When presented with the same kind of evidence in subsequent interactions I will not trust because trusting has been shown to not work under these conditions. If however, I have predominately positive experiences as a youth then I will more readily trust those that I encounter. Expanding on his account we can see how this can lead us to have different responses to the same evidence depending on our previous life experiences.

Hardin compares the return on interaction from both a high capacity for trusting and a low capacity for trusting. A person with a high capacity to trust will enter into more

[342] John Badham, "Wargames," (United States: United Artists, 1983).
[343] Hardin, *Trust and Trustworthiness*, 119.

relationships than a person with a low capacity. But the high capacity truster will more often be let down by trusting whereas the low capacity truster will rarely be let down. This, though it sounds bad, actually provides the high capacity truster with more opportunity to adjust the ease with which they trust and so will more quickly approach the optimum level of trust.[344] The low capacity truster, on the other hand, will have few opportunities to adjust the ease with which they trust. Hardin recognizes that his model fails to account for either the size of the potential loss or the importance that we might place on specific interactions. Some losses might be more severe than others and might therefore cause of to become much more wary should they go wrong. Likewise, interacting with certain individuals might have a much greater impact on our lives; the interactions that we have with our parents should have more affect on us that the interactions we have with strangers, even adjusting for the greater frequency of interactions with our parents. He also recognizes that there might be strategic reasons for "as-if trust behavior"[345] that the model does not account for. But even with these and other drawbacks Hardin feels that the model manages to capture all the important elements of an account of trusting.

His learning account of trust gives us a good basis for understanding the problems faced by those who are 'at-risk' in regards to trust relations. That is, those who are most need to benefits of trusting others are least likely to trust others; those who have been abused or taken advantage of will be less likely to open themselves up to the possibility of being hurt again. It produces a potentially vicious cycle of distrust. But this knowledge should help us to overcome the problem. We know that these individuals will require

[344] Ibid., 122.
[345] Ibid., 124.

more evidence from us before they can participate in a trust relation. Insofar as trust is cognitive the knowledge that the capacity to trust is acquired in this way should be useful to those who are low-capacity trusters; they should be able to recognize their deficiency and compensate for it. This gives hope where trust is concerned. Even if one is disinclined to trust it "can be cultivated. It can be initiated, generated, and repaired."[346]

Conclusion

So, there it is. We have reached the conclusion that trust is not so much a thing or an attitude but it is instead rather more like a way of being. We can now re-envision Baier's claim that "[w]e inhabit a climate of trust as we inhabit the atmosphere and notice as we notice air, only when it becomes scare or is polluted."[347] We can understand this to mean not that there is, in actuality, some atmosphere of trust but instead that it is as Daryl Koehn suggests: "[a]nyone who is human trusts simply by virtue of being human."[348] We live in an environment of trust. We are always in relationships with others, we view them in particular ways that allow us to interpret their behavior to have meaning and import to our lives. Those behaviors inform our beliefs about others and cause those beliefs to evolve and change. When we interpret the behavior of another in a positive way, one that supports our interests, we are trusting him. When we interpret his behavior in a negative way, we are distrusting him.

On this account trust is of paramount importance to human interaction because it affects every relationship we have and every decision we make. It is the framework we have for understanding the world and as such it is a precursor to making ethical

[346] Solomon and Flores, *Building Trust in Business, Politics, Relationships, and Life*, 107.
[347] Baier, *Moral Prejudices : Essays on Ethics*, 98.
[348] Koehn, "What Can Eastern Philosophy Teach Us About Business Ethics?," 74.

judgments. Recognizing this will allow us to take much more care in our relationships and will allow us to begin to build positive connections with others. Institutions can be structured to support, rather than deny, the importance of human interaction and human agency. Individuals can be encouraged to expand the interpretations that the have of the behaviors of others, which will allow them to be more trusting. There is a pithy way to end this, I am sure, but I would rather leave the end open, a kind of unfinished sentence that exhibits the ongoing trust relationship that we all share.

References

Adler, Jonathan E. "Testimony, Trust, Knowing." *The Journal of Philosophy* 91, no. 5 (1994): 264-75.

Allais, M. "Le Comportement De L'homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes De L'ecole Americaine." *Econometrica* 21, no. 4 (1953): 503-46.

Aristotle. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis, Ind.: Hackett Pub. Co., 1985.

Axelrod, Robert M. *The Evolution of Cooperation*. New York: Basic Books, 1984.

Bacharach, Michael, Gerardo Guerra, and Daniel John Zizzo. "The Self-Fulfilling Property of Trust: An Experimental Study." *Theory and Decision* 63, no. 4 (2007): 349-88.

Badham, John. "Wargames." 114 min. United States: United Artists, 1983.

Baier, Annette. *Moral Prejudices : Essays on Ethics*. Cambridge, Mass.: Harvard University Press, 1994.

Barbalet, Jack. "Trust and Uncertainty: The Emotional Basis of Rationality." In *SCARR Network Trust Conference*. London, 2005.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Dover ed. Mineola, N.Y.: Dover Publications, 2007.

Bhattacharya, Rajeev, Timothy M. Devinney, and Madan M. Pillutla. "A Formal Model of Trust Based on Outcomes." *The Academy of Management Review* 23, no. 3 (1998): 459-72.

Blackburn, Simon. *Being Good : An Introduction to Ethics*. Oxford: Oxford University Press, 2001.

Bok, Sissela. *Lying : Moral Choice in Public and Private Life*. 2nd Vintage Books ed. New York: Vintage Books, 1999.

Carolan, Michael S. "Risk, Trust and 'the Beyond' of the Environment: A Brief Look at the Recent Case of Mad Cow Disease in the United States." *Environmental Values* 15, no. 2 (2006): 233-52.

Castelfranchi, Cristiano, and Rino Falcone. "Social Trust: A Cognitive Approach." In *Trust and Deception in Virtual Societies*, edited by Cristiano Castelfranchi and Yao-Hua Tan, 55--90. Dordrecht ; Boston: Kluwer Academic Publishers, 2001.

Dawkins, Richard. *The Selfish Gene*. 30th anniversary ed. Oxford ; New York: Oxford University Press, 2006.

Deutsch, Morton. "Trust and Suspicion." *The Journal of Conflict Resolution* 2, no. 4 (1958): 265-79.

Doris, John M. *Lack of Character : Personality and Moral Behavior*. Cambridge, U.K. ; New York: Cambridge University Press, 2002.

Eastwood, Clint. "Heartbreak Ridge." 130 min. United States: Warner Bros., 1986.
Faulkner, Paul. "On Telling and Trusting." *Mind* 116, no. 464 (2007): 875-902.
Fricker, Elizabeth. "Second-Hand Knowledge." *Philosophy and phenomenological research.* 73, no. 3 (2006): 592.
Fukuyama, Francis. *Trust : The Social Virtues and the Creation of Prosperity*. New York: Free Press, 1995.
Gambetta, Diego, ed. *Trust : Making and Breaking Cooperative Relations*. New York, NY, USA: B. Blackwell, 1988.
Gibson, Kevin. "Games Students Play: Incorporating the Prisoner's Dilemma in Teaching Business Ethics." *Journal of Business Ethics* 48, no. 1 (2003): 53-64.
Giffin, Kim. "Interpersonal Trust in the Helping Professions." *The American Journal of Nursing* 69, no. 7 (1969): 1491-92.
Glass, Bentley. "The Ethical Basis of Science." *Science* 150, no. 3701 (1965): 7.
Gleick, James. *Chaos : Making a New Science*. New York, N.Y., U.S.A.: Viking, 1987.
Goodin, Robert E. *Protecting the Vulnerable : A Reanalysis of Our Social Responsibilities*. Chicago: University of Chicago Press, 1985.
Govier, Trudy. *Dilemmas of Trust*. Montreal ; Ithaca: McGill-Queen's University Press, 1998.
———. "Trust, Distrust, and Feminist Theory." *Hypatia* 7, no. 1 (1992): 16-33.
Guseva, Alya, and Akos Rona-Tas. "Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared." *American Sociological Review* 66 (2001): 623-46.
Hardin, Russell. "Conceptions and Explanations of Trust." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xxviii, 403 p. New York: Russell Sage Foundation, 2001.
———. "Conceptions and Explanations of Trust." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xiii, 253 p. New York: Russell Sage Foundation, 2005.
———. *Trust and Trustworthiness*, The Russell Sage Foundation Series on Trust ;. New York: Russell Sage Foundation, 2002.
———. "Trustworthiness." *Ethics* 107, no. 1 (1996): 26-42.
Hayenhjelm, Madeleine. "Trusting and Taking Risks : A Philosophical Inquiry." Royal Institute of Technology, 2007.
Held, Virginia. "On the Meaning of Trust." *Ethics* 78, no. 2 (1968): 156-59.
———. "Rationality and Social Value in Game-Theoretical Analyses." *Ethics* 76, no. 3 (1966): 215-20.
Hieronymi, Pamela. "The Reasons of Trust." *Australasian Journal of Philosophy* 86, no. 2 (2008): 213-36.
Hill, Claire A., and Erin A. O'Hara. *A Cognitive Theory of Trust*: SSRN, 2005.
Hobbes, Thomas, and J. C. A. Gaskin. *Leviathan*, The World's Classics. Oxford ; New York: Oxford University Press, 1996.
Hughes, John. "Uncle Buck." 100 minutes. United States: Universal Pictures, 1989.
Kant, Immanuel. *Ethical Philosophy : The Complete Texts of Grounding for the Metaphysics of Morals and Metaphysical Principles of Virtue*. Indianapolis, IN: Hackett Pub., 1994.

———. *Grounding for the Metaphysics of Morals*. Translated by James W. Ellington. 3rd ed. Indianapolis: Hackett Pub. Co., 1993.
———. *Perpetual Peace, and Other Essays on Politics, History, and Morals*. Translated by Ted Humphrey. Indianapolis: Hackett Pub. Co., 1983.
Kennedy, John Fitzgerald. "Radio and Television Report to the American People on the Soviet Arms Buildup in Cuba." http://www.jfklibrary.org/Historical+Resources/Archives/Reference+Desk/Speeches/JFK/003POF03CubaCrisis10221962.htm.
Kennedy, Robert F. *Thirteen Days : A Memoir of the Cuban Missile Crisis*. New York: W.W. Norton, 1969.
Kierkegaard, Søren. *Fear and Trembling*. Translated by Alastair Hannay, Penguin Classics. New York: Penguin Books, 1985.
King, Martin Luther. *Strength to Love*. 1st ed. New York,: Harper & Row, 1963.
Koehn, Daryl. "What Can Eastern Philosophy Teach Us About Business Ethics?" *Journal of Business Ethics* 19, no. 1 (1999): 71-79.
Kramer, Stanley. "On the Beach." 134 min. The United States of America: Lomitas Productions. Released by United Artists, 1959.
Kuhn, Steven. "Prisoner's Dilemma." The Metaphysics Research Lab Center for the Study of Language and Information Stanford University, http://plato.stanford.edu/archives/win2007/entries/prisoner-dilemma/.
Kydd, Andrew. "Trust, Reassurance, and Cooperation." *International Organization* 54 (2000): 325-58.
LaFollette, Hugh. *Personal Relationships : Love, Identity, and Morality*. Oxford, UK ; Cambridge, Mass.: Blackwell, 1996.
LaFollette, Marcel C. "The Evolution of The "Scientific Misconduct" Issue: An Historical Overview." *Experimental Biology and Medicine* 224, no. 4 (2000): 211-15.
Lahno, Bernd. "On the Emotional Character of Trust." *Ethical Theory and Moral Practice* 4, no. 2 (2001): 171-89.
Laundau, Les. "Counterpoint." In *Star Trek: Voyager*, 45 minutes. United States: Paramount Television, 1998.
Lucas, George. "Star Wars." (ca. 120 min.). United States: Twentieth Century-Fox Film Corporation : Lucasfilm Ltd., 1977.
Luhmann, Niklas. "Familiarity, Confidence, Trust." In *Trust : Making and Breaking Cooperative Relations*, edited by Diego Gambetta, 246 p. New York: B. Blackwell, 1988.
———. *Trust and Power : Two Works*. New York: J. Wiley, 1979.
MacIntyre, Alasdair C. *After Virtue : A Study in Moral Theory*. 2nd ed. Notre Dame, Ind.: University of Notre Dame Press, 1984.
Messick, David M., and Roderick Moreland Kramer. "Trust as a Form of Shallow Morality." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xxviii, 403 p. New York: Russell Sage Foundation, 2001`.
Möllering, Guido. "The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension." *Sociology* 35, no. 2 (2001): 403-20.

Molm, Linda D., Nobuyuki Takahashi, and Gretchen Peterson. "Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition." *American Journal of Sociology* 105 (2000): 1396-427.
Nee, Victor, and Jimy Sanders. "Trust in Ethnic Ties: Social Capital and Immigrants." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xxviii, 403 p. New York: Russell Sage Foundation, 2001.
O'Neill, Onora. *A Question of Trust*. Cambridge ;; New York: Cambridge University Press, 2002.
Parales-Quenza, Carlos. "Astuteness, Trust, and Social Intelligence." *Journal for the Theory of Social Behaviour* 36, no. 1 (2006): 39-56.
Plato, G. M. A. Grube, and John M. Cooper. *The Trial and Death of Socrates*. 3rd ed. Indianapolis, IN: Hackett Pub., 2000.
Plato, G. M. A. Grube, and C. D. C. Reeve. *Republic*. Indianapolis: Hackett Pub. Co., 1992.
Poundstone, William. *Prisoner's Dilemma*. 1st Anchor Books ed. New York: Anchor Books, 1993.
Pry, Peter Vincent. *War Scare : Russia and America on the Nuclear Brink*. Westport, Conn.: Praeger, 1999.
Quine, W. V., and J. S. Ullian. *The Web of Belief*. 2d ed. New York: Random House, 1978.
Schwarz, Frederick Charles. *You Can Trust the Communists*. Englewood Cliffs, N.J.,: Prentice-Hall, 1960.
Shermer, Michael. *The Science of Good and Evil : Why People Cheat, Gossip, Care, Share, and Follow the Golden Rule*. 1st ed. New York: Times Books, 2004.
Solomon, Robert C., and Fernando Flores. *Building Trust in Business, Politics, Relationships, and Life*. New York: Oxford University Press, 2001.
Spielberg, Steven. "Raiders of the Lost Ark." 115 min. United States: Paramount Pictures, 1981.
Swinth, Robert L. "The Establishment of the Trust Relationship." *The Journal of Conflict Resolution* 11, no. 3 (1967): 335-44.
Taylor, Richard. *Ethics, Faith, and Reason*. Englewood Cliffs, N.J.: Prentice-Hall, 1985.
Thompson, George. "Game Theory And "Social Value" States." *Ethics* 75, no. 1 (1964): 36-39.
Tullock, Gordon. "The Prisoner's Dilemma and Mutual Trust." *Ethics* 77, no. 3 (1967): 229-30.
Wagner, Richard. *The Rhinegold*. Translated by H. and F. Corder, Grand Opera Librettos. Boston: O. Ditson Co., 1904.
Weber, Michael. "The Resilience of the Allais Paradox." *Ethics* 109, no. 1 (1998): 94-118.
Wilson, Edward O. *Sociobiology : The New Synthesis*. Cambridge, Mass.: Belknap Press of Harvard University Press, 1975.
Wisdom, J. "Gods." *Proceedings of the Aristotelian Society* 45 (1944): 185-206.
Wolff, Robert Paul. "Reflections on Game Theory and the Nature of Value." *Ethics* 72, no. 3 (1962): 171-79.

Wollstonecraft, Mary. *A Vindication of the Rights of Men*, Great Books in Philosophy. Amherst, N.Y.: Prometheus Books, 1996.
Yamagishi, Toshio, Satoshi Kanazawa, Rie Mashima, and Shigeru Terai. "Separating Trust from Cooperation in a Dynamic Relationship: Prisoner's Dilemma with Variable Dependence." *Rationality and Society* 17, no. 3 (2005): 275-308.
Yates, J. Frank, ed. *Risk-Taking Behavior*. New York: Wiley, 1992.
Young, Iris Marion. "Displacing the Distributive Paradigm." In *Ethics in Practice : An Anthology*, edited by Hugh LaFollette, xvi, 745 p. Malden, MA: Blackwell Pub., 2007.
Zemeckis, Robert. "Romancing the Stone." 105 minutes. United States: Twentieth Century-Fox Film Corporation, 1984.

Bibliography

Adler, Jonathan E. "Testimony, Trust, Knowing." *The Journal of Philosophy* 91, no. 5 (1994): 264-75.

Aristotle, Stephen Everson, Aristotle, and Aristotle. *The Politics, and the Constitution of Athens*. Rev. student ed, Cambridge Texts in the History of Political Thought. Cambridge ; New York: Cambridge University Press, 1996.

Audi, Robert. "Faith, Belief, and Rationality." *Philosophical Perspectives* 5 (1991): 213-39.

Austin, J. L., J. O. Urmson, and G. J. Warnock. *Philosophical Papers*. 3d ed. Oxford [Eng.] ; New York: Oxford University Press, 1979.

Austin, J. L., and G. J. Warnock. *Sense and Sensibilia*. Oxford,: Clarendon Press, 1962.

Bacharach, Michael, and Gambetta Diego. "Trust as Type Detection." In *Trust and Deception in Virtual Societies*, 1-26: Kluwer Academic Publishers, 2001.

Baier, Annette. *Postures of the Mind : Essays on Mind and Morals*. Minneapolis: University of Minnesota Press, 1985.

———. *A Progress of Sentiments : Reflections on Hume's Treatise*. Cambridge, Mass.: Harvard University Press, 1991.

Bantz, David A. "The Philosophical Basis of Cost-Risk-Benefit Analyses." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (1982): 227-42.

Barbalet, Jack. "Trust and Uncertainty: The Emotional Basis of Rationality." In *SCARR Network Trust Conference*. London, 2005.

Barber, Bernard. *The Logic and Limits of Trust*. New Brunswick, N.J.: Rutgers University Press, 1983.

Barney, Jay B., and Mark H. Hansen. "Trustworthiness as a Source of Competitive Advantage." *Strategic Management Journal* 15, no. SPI (1994): 175.

Beah, Ishmael. *A Long Way Gone : Memoirs of a Boy Soldier*. 1st ed. New York: Farrar, Straus and Giroux, 2007.

Becker, Lawrence C. "Trust as Noncognitive Security About Motives." *Ethics* 107, no. 1 (1996): 43-61.

Bhattacharya, Rajeev, Timothy M. Devinney, and Madan M. Pillutla. "A Formal Model of Trust Based on Outcomes." *The Academy of Management Review* 23, no. 3 (1998): 459-72.

Bickford, Susan. "Beyond Friendship: Aristotle on Conflict, Deliberation, and Attention." *The Journal of Politics* 58, no. 2 (1996): 398-421.

Birmingham, Robert L. "The Prisoner's Dilemma and Mutual Trust: Comment." *Ethics* 79, no. 2 (1969): 156-58.

Blair, Margaret M., and Lynn A. Stout. "Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law." *University of Pennsylvania Law Review* 149 (2001): 1735-810.
Bo, Pedro Dal. "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *The American Economic Review* 95, no. 5 (2005): 1591-604.
Braithwaite, V. A., and Margaret Levi, eds. *Trust and Governance*, The Russell Sage Foundation Series on Trust ;. New York: Russell Sage Foundation, 1998.
Buck, Rinker. *Flight of Passage*. 1st ed. New York: Hyperion, 1997.
Byron, Michael. "Rationality Is Not Fair." *Analysis* 55, no. 4 (1995): 252-60.
———. "Satisficing and Optimality." *Ethics* 109, no. 1 (1998): 67-93.
Childress, James F. "Nonviolent Resistance: Trust and Risk-Taking." *The Journal of Religious Ethics* 1 (1973): 87-112.
Chiles, Todd H., and John F. McMackin. "Integrating Variable Risk Preferences, Trust, and Transaction Cost Economics." *The Academy of Management Review* 21, no. 1 (1996): 73-99.
Cocking, Dean, and Jeanette Kennett. "Friendship and the Self." *Ethics* 108, no. 3 (1998): 502-27.
Cole, Richard L. "Toward as Model of Political Trust: A Causal Analysis." *American Journal of Political Science* 17, no. 4 (1973): 809-17.
Cook, Karen S., Toshio Yamagishi, Coye Cheshire, Robin Cooper, Masafumi Matsuda, and Rie Mashima. "Trust Building Via Risk Taking: A Cross-Societal Experiment." *Social psychology quarterly.* 68, no. 2 (2005): 121.
Cooper, John M. "Friendship and the Good in Aristotle." *The Philosophical Review* 86, no. 3 (1977): 290-315.
Cosmides, Leda, and John Tooby. "Are Humans Good Intuitive Statisticians Aher All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty." *Cognition.* 58, no. 1 (1996): 1.
Cullity, Garrett. "As You Were?" *Philosophical Explorations* 9, no. 1 (2006): 117-31.
Darwall, Stephen L. *The British Moralists and the Internal "Ought", 1640-1740*. Cambridge ; New York: Cambridge University Press, 1995.
Davion, Victoria. "Rape, Group Responsibility and Trust." *Hypatia* 10, no. 2 (1995): 153-56.
Dees, Richard H. "Trust and the Rationality of Toleration." *NoÃ»s* 32, no. 1 (1998): 82-98.
Demos, Raphael. "Lying to Oneself." *The Journal of Philosophy* 57, no. 18 (1960): 588-95.
Dick, Philip K. *Do Androids Dream of Electric Sheep?* New York: Ballantine Books, 1996.
———. *Flow My Tears, the Policeman Said*. 1st Vintage Books ed. New York: Vintage Books, 1993.
———. *The Man in the High Castle*. 1st Vintage Books ed. New York: Vintage Books, 1992.
———. *A Scanner Darkly*. 1st Vintage Books ed. New York: Vintage Books, 1991.
Dimock, Susan. "Retributivism and Trust." *Law and Philosophy* 16, no. 1 (1997): 37-61.

Doney, P. M., and J. R. Cannon. "An Examination of the Nature of Trust in Buyer-Seller Relationships." *Journal of Marketing* 61, no. 2 (1997): 35-51.
Dostal, Robert J. "The World Never Lost: The Hermeneutics of Trust." *Philosophy and Phenomenological Research* 47, no. 3 (1987): 413-34.
Ewin, R. E. "Loyalty and Virtues." *The Philosophical Quarterly* 42, no. 169 (1992): 403-19.
Farnham, Barbara. "Reagan and the Gorbachev Revolution: Perceiving the End of Threat." *Political Science Quarterly* 116, no. 2 (2001): 225-52.
Faulkner, Paul. "What Is Wrong with Lying?" *Philosophy and phenomenological research.* 75, no. 3 (2007): 535.
Fine, Gary Alan, and Lori Holyfield. "Secrecy, Trust, and Dangerous Leisure: Generating Group Cohesion in Voluntary Organizations." *Social psychology quarterly.* 59, no. 1 (1996): 22.
Foucault, Michel. *The Care of the Self*, The History of Sexuality, Vol. 3. New York: Vintage Books, 1988.
———. *Madness and Civilization : A History of Insanity in the Age of Reason*. New York: Vintage Books, 1988.
Foucault, Michel, and Paul Rabinow. *The Foucault Reader*. 1st ed. New York: Pantheon Books, 1984.
Freeland, Cynthia A. "Aristotelian Actions." *NoÃ»s* 19, no. 3 (1985): 397-414.
French, Peter A. *Cowboy Metaphysics : Ethics and Death in Westerns*. Lanham, Md: Rowman & Littlefield, 1997.
———. *Responsibility Matters*. Lawrence, Kan.: University Press of Kansas, 1992.
Friend, C. M. "Trust and the Presumption of Transluency." *SOCIAL THEORY AND PRACTICE* 27 (2001): 1-18.
Fukuyama, F. "Social Capital, Civil Society and Development." *Third World Quarterly* 22 (2001): 7-20.
Gambetta, Diego. "Can We Trust Trust?" In *Trust : Making and Breaking Cooperative Relations*, edited by Diego Gambetta, 213-37. New York, NY, USA: B. Blackwell, 1988.
———, ed. *Trust : Making and Breaking Cooperative Relations*. New York, NY, USA: B. Blackwell, 1988.
Gambetta, Diego, and Heather Hamill. *Streetwise : How Taxi Drivers Establish Their Customers' Trustworthiness*, The Russell Sage Foundation Series on Trust. New York: Russell Sage, 2005.
Gauthier, David P. *Morals by Agreement*. Oxford: Oxford University Press, 1986.
Gert, Bernard, and Timothy J. Duggan. "Free Will as the Ability to Will." *NoÃ»s* 13, no. 2 (1979): 197-217.
Gibson, Kevin. "Games Students Play: Incorporating the Prisoner's Dilemma in Teaching Business Ethics." *Journal of Business Ethics* 48, no. 1 (2003): 53-64.
Giffin, Kim. "Interpersonal Trust in the Helping Professions." *The American Journal of Nursing* 69, no. 7 (1969): 1491-92.
Gigerenzer, Gerd. "Content-Blind Norms, No Norms, or Good Norms? A Reply to Vranas." *Cognition* 81, no. 1 (2001): 93-103.

———. "From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology." *Psychological Review* 98, no. 2 (1991): 254-67.
———. *Gut Feelings : The Intelligence of the Unconscious*. New York: Viking, 2007.
———. "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)." *Psychological review.* 103, no. 3 (1996): 592.
Gigerenzer, Gerd, and D. G. Goldstein. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological review* 103, no. 4 (1996): 650-69.
Gilbert, Margaret. "Is an Agreement an Exchange of Promises?" *The Journal of Philosophy* 90, no. 12 (1993): 627-49.
Gleick, James. *Chaos : Making a New Science*. New York, N.Y., U.S.A.: Viking, 1987.
Goldstein, D. G., and G. Gigerenzer. "Models of Ecological Rationality: The Recognition Heuristic." *Psychological review* 109, no. 1 (2002): 75-90.
Goodin, Robert E. *Protecting the Vulnerable : A Reanalysis of Our Social Responsibilities*. Chicago: University of Chicago Press, 1985.
Govier, Trudy. "Self-Trust, Autonomy, and Self-Esteem." *Hypatia* 8, no. 1 (1993): 99-120.
———. "Trust, Distrust, and Feminist Theory." *Hypatia* 7, no. 1 (1992): 16-33.
Guignon, Charles B. *On Being Authentic*. 1st ed, Thinking in Action. London ; New York: Routledge, 2004.
Gul, Faruk. "A Nobel Prize for Game Theorists: The Contributions of Harsanyi, Nash and Selten." *The Journal of Economic Perspectives* 11, no. 3 (1997): 159.
Guseva, Alya, and Akos Rona-Tas. "Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared." *American Sociological Review* 66 (2001): 623-46.
Hardin, Russell. *Distrust*, Russell Sage Foundation Series on Trust ;. New York: Russell Sage Foundation, 2004.
———. *Trust*. Cambridge, U.K.: Polity, 2006.
Harsanyi, John C. "Bayesian Decision Theory and Utilitarian Ethics." *The American Economic Review* 68, no. 2 (1978): 223-28.
———. "Games with Incomplete Information." *The American Economic Review* 85, no. 3 (1995): 291-303.
Hayenhjelm, Madeleine. "Trusting and Taking Risks : A Philosophical Inquiry." Royal Institute of Technology, 2007.
Hieronymi, Pamela. "Controlling Attitudes." *Pacific Philosophical Quarterly 87, no* 1 (2006): 45-74.
———. "The Reasons of Trust." *Australasian Journal of Philosophy* 86, no. 2 (2008): 213-36.
Hill, Claire A., and Erin A. O'Hara. *A Cognitive Theory of Trust*: SSRN, 2005.
Holton, Richard. "Deciding to Trust, Coming to Believe." *Australasian Journal of Philosophy* 72, no. 1 (1994): 63 - 76.
Honderich, Ted, ed. *Morality and Objectivity : A Tribute to J.L. Mackie*, International Library of Philosophy. London ; Boston: Routledge & Kegan Paul, 1985.
Horsburgh, H. J. N. "The Ethics of Trust." *The Philosophical Quarterly* 10, no. 41 (1960): 343-54.
———. "Trust and Collective Security." *Ethics* 72, no. 4 (1962): 252-65.

Hosmer, Larue Tone. "Trust: The Connecting Link between Organizational Theory and Philosophical Ethics." *The Academy of Management Review* 20, no. 2 (1995): 379-403.
Hume, David, and Tom L. Beauchamp. *An Enquiry Concerning the Principles of Morals*, Oxford Philosophical Texts. Oxford ; New York: Oxford University Press, 1998.
Hume, David, L. A. Selby-Bigge, and P. H. Nidditch. *A Treatise of Human Nature*. 2d ed. Oxford: Oxford University Press, 1978.
Hummels, Harry, and Hans E. Roosendaal. "Trust in Scientific Publishing." *Journal of Business Ethics* 34, no. 2 (2001): 87-100.
Jeske, Diane. "Families, Friends, and Special Obligations." *Canadian journal of philosophy* 28, no. 4 (1998): 527 (30 pages).
———. "Friendship, Virtue, and Impartiality." *Philosophy and Phenomenological Research* 57, no. 1 (1997): 51-72.
Johnson, Devon, and Kent Grayson. "Cognitive and Affective Trust in Service Relationships." *Journal of Business Research* 58, no. 4 (2005): 500-07.
Jones, Karen. "Trust as an Affective Attitude." *Ethics* 107, no. 1 (1996): 4-25.
Kahn, Charles H. "Aristotle and Altruism." *Mind* 90, no. 357 (1981): 20-40.
Kapur, Neera Badhwar. "Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship." *Ethics* 101, no. 3 (1991): 483-504.
Kaye, Sharon. "True Friendship and the Logic of Lying." *The Journal of Value Inquiry* 39, no. 3 (2005): 475-85.
Kee, Herbert W., and Robert E. Knox. "Conceptual and Methodological Considerations in the Study of Trust and Suspicion." *The Journal of Conflict Resolution* 14, no. 3 (1970): 357-66.
Kirschenmann, Peter. "Reciprocity in the Uncertainty Relations." *Philosophy of Science* 40, no. 1 (1973): 52-58.
Kitcher, Philip. "The Evolution of Human Altruism." *Journal of Philosophy* 90, no. 10 (1993): 497-516.
Koehn, Daryl. "The Nature of and Conditions for Online Trust." *Journal of Business Ethics* 43, no. 1/2 (2003): 3-19.
Korman, D. "The Failure of Trust-Based Retributivism." *LAW AND PHILOSOPHY* 22, no. 6 (2003): 561-75.
Korsgaard, Christine M., and Onora O'Neill. *The Sources of Normativity*. Cambridge ; New York: Cambridge University Press, 1996.
Kramer, Roderick Moreland, and Karen S. Cook. *Trust and Distrust in Organizations : Dilemmas and Approaches*, The Russell Sage Foundation Series on Trust ;. New York: Russell Sage Foundation, 2004.
Kuhn, Steven T. "Agreement Keeping and Indirect Moral Theory." *The Journal of Philosophy* 93, no. 3 (1996): 105-28.
Lamprecht, Sterling P. "The Fable of the Bees." *The Journal of Philosophy* 23, no. 21 (1926): 561-79.
Laundau, Les. "Counterpoint." In *Star Trek: Voyager*, 45 minutes. United States: Paramount Television, 1998.
Leeds, Roger. "Breach of Trust." *Harvard International Review* 25, no. 3 (2003): 76-82.

Lewicki, Roy J., Daniel J. McAllister, and Robert J. Bies. "Trust and Distrust: New Relationships and Realities." *The Academy of Management Review* 23, no. 3 (1998): 438-58.
Lewis, J. David, and Andrew Weigert. "Trust as a Social Reality." *Social Forces* 63, no. 4 (1985): 967-85.
Lewis, J. David, and Andrew J. Weigert. "Social Atomism, Holism, and Trust." *The Sociological Quarterly* 26, no. 4 (1985): 455-71.
Lewis, Meirlys. "On Forgiveness." *The Philosophical Quarterly* 30, no. 120 (1980): 236-45.
Liddell, Brendan E. A. "Mr. Harsanyi on Hypothetical Imperatives." *Mind* 68, no. 272 (1959): 527-29.
Lingis, Alphonso. *Trust*, Theory out of Bounds. Minneapolis: University of Minnesota Press, 2004.
Locke, John, and Peter Laslett. *Two Treatises of Government*. Student ed, Cambridge Texts in the History of Political Thought. Cambridge [England] ; New York: Cambridge University Press, 1988.
Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch. "Risk as Feelings." *Psychological bulletin* 127, no. 2 (2001): 267-86.
Longino, Helen E. "What's Really Wrong with Quantitative Risk Assessment?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1986 (1986): 376-83.
MacIntyre, Alasdair C. *After Virtue : A Study in Moral Theory*. 2nd ed. Notre Dame, Ind.: University of Notre Dame Press, 1984.
———. *Selected Essays*. Cambridge, UK ; New York: Cambridge University Press, 2006.
———. *Three Rival Versions of Moral Enquiry : Encyclopaedia, Genealogy, and Tradition : Being Gifford Lectures Delivered in the University of Edinburgh in 1988*. Notre Dame, Ind.: University of Notre Dame Press, 1990.
Malhotra, Deepak, and J. Keith Murnighan. "The Effects of Contracts on Interpersonal Trust." *Administrative Science Quarterly* 47, no. 3 (2002): 534-59.
Malle, Bertram, Joshua Affiliation University of Oregon Knobe, and University Stanford. "The Folk Concept of Intentionality." *Journal of Experimental Social Psychology 33, no* 2 (1997): 101 (21 pages).
Mara, Gerald M. "Thucydides and Plato on Democracy and Trust." *The Journal of Politics* 63, no. 3 (2001): 820-45.
Marx, Karl, and Lawrence Hugh Simon. *Selected Writings*. Indianapolis: Hackett, 1994.
McAllister, Daniel J. "Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations." *The Academy of Management Journal* 38, no. 1 (1995): 24-59.
McClennen, Edward F. "Utility and Equity: Sen Vs. Harsanyi." *The Journal of Philosophy* 78, no. 10 (1981): 600-01.
McGeer, Victoria. "Trust, Hope and Empowerment." *Australasian Journal of Philosophy* 86, no. 2 (2008): 237 - 54.

Messick, David M., and Roderick Moreland Kramer. "Trust as a Form of Shallow Morality." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xxviii, 403 p. New York: Russell Sage Foundation, 2001`.
Mill, John Stuart, and Roger Crisp. *Utilitarianism*, Oxford Philosophical Texts. Oxford ; New York: Oxford University Press, 1998.
Mill, John Stuart, and Elizabeth Rapaport. *On Liberty*. Indianapolis: Hackett Pub. Co., 1978.
Miller, Cecil H. "Kant's Good Will and the Scholar." *Ethics* 80, no. 1 (1969): 62-65.
Mintoff, Joe. "Rational Cooperation, Intention, and Reconsideration." *Ethics* 107, no. 4 (1997): 612-43.
Möllering, Guido. "The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension." *Sociology* 35, no. 2 (2001): 403-20.
Molm, Linda D., Nobuyuki Takahashi, and Gretchen Peterson. "Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition." *American Journal of Sociology* 105 (2000): 1396-427.
Nagel, Thomas. "Hobbes's Concept of Obligation." *The Philosophical Review* 68, no. 1 (1959): 68-83.
Nee, Victor, and Jimy Sanders. "Trust in Ethnic Ties: Social Capital and Immigrants." In *Trust in Society*, edited by Karen S. Cook, Russell Hardin and Margaret Levi, xxviii, 403 p. New York: Russell Sage Foundation, 2001.
Nesse, Randolph M., ed. *Evolution and the Capacity for Commitment*, Russell Sage Foundation Series on Trust ;. New York: Russell Sage Foundation, 2001.
Newton, Kenneth. "Trust, Social Capital, Civil Society, and Democracy." *International Political Science Review / Revue internationale de science politique* 22, no. 2 (2001): 201-14.
Nickel, Philip. "Trust and Obligation-Ascription." *Ethical Theory and Moral Practice* 10, no. 3 (2007): 309-19.
Nietzsche, Friedrich Wilhelm, and Walter Arnold Kaufmann. *Beyond Good and Evil : Prelude to a Philosophy of the Future*. New York: Vintage Books, 1989.
Nilson, Douglas C., and Linda Burzotta Nilson. "Trust in Elites and Protest Orientation: An Integrative Approach." *Political Behavior* 2, no. 4 (1980): 385-404.
Oelsner, Andrea. "Friendship, Mutual Trust and the Evolution of Regional Peace in the International System." *Critical Review of International Social and Political Philosophy* 10, no. 2 (2007): 257 - 79.
Parry, Geraint. "Trust, Distrust and Consensus." *British Journal of Political Science* 6, no. 2 (1976): 129-42.
Perrin, Andrew J., and Sondra J. Smolek. "Who Trusts? Race, Gender, and the September 11 Rally Effect among Young Adults." *Social Science Research* 38, no. 1 (2009): 134-45.
Pettit, Philip. "The Cunning of Trust." *Philosophy and Public Affairs* 24, no. 3 (1995): 202-25.
Pufendorf, Samuel, and Craig L. Carr. *The Political Writings of Samuel Pufendorf*. New York: Oxford University Press, 1994.
Regan, Donald. "On Preferences and Promises: A Response to Harsanyi." *Ethics* 96, no. 1 (1985): 56-67.

Rengger, Nicholas. "The Ethics of Trust in World Politics." *International Affairs (Royal Institute of International Affairs 1944-)* 73, no. 3 (1997): 469-87.
Rolin, Kristina. "Gender and Trust in Science." *Hypatia* 17, no. 4 (2002): 95-118.
Rosthal, Robert. "Moral Weakness and Remorse." *Mind* 76, no. 304 (1967): 576-79.
Rousseau, Jean-Jacques, and Donald A. Cress. *Basic Political Writings*. Indianapolis: Hackett Pub. Co., 1987.
Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, Mass.: Belknap Press of Harvard University Press, 1998.
Schneewind, J. B. *The Invention of Autonomy : A History of Modern Moral Philosophy*. Cambridge ; New York, NY, USA: Cambridge University Press, 1998.
Schwarzenbach, Sibyl A. "On Civic Friendship." *Ethics* 107, no. 1 (1996): 97-128.
Seligman, A. *The Problem of Trust*. Princeton, N.J.: Princeton University Press, 1997.
Seligman, A. B. "Trust and Sociability: On the Limits of Confidence and Role Expectations." *American Journal of Economics and Sociology* 57, no. 4 (1998): 391-404.
Sellman, Derek. "Trusting Patients, Trusting Nurses." *Nursing Philosophy* 8, no. 1 (2007): 28-36.
Sheppard, Blair H., and Dana M. Sherman. "The Grammars of Trust: A Model and General Implications." *The Academy of Management Review* 23, no. 3 (1998): 422-37.
Sherman, Nancy. "Aristotle on Friendship and the Shared Life." *Philosophy and Phenomenological Research* 47, no. 4 (1987): 589-613.
———. "The Virtues of Common Pursuit." *Philosophy and Phenomenological Research* 53, no. 2 (1993): 277-99.
Shrader-Frechette, Kristin. "Risk Assessment and Uncertainty." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1988 (1988): 504-17.
Smith, Thomas W. "Aristotle on the Conditions for and Limits of the Common Good." *The American Political Science Review* 93, no. 3 (1999): 625-36.
Stocker, Michael. "Values and Purposes: The Limits of Teleology and the Ends of Friendship." *The Journal of Philosophy* 78, no. 12 (1981): 747-65.
Stroud, Barry. *Hume*, The Arguments of the Philosophers. London ; New York: Routledge, 1999.
Swinth, Robert L. "The Establishment of the Trust Relationship." *The Journal of Conflict Resolution* 11, no. 3 (1967): 335-44.
Thomas, D. O. "The Duty to Trust." *Proceedings of the Aristotelian Society* 79 (1978): 89-101.
Tong, Rosemarie. *Feminist Thought : A Comprehensive Introduction*. Boulder, Colo.: Westview Press, 1989.
Ullmann-Margalit, Edna. "Trust out of Distrust." *The Journal of Philosophy* 99, no. 10 (2002): 532-48.
Uslaner, Eric M. *The Moral Foundations of Trust*. New York: Cambridge University Press, 2002.
Wallace, Donnel, and Paul Rothaus. "Communication, Group Loyalty, and Trust in the Pd Game." *The Journal of Conflict Resolution* 13, no. 3 (1969): 370-80.

Wellman, Christopher Heath. "Friends, Compatriots, and Special Political Obligations." *Political Theory* 29, no. 2 (2001): 217-36.
White, Nicholas. "Conflicting Parts of Happiness in Aristotle's Ethics." *Ethics* 105, no. 2 (1995): 258-83.
Wicks, Andrew C., Shawn L. Berman, and Thomas M. Jones. "The Structure of Optimal Trust: Moral and Strategic Implications." *The Academy of Management Review* 24, no. 1 (1999): 99-116.
Will Carrington, Heath. "Von Neumann/Morgenstern Decision Making and Harsanyi's Theory of Justice: After Rawls and Nozick, If It Risks Personal Freedom and Individual Liberty, Can It Be Really Just?" *American Journal of Economics and Sociology* 47, no. 3 (1988): 355-62.
Wolff, Robert Paul. "Reflections on Game Theory and the Nature of Value." *Ethics* 72, no. 3 (1962): 171-79.
Wollstonecraft, Mary. *A Vindication of the Rights of Woman*. Rev. ed, Penguin Classics. London ; New York: Penguin Books, 2004.
Yamagishi, Toshio, Satoshi Kanazawa, Rie Mashima, and Shigeru Terai. "Separating Trust from Cooperation in a Dynamic Relationship: Prisoner's Dilemma with Variable Dependence." *Rationality and Society* 17, no. 3 (2005): 275-308.
Yarros, Victor S. "Theoretical and Practical Nietzscheism." *The American Journal of Sociology* 6, no. 5 (1901): 682-94.
———. "The Trust Problem Restudied." *The American Journal of Sociology* 8, no. 1 (1902): 58-74.
Young, Iris Marion. "Recent Theories of Justice." *Social Theory & Practice* 18, no. 1 (1992): 63-79.

About the Author

C.S.Hudspeth, son of a mother and a father and brother to a sister, while having left his home state for another in order to pursue his undergraduate degree at a fine institution and having returned to the state of his youth for graduate school in a more northern city than the very southern, even tropical, city in which he grew up, found, one day, that he had, really and truly, finally, finished writing a document that he had begun to take for granted as existing daily in his life, which surprised and perplexed him but left him, overall, with a sense of accomplishment and well-being that he suspected would remain with him for most, if not all, of the remainder of his days, a prospect that, the thought of which, brought a smile to his face, warmth to his heart, and tears to his eyes.