THE EPISTEMOLOGICAL IMPORTANCE OF TRUST IN SCIENCE

by

Karen Louise Frost-Arnold

B.A., Wellesley College, 1999

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Karen Frost-Arnold

It was defended on

July 29th, 2008

and approved by

Lisa Parker, Associate Professor, Department of Human Genetics

Nicholas Rescher, University Professor, Department of Philosophy

Laura Ruetsche, Associate Professor, Department of Philosophy

Dissertation Co-Advisor: Sandra Mitchell, Professor, Department of History and Philosophy of Science

Dissertation Co-Advisor: Kieran Setiya, Associate Professor, Department of Philosophy

THE EPISTEMOLOGICAL IMPORTANCE OF TRUST IN SCIENCE

Karen Frost-Arnold Ph.D.

University of Pittsburgh, 2008

I argue that trust is epistemically important because it is the foundation of social practices that confer significant epistemic benefits on scientific communities.

I begin by showing the limitations of the dominant rational choice account of trust, which maintains that trust is rational when the truster has good reason to believe that it is in the trusted's self-interest to act trustworthily. These limitations motivate my alternative account of trust, which recognizes non-self-interested motivations for acting trustworthily, such as having a sense of duty. The first part of the account captures the cognitive aspect of trust. When we trust, we take a particular cognitive attitude towards the claim that the trusted will do what we expect her to do; I argue that this cognitive attitude can be either belief or acceptance, in the sense outlined by Michael Bratman. The second part of my account captures the emotional and moral aspects of trust by providing a framework to understand the connection between trust and betrayal—the feeling that usually results from being let down by a person one trusts. I provide an account of betrayal as a reactive emotion that connects it to beliefs about relational obligations. Thus when we trust, we depend on the trusted because we believe that our relationship with the trusted morally obliges her to act as expected.

Using this account of trust, I argue that scientific communities can garner significant epistemic benefits when scientists are trustworthy and when they trust each other. Applying a framework adapted from Alvin Goldman's work on social epistemology, I argue that trust fosters epistemically beneficial sharing between scientists. These arguments are supported by a case study of the role that trust played in the achievements made by the community of early 20th Century Drosophilists. Finally, using recent examples of scientific fraud in cloning research and public policy responses to much-publicized ‘crises in trust’, I argue that the epistemic success of science results, in part, from science’s ability to balance competition and cooperation, trust and distrust, self-interest and other-interest.

TABLE OF CONTENTS

PREFACE

I would like to acknowledge the dedicated work of my co-advisors, Sandra Mitchell and Kieran Setiya, and my whole dissertation committee, including Lisa Parker, Nicholas Rescher and Laura Ruetsche. It has meant a lot to me to have such a trustworthy group of mentors.

I first became fascinated in the subject of trust when I read Annette Baier's "Trust and Anti-Trust" during Lisa Parker's seminar on Feminist Bioethics. As well as sparking my interest in trust, Lisa Parker has been an extremely careful reader of my work, and I am grateful for the many difficult questions she has asked along the way.

Joe Camp and Sandra Mitchell were kind enough to do a joint independent study with me to help me develop my interest in trust into a dissertation project. My regular meetings with the two of them in his sweltering office with the broken radiator were challenging and exciting. I would like to thank Joe Camp for supporting me early in my graduate student career and for teaching me to think beyond the obvious and ordinary answers.

Sandra Mitchell has mentored me since my first semester of graduate school. I am grateful for her patience and unflagging interest in whatever work I was doing. I cannot measure how much I have learned from being in her seminars and reading groups and from our many conversations.

Kieran Setiya, who stepped in as my dissertation co-advisor upon Joe Camp's retirement, saved me from confusion countless times. I was lost in the trust literature until he pointed me towards certain key pieces and helped me clarify my thoughts. I am thankful for his dedication and crystal clear thinking.

Nicholas Rescher has been an inspiration for this project both through his own work and our conversations that I enjoyed so much. He has consistently helped me to see the forest instead of only the trees, and I am grateful for his many insights into the nature of trust.

When I have gotten myself in over my head, Laura Ruetsche has repeatedly helped me figure my way out. This dissertation might have remained stuck on a particular problem in social epistemology if she had not reminded me of an article she taught in one of my first year seminars.

I am also grateful to my wonderful family and friends for all their support. Charliene Hooker, Bill Hamilton and David Krueger-Duncan were kind enough to provide painstaking proofreading. Emily Nacol has been a supportive friend throughout college and graduate school. And most importantly, without the love and encouragement of my husband, Greg Frost-Arnold, and my parents, Richard and Gillian Arnold, none of this would have been possible.

0.0 INTRODUCTION

But it is impossible to go through life without trust:
that is to be imprisoned in the worst cell of all, oneself.
—Graham Greene *The Ministry of Fear*

As social beings, we are dependent on others. Most of the goods in life (e.g. food, health, security, intimacy, and education) cannot be brought about by one individual acting alone. In addition, once one has attained some of these goods, one cannot sustain or keep them secure without the help or, at the very least, non-interference of others (Baier 1994, p.95). This means that we are constantly making plans based on the assumption that others will play their part in providing, sustaining, or leaving alone something on which we depend. In particular, we constantly trust others to perform or refrain from performing particular actions.

Scientists are no different from the rest of humanity in this regard. The most forceful arguments for a role for trust in scientific knowledge have come from philosophers who focus on the importance of collaboration for science (Hardwig 1991 and Webb 1993). They note that modern science is increasingly a social institution in which individuals work together to produce scientific knowledge.[1] One indicator of the degree of dependence in science is the number of

[1] Of course, science has always been a social practice if only in the fact that individuals respond to and attempt to build on the results of predecessors.

scientists who work together on a project. In many fields it is common for published papers to include several authors. Laboratories are often complex groups of scientists working together on a project at a variety of different levels. Senior scientists frequently depend on graduate students and junior scientists to produce much of the data. Scientists also collaborate with colleagues from different labs at different institutions around the world. Another piece of evidence for dependence in science is the extensive role of testimony in scientific knowledge. Much of what scientists claim to know is based on testimony from others. Testimony obviously plays a central role in the education of scientists. It is also present in the ways scientists learn about new developments in their fields, through reading journals and attending conferences. Therefore, scientists are dependent on each other, and so need to trust each other. Thus, at the broadest level of analysis, trust is a response to dependence, and this connection between trust and dependence makes trust a particularly useful concept for investigating the social character of science.

In the past few decades, philosophers of science have become increasingly interested in understanding the social character of science. Even though a significant amount of work has been done in the sociology of science and in social epistemology, relatively little has been said about the relationship between the morality of science and its epistemology. The present project attempts to fill this gap by asking what ethics can contribute to the epistemology of science. The ethical concept I focus on is *trust*. The guiding questions of my project are: Is trust epistemically important? Does trust between scientists help the community meet the epistemic aims of science?[2] The inclusion of the phrase ‘the community’ indicates that this is an exercise in social epistemology. As such, this project takes social practices as the focal unit of epistemic analysis and investigates the cognitive attitudes of individuals insofar as they motivate individuals to

[2] The nature of these epistemic aims will be specified in chapter three.

engage in epistemically significant social practices. I argue that the cognitive attitude of *trust is epistemically important because it is the foundation of social practices that confer significant epistemic benefits on scientific communities*.

This analysis of the role of trust in science brings us closer to a more complete social epistemology of science. Interestingly, many current social epistemologies share fundamental assumptions with the dominant accounts of trust. Both approaches ground their analyses in a particular view of human nature: that the agents under study are self-interested rational beings. While there are good theoretical reasons for making this assumption, it nonetheless blinds us to a certain range of the phenomena under study. In accounts of trust, assuming that humans are self-interested blinds us to the ways in which we depend upon others to act out of a sense of moral obligation.[3] In social epistemology, the assumption of self-interest prevents us from investigating the epistemic significance of the non-self-interested motivations that scientists do in fact have. In particular, it ignores the role that trust in dutiful scientists, who are motivated by their interest in others, has in producing better science. To show why a social epistemology of trust is needed to complete the picture, a discussion of the aims and current state of social epistemology is required.

0.1 THE SOCIO-HISTORICAL CHALLENGE AND SOCIAL EPISTEMOLOGY

The past few decades of socio-historical study of science have challenged the traditional view of the field. The notion that science owes its privileged status as the paradigm of objective,

[3] The limits of this approach to trust are presented in detail in chapter one.

rational inquiry to scientists' disinterested dedication to the pursuit of truth has been perceived to be undermined by sociological and historical research showing that scientists are influenced by non-epistemic values. Sociologists, philosophers and historians have argued that scientific practice is not driven primarily by desire for the truth and responsiveness to the evidence, but is instead directed by such factors as: (a) individual desire for reputation, (b) competition between laboratories for grant funding, (c) personal loyalties and relationships, (d) political, social and moral beliefs, (e) gender, race and economic status, (f) power relations, (g) and cultural paradigms—among others. In response, some sociologists, historians and philosophers have claimed that the traditional view is a myth, on the grounds that these factors were traditionally thought of as biasing and science ought to be viewed as no less susceptible to these biasing factors than other areas of human activity.[4] This response has been perceived to undermine the objectivity and rationality of science. However, others have argued that since science is a paradigm of objective, rational inquiry, it is the traditional conceptions of objectivity and rationality that must be questioned.[5] Philosophers of science of this latter persuasion have taken up the challenge of providing an explanation of the objectivity and rationality of science that takes the sociological and historical claims into account.

[4] The Edinburgh School's Strong Programme is the most influential version of this response to the socio-historical challenge (e.g. Barnes and Bloor 1982). In addition, Feyerabend argues against the privileged status of science (Feyerabend 1975).

[5] For detailed discussions of this move, see (Kitcher 1993 and 2001), as well as (Longino 1990 and 2002). The social epistemologists described below provide one example of this response to the socio-historical challenge. Standpoint epistemologists (e.g. Haraway (1991), Harding (1991)) also respond to the challenge by proposing a new vision of objectivity. Kuhn's essay "Objectivity, Value Judgment, and Theory Choice" (1977) is another frequently cited attempt to respond to the socio-historical challenge by reformulating our notion of objectivity.

One response to this socio-historical challenge has been to move towards a social epistemology of science according to which the objectivity[6] of science appears at the community level. Such social epistemologists attempt to specify the conditions under which groups of individuals interact in ways that are communally objective.[7] This is not to say that individual scientists are always completely biased, but rather that if we are to demonstrate the objectivity of science, we need to show that the structures and institutions that govern the interactions between scientists produce objectivity. This type of social epistemology rejects the notion that science's objectivity results from its being composed of individual scientists who each eschew their subjective viewpoint and are unmoved by biasing factors like self-interest, personal relationships, socioeconomic status, race and gender, etc. Thus, these social epistemologists accommodate the observations of sociologists by providing an alternative to the traditional view of individual scientists as disinterested truth-seekers. They acknowledge both that scientists are motivated by interests other than an impersonal desire for true belief and that their scientific decisions are made on the basis of more than just the weight of the evidence. Thus, social epistemology requires that we move away from a conception of objectivity as, "some special relation (of detachment, hardheadedness) they [individual scientists] may bear to their observations" (Longino 1990, p.79).

Social epistemologists argue that interested, idiosyncratic parties can still create objective science, as long as they are organized in what Kitcher calls 'epistemically well-designed social systems' (Kitcher 1993, p.303). In epistemically well-designed social systems, biased,

[6] Henceforth, I will frame the discussion in terms of objectivity instead of rationality. However, there is clearly a close connection between objectivity and one type of rationality since subjective, biasing factors are often considered to be irrational influences.

[7] Philosophers who make this move to social epistemology include: Goldman (1992), Hull (1988), Kitcher (1993), Longino (1990 and 2002), and Railton (1994).

subjective assumptions are washed out, so that the products of the systems do not reflect the limited, subjective concerns of self-interest, personal loyalties, social and political agendas, race, gender and socioeconomic status, etc. Instead of being a relation that individuals bear to the world, objectivity is the result of surviving the social processes that wash out these biasing factors. Science can be objective, in the sense of not being biased, tainted or inappropriately partial, if the scientific community is well-organized. For example, Longino argues that "criticism from alternative points of view is required for objectivity and that the subjection of hypotheses and evidential reasoning to critical scrutiny is what limits the intrusion of individual subjective preference into scientific knowledge" (Longino 1990, p.76). For Longino, an epistemically well-designed scientific community is one with effective methods for hypotheses to be subjected to extensive and varied critical scrutiny. On her account, science is objective to the extent that the scientific community is organized in ways that produce this critical scrutiny. Her account of objectivity illustrates the social epistemologists' strategy of acknowledging the existence of biasing factors identified by the sociologists and historians, while maintaining that science produced by well-designed communities can still be objective. Social epistemologies differ in their accounts of the processes and structures that best wash out biasing factors, but they share the conception of objectivity as a property of a well-designed community.

0.1.1 Social epistemologies of self-interest

One common form of this move to social epistemology investigates how self-interested scientists can interact in ways that are communally objective (e.g. Goldman 1992, Hull 1988, Kitcher 1993, and Railton 1994). This approach takes to heart the observations of the role that desire for reputation, credit and funding play in science (see Latour and Woolgar 1979). Peter

Railton argues that “objectivity arises not so much at the level of the individual investigator as at the social level, a perhaps unintended consequence of competition for funds, glory, and other scarce resources in a circumstance in which innovation that enhances prediction and control is rewarded” (Railton 1994, p.83). Thus objectivity is a result of an “invisible mind” that acts much like the “invisible hand” of economics (Railton 1994, p.84). A competitive community in which there are rewards for new, subsequently replicable results is one in which self-interested scientists are strongly motivated to stake out their own lines of research. Thus the community benefits from a greater range of hypotheses investigated and tested. The claims produced by such a community will be more objective than those in a community in which only a small range of hypotheses are tested.

Philip Kitcher also argues that “the very factors that are frequently thought of as interfering with the (epistemically well-designed) pursuit of science—the thirst for fame and fortune, for example—might actually play a constructive role in our community epistemic projects, enabling us, as a group, to do far better than we would have done had we behaved as independent epistemically pure individuals” (Kitcher 1993, p.351). Kitcher models many aspects of scientific practice formally. He uses this formal analysis to show that, under certain conditions, a community of *pure epistemic agents*, those whose primary goal is to attain an epistemically valuable state (e.g. true belief), will do worse than a community of epistemically *sullied agents*, those who are motivated both by a desire for priority and a desire to solve problems (Kitcher 1993, p.351). This is because pure agents will tend towards uniformity, whereas the community of sullied agents will be more diverse in their research strategies. Why? Since pure agents are all motivated by a desire to attain an epistemically desirable state, they will all pursue the line of research that appears most promising (perhaps by being better developed or more successful) (Kitcher 1994, p.344). In contrast, epistemically sullied agents have more reason to take risks because they want to be the first to solve a problem. Therefore, like Railton, Kitcher thinks that we need not be dismayed when sociologists tell us that scientists are driven by self-interest (in this case, desire for credit). If science is well-organized to provide the right incentives to such agents, it can still be a rational, objective enterprise.

0.1.2 A social epistemology of trust

These social epistemologies of self-interest are interesting and useful analyses of the social aspect of the epistemology of science, especially since, as a matter of fact, scientists can be motivated by considerations of self-interest. However, they are not the only possible form that social epistemology can take to respond to the socio-historical challenge to provide an account of the objectivity of science that takes seriously the power of biasing factors to influence scientists. Consider the following quotation from Railton:

> what forms of interaction with the world might selectively reward reliability in belief-formation, and thereby tend over time to enhance it? Strikingly, a disinterested, contemplative, appreciative mode of inquiry would not appear a promising candidate to fill this bill…. Equally strikingly, an interested, instrumental, manipulative, ambitious, competitive mode of inquiry might better fill the bill of producing novel forms of feedback that can frustrate subjective projection—at least, up to a point. (Railton 1994, pp.81-2)

In this passage, Railton moves from an appreciation of the limits of disinterested inquiry to the suggestion that we investigate whether self-interested inquiry might better provide objectivity. Suppose one agrees with his assessment of the failings of disinterested inquiry; is the only solution to substitute self-interested inquiry for disinterested inquiry? The answer is *no*, because *dis*-interested and *self*-interested are not exhaustive categories. One might suggest *other*-interested inquiry as an alternative. People often have other-interested motivations; in other words, they are often motivated to act for the good of others even when it is not in their self-interest to do so. Often we do something because we believe we have an obligation to do it, and that sense of obligation is motivating for us. In chapter two, I present an account of trust which shows that other-interested individuals are trustworthy individuals. Thus there is conceptual space for a social epistemology that examines how communities of other-interested scientists can be well-organized to act in rational ways that produce objective knowledge. This conceptual possibility, which provides the tools for analyzing the role of trust in science, has been largely ignored by philosophers of science.[8, 9]

What is the relationship between other-interest and trust? There are many types of motivations for acting for the good of others that are not self-interested reasons, including caring, a sense of honor and a sense of duty. It is this last type of other-interest, being motivated by a sense of duty or obligation, that is connected to trust. In chapter two I present an account of trust on which one adopts the cognitive attitude of trust because one believes that one's relationship with the trusted morally obliges her to do that which one trusts her to do. Trustworthy individuals are those who are motivated to live up to the obligations inherent in their relationships with others. I call these moral obligations 'relational obligations'. Thus, other-interested individuals who are motivated by a sense of duty to live up to their relational obligations are trustworthy. Therefore, a social epistemology of trust investigates the role that dependence on the other-interestedness of scientists plays in producing objective science.

[8] The notable exception is Steven Shapin's work, especially (1994). However, his work centers on the science of the 17th Century. There has, to my knowledge, been little exploration of this approach in relation to modern science. Thagard (1997) also analyzes the epistemology of collaboration, but he does not specifically address other-directed motivations for collaboration.

[9] Not only has philosophy of science ignored this possibility; but, as I argue in chapter four, by focusing on self-interested motivations, philosophy of science justifies institutional structures that may actually undermine other-interestedness in science.

Like other social epistemologies, a social epistemology of trust shows that what appear to be biasing factors can, when scientists interact in a well-ordered community, provide epistemic benefits to the community. Just as it is natural for individuals to be motivated, to some degree, by self-interest, it is natural for people to form personal ties and become involved in a variety of different social relationships (Rachels 1975, p. 326). Any particular scientist will have different personal relationships with different members of the scientific community. Some will be her friends, some her students, others her collaborators, and some will be members of her clique or 'social circle' of scientists who informally exchange information about her research area.[10] Initially, it appears that the personal relationships and personal loyalties of scientists might bias their work. A sense of duty to one's friends, colleagues or students could cause scientists to give special, unwarranted credence to the work of scientists with whom they have personal, trusting relationships. In addition, the existence of cliques and social circles might harm the objectivity of science. If scientists share, collaborate, and discuss their work only or primarily with trusted members of their social circle, the kind of transformative criticism that Longino argues increases objectivity may not take place. Assumptions may remain unquestioned as information is not shared evenly throughout the community.

In general, it appears that scientists' trust in the other-interestedness of their colleagues threatens objectivity by undermining two of the norms that Robert K. Merton identified in the 1940's as constituting the 'ethos of science' (Merton 1942/1996). Merton's norms are usually taken to provide the classic statement of the traditional view of science, which has since come under attack by sociologists and historians. Merton's first norm is *universalism*, which means that

[10] Diana Crane (1972) and Karin Knorr Cetina (1999) provide detailed analyses of social circles and informal communication in science.

> …truth claims, whatever their source, are to be subjected to *preestablished impersonal criteria*; consonant with observation and with previously confirmed knowledge. The acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonists; their race, nationality, religion, class, and personal qualities are as such irrelevant. Objectivity precludes particularism. (Merton 1942/1996, p.269)

As I have suggested, personal relationships can promote particularism, and trusting relationships may lead scientists to subject scientific claims to personal, rather than impersonal, criteria. Another one of Merton's norms is *communism*, by which he means that the findings of scientific work are to be shared communally rather than hoarded as the private property of the discoverer: "The institutional conception of science as part of the public domain is linked with the imperative for communication of findings. Secrecy is the antithesis of this norm; full and open communication its enactment" (Merton 1942/1996, p.272). Sharing information only with trusted members of one's clique seems to fly in the face of full and open communication. Therefore, trusting relationships might initially seem to bias scientists, and thereby undermine the objectivity of science. One of the aims of social epistemology is to show how science can be objective at the community level despite the existence of factors that bias scientists at the individual level. In sum, there is room for a social epistemology that investigates how to design an epistemically well-functioning community of scientists who are often motivated by concern for others and who often trust their colleagues to be similarly motivated. Such an epistemology would both fill a conceptual gap and move the discussion toward a more complete understanding of the actual impact of social relations on the epistemology of science.

A social epistemology of trust needs to draw on a particular account of trust. Unfortunately, as previously mentioned, the dominant accounts of trust are unsuitable for providing a social epistemology of trust in other-interested scientists. This is because the dominant accounts take what I call the 'rational choice approach' by analyzing trust as reasonable belief in the self-interested actions of the trusted. As I argue in chapter one, this approach has several limitations in its own right. In addition, such an account of trust will not furnish the theoretical tools needed for an alternative to the social epistemologies of self-interest. For this reason, in chapter two, I provide my own account of trust, which is better suited to the purposes of the epistemological project at hand. Trust is a broad concept. Therefore, some preliminary remarks about the kind of trust at issue and the types of questions that are relevant to this project are required.

0.2 TRUST: PRELIMINARY DISTINCTIONS

Trust, in the sense discussed here, is a three-part relation in which *A* trusts *B* to do *C*. This type of trust always involves a truster, *A*, a trusted, *B*, and an action, *C*, that the truster trusts the trusted to perform. This means that I will not be concerned with questions about how trusting scientists are in general (this would be to analyze trust as a one-part relation), nor will I be considering questions about what it means for *A* to have generalized trust, or faith, in *B* (trust as a two-part relation). Instead, it will be necessary to investigate the nature of the cognitive attitude that *A* takes when she trusts *B* to do *C*. I will be asking questions about what trust of this sort is. In answering this question it is also helpful to examine the nature of our reasons for trusting, since an account of the rationality of trust presupposes some notion of what kind of a phenomenon trust is. Similarly, an account of what trust is can naturally suggest a particular set of standards for evaluating trust. While there are many other important questions one could ask about the morality of trust (e.g. When is trust blameworthy?), answering these questions is not directly relevant to showing that trust has epistemic significance in science. Hopefully, demonstrating the importance of trust in science will encourage others to take up these worthwhile questions.

If trust is a three-part relation between the truster, the trusted and an action the trusted is expected to perform, it is important to determine what kinds of things one can put one's trust in. Even a cursory survey of common usage of 'trust' shows that a wide variety of things are referred to as being trusted to perform some action. The paradigmatic cases of trust are when we trust people to perform actions, e.g. trust in one's spouse to be faithful, trust in the babysitter to care for one's child, trust in one's colleague to turn in the report on time, and trust in one's doctor to prescribe the right medication. Thus, discussions of trust center typically around questions of the nature of, and reasons for, adopting a cognitive attitude of trust towards the proposition that someone will do something. While people are the paradigmatic objects of trust, animals, natural objects, artificial objects and machines, and institutions are all also commonly referred to as objects of trust, e.g. trust in one's cat not to eat the fish, trust in the wetlands to prevent flooding, trust in one's car to start in the morning and trust in the medical system[11]. In one sense, it is reasonable to say that we trust these objects as well as people because we can be dependent on all of these things. We make plans based on the assumption that our alarm clock will wake us up, and we make plans based on the assumption that our partner will not unplug the alarm clock. However, there is another sense in which our use of 'trust' to refer to our attitude

[11] Trust in institutions is an interesting mixed case. On the one hand, trust in the medical system involves 'trust in abstract objects' (to use Anthony Giddens' term) (Giddens 1990, p.114). We trust the procedures and social structures that govern the interaction of members of the medical profession and the public. On the other hand, trust in the medical system involves trusting those individual people who play specified roles in the system. While a comprehensive analysis of what it means to trust people who play institutionalized roles is outside the scope of this project, the account of trust developed in chapter two relies heavily on a notion of relational obligations that makes a step towards recognizing that our trust in people is tied up with our beliefs about the kinds of actions they are obligated to perform as a result of their norm-governed relationships with us.

towards non-persons is usually figurative. As I argue in chapter one, a central feature of trust is that trusting someone to do something makes the truster vulnerable to feeling *betrayed* if the trusted does not act as expected. Thus, one difference between our trust in people and our attitude towards non-persons is that we tend to feel merely disappointed rather than betrayed when non-persons fail to live up to our expectations.

The social epistemology of trust in science provides an analysis of the epistemic significance of scientists' trust in their colleagues to perform particular actions. The paradigmatic examples of trust in science include the following: the head of a laboratory trusts her postdoc to perform the experiments as directed and not fabricate results, the postdoc trusts the senior advisor not to steal her ideas, and one partner in an interdisciplinary collaboration trusts her partner to conduct her part of the project with due diligence. In chapter three, I argue that these kinds of trust, and more, create epistemic benefits to the scientific community as a whole.

Finally, in my analysis of the role of trust in science, I do not address any of the important and complex questions about trust and competence. In most situations, it makes no sense for *A* to trust *B* to do *C* if *A* has overwhelming evidence that *B* is incapable of doing *C*. For this reason, questions of competence are an important part of the investigation of the nature of trustworthiness and the rationality of trust. However, in this analysis, I follow Russell Hardin in focusing on the question of motivation for acting trustworthily (Hardin 2002, p.8). Thus, I will not address any of the questions about how scientists can determine whether their colleagues are competent to carry out the actions they are trusted to perform or how they determine whether their colleagues are sloppy or careless. The reason for sidestepping the issue of competence is that it is not directly relevant to the concerns of social epistemology. The relevant aim of social epistemology is to show how factors that appear to provide biased motivations to scientists can actually contribute to objectivity at the community level. Thus it is questions of motivation and not competence that are central to this analysis of trust in science. Having introduced the necessary preliminary distinctions to focus the concept of trust under discussion, I conclude this introduction with a brief outline of the structure of the argument to be presented.

0.3 THE STRUCTURE OF THE ARGUMENT

Trust and trustworthiness have received attention from numerous scholars representing a variety of fields. At present, it is possible to identify one dominant approach to the analysis of trusting relationships, which I call the *rational choice approach*. Chapter one outlines the rational choice approach and elucidates why it is of limited use in analyzing trusting relationships in general, and trusting relationships between scientists in particular. In short, rational choice theorists argue that trust is rational when the truster has good reason to believe that it is in the trusted's self-interest to act trustworthily. The rational choice approach provides useful analytical tools; however, it is limited because it fails to recognize the role of other motivations for acting trustworthily such as having a sense of duty. By ignoring these motivations, the rational choice approach provides only a partial explanation of why people trust and why they act trustworthily. Of particular importance is how these limitations prevent the rational choice approach from accounting for cases of trust on the part of powerless individuals.

This analysis of the lacunae in the dominant approach to trust and trustworthiness motivates an alternative account of trust offered in chapter two. My account is as follows:

> ***(T)*** A *trusts* B *to do* C *iff (1)* A *takes the proposition that* B *will do* C *as part of* A*'s adjusted cognitive background, (2) in part, because* A *believes that their relationship morally obliges* B *to do* C*.*

The first part of **(T)** captures the cognitive aspect of trust. When we trust, we take a particular cognitive attitude towards the claim that the trusted will do what we expect her to do. There is serious debate in the trust literature about the nature of this cognitive attitude—is it belief or some other cognitive attitude? Drawing on the work of Michael Bratman (1992), I argue that the cognitive attitude we take when we trust can be either belief or acceptance. The second part of **(T)** captures the emotional and moral aspects of trust by providing a framework to understand the connection between trust and betrayal, which is the feeling that usually results from being let down by a person one trusts. Following Annette Baier (1994), many authors identify vulnerability to betrayal as the essential difference between trust and mere reliance. Unfortunately, little has been said about what kind of an emotion betrayal is. I provide an account of betrayal as a reactive emotion that connects it to beliefs about relational obligations. Thus when we trust, we depend on the trusted because we believe that our relationship with the trusted obliges her to act as expected. This belief makes us vulnerable to feelings of betrayal if the trusted fails to fulfill these obligations.

In chapter three, I use this account of trust to analyze the role of trust in science. I argue that scientific communities can garner several significant epistemic benefits when scientists are trustworthy and when they trust each other to act trustworthily. Using a framework adapted from Alvin Goldman's work on social epistemology (1992), I show how the epistemic benefits of trusting relationships between scientists can be evaluated in terms of increased reliability, power, speed, efficiency and fertility. In particular, I argue that trust fosters epistemically beneficial sharing between scientists. Communities in which scientists share materials and technology, technical gossip and shop talk have increased reliability, power, speed, efficiency and fertility. These arguments are supported, in part, by a case study of the role that trust played in the achievements made by the community of Drosophilists that grew out of Thomas Hunt Morgan's laboratory.

Finally, in chapter four, I argue that the epistemic success of science results, in part, from science's ability to balance competition and cooperation, trust and distrust, self-interest and other-interest. Having followed the arguments of chapter three, one might object that while trust does have some epistemic benefits, those benefits are overshadowed by the harms that trust does to the epistemic projects of science. In particular, one might worry that a scientific community that values trust may increase conformism, bias and epistemic laziness. I respond to these objections by showing that not only are the epistemic benefits of trust not overshadowed by its epistemic harms, but that competition between self-interested scientists also has negative epistemic consequences and that these negative effects can be mitigated by trusting relationships. I use recent examples of scientific fraud in cloning research and public policy responses to much-publicized "crises in trust" to illustrate the need for balancing trust with competition while valuing diversity in science. I conclude with an outline of some positive proposals for how to design scientific communities that maximize the epistemic benefits of trust while minimizing its drawbacks.

1.0 LIMITATIONS OF THE RATIONAL CHOICE APPROACH TO TRUST

In recent decades, trust and trustworthiness have received attention from numerous scholars representing a variety of fields. Since the philosophers, psychologists, sociologists and political theorists who have studied trust and trustworthiness ask different questions and use diverse theoretical tools, it is difficult to classify the existing literature into a neat and tidy taxonomy. However, it is possible to identify one dominant approach, which I call the 'rational choice approach'. This chapter outlines the rational choice approach and elucidates why it is of limited use in analyzing trusting relationships in general, and trusting relationships between scientists in particular. This analysis of the lacunae in the dominant approach to trust and trustworthiness motivates an alternative account of trust, offered in chapter two.

There are a number of questions that a theory of trust could be intended to answer. First, 'What is trust?' One might want an account of what trust is that distinguishes it from other social phenomena. Second, 'Why do people trust each other (when they do)?' A theory of trust could attempt to elucidate the psychology of trust. Third, 'When is it rational to trust, and what makes it rational?' An evaluation of the legitimacy of the reasons people have for trusting could also be part of a theory of trust. Fourth, 'What are the standards for moral evaluation of trust?' One might want to know when trust is ethically suspect or praiseworthy.

The existing philosophical, political and sociological literature on trust contains attempts to answer all four of these questions. Often the questions are not clearly distinguished in an author’s attempt to tackle the subject of trust. This is understandable, since the four questions are connected. An account of what trust is can make it natural to suggest a particular set of standards for evaluating trust. Similarly, an account of the rationality of trust presupposes some notion of what kind of a phenomenon trust is. Answering one question about trust leads one to take positions on other central questions about it. In addition, as Russell Hardin has noted, much of the literature on trust is actually primarily about trustworthiness (Hardin 2002, p. 29). Hence, these four questions about trust can also be asked about trustworthiness. What is trustworthiness? Why do people act trustworthily? When is trustworthiness rational? How should we morally evaluate trustworthiness? For these reasons, the trust literature is somewhat complicated, and identifying established positions and persistent points of disagreement requires a significant degree of abstraction and reconstruction. In identifying the rational choice approach to trust, I am not referring to a position with which authors have already self-identified. Instead, I provide a clear explanation of the most prominent line of argument that I find in the trust literature. While the authors I will discuss are tackling different questions about trust and trustworthiness, I find among their arguments a distinctive set of assumptions and methods.

1.1 THE RATIONAL CHOICE APPROACH TO TRUST

Rational choice theorists share some philosophical motivations. They all have 1) an interest in the question ‘What makes trust rational?’, and 2) a preference for a minimalist approach, which leads them to provide answers to the question that center around self-interested rationality. In short, rational choice theorists argue that trust is rational when the truster has good reason to believe that it is in the trusted’s self-interest to act trustworthily. The writings of four proponents of the rational choice approach— Adler (1994), Blais (1987 and 1990), Hardin (2002), and Rescher (1989)—share these two features.

First, as attempts to answer the question, 'What makes trust rational?', rational choice accounts of trust are particularly concerned to show why it makes sense to trust, given the risks involved and the temptations of uncooperative behavior. When one trusts (or acts on trust)[12] one puts oneself, to some degree, in the hands of the trusted. Hardin puts it this way: "As virtually all writers on trust agree, acting on trust involves giving discretion to another to affect one's interests. This move is inherently subject to the risk that the other will abuse the power of discretion" (Hardin 2002, p.11). In his discussion of trusting testimony, Jonathan Adler defines trust in terms of risk: "It seems sufficient for a seeker to be extending trust that the seeker is at risk, if the testimony is false" (Adler 1994, p.266). It is recognition of this risk, coupled with attention to the temptations for abuse of power by the trusted, that makes the question of the rationality of trust pressing for rational choice theorists. Nicholas Rescher talks of the temptations for abuse of power in connection with the need for sharing of information in science: "Since information is power, there is a constant temptation to monopolize it. But information monopolies, however advantageous for some few favorably circumstanced beneficiaries, exact an awful price from the community as a whole" (Rescher 1989, p.34). The temptation to abuse the power entrusted to oneself is familiar from a host of everyday trusting situations. A

[12] I add this qualification because Russell Hardin is one rational choice theorist who insists that trusting and acting on trust are distinct (Hardin 2002, pp.58-60). Hardin is careful to note that on his account, trust is an involuntary cognitive attitude. For him, acting on trust involves risk-taking, but "trust is not itself a matter of deliberately taking a risk because it is not a matter of making a choice" (Hardin 2002, p.12). Other rational choice theorists are not so careful in observing the trusting/acting on trust distinction. Michel Blais, for example, often talks of trust as the act of cooperation where Hardin would prefer him to talk of the act of cooperation as an act based on trust (Blais 1987 and 1990). I will attempt to describe the rational choice approach in a way that presupposes neither adherence to nor rejection of this distinction.

babysitter may be tempted to steal the loose change lying on the kitchen counter of her employer's house, a friend may be tempted to gain popularity by spreading her friend's secret, and a teacher may be tempted to pass off a student's idea as her own. Rational choice theorists want to know how it could be rational to trust given all the risks involved in making oneself vulnerable to abuse of power.

Second, rational choice theorists adopt a sparse set of theoretical tools to answer this question. With a minimalist spirit, they make as few assumptions as possible about the individuals involved in trusting relationships. All they assume about these individuals is that they are self-interested and possess instrumental rationality. This is not to say that the individuals are exclusively self-interested but merely that no altruistic motives are taken for granted. Their challenge is, therefore, to show how it could be rational for such people to trust each other. This is obviously very much a Hobbesian project, and as such it is fitting that rational choice theorists often use the simple models of game theory, such as the Prisoner's Dilemma, that have figured prominently in interpretations of Hobbes. I find it useful to compare the rational choice accounts of trust to two solutions to the problem of cooperation between Hobbesian agents. First, as Hobbes notes in his response to the fool, Hobbesian agents can find it in their self-interest to cooperate in the state of nature when cooperation will maintain a useful friendship (Hobbes 1650/1994, *xv*.4). Second, cooperative behavior is rational for Hobbesian agents when external constraints imposed by the sovereign make cooperation in one's self-interest. Rational choice theorists provide similar solutions to the problem of trust.

Hardin provides a schema for the first type of solution by giving a rational choice account of trust in general rather than focusing specifically on trust in science, as do Rescher, Blais and Adler. Hardin calls his account the "encapsulated-interest view." Like the other rational choice theorists, Hardin assumes little about the individuals involved in a trust relationship; all we assume about them is that they are self-interested and instrumentally rational: "[trust is] essentially rational expectations about the self-interested behavior of the trusted" (Hardin 2002, p.6). For Hardin, trust is simply belief that the trusted will be trustworthy. The question that interests him is: What are the reasons a truster has for believing that the other party will be trustworthy? In other words, what makes it rational for trusters to trust? His answer is as follows:

> On this account, I trust you because I think it is in your interest to take my interests in the relevant matter seriously in the following sense: You value the continuation of our relationship, and you therefore have your own interests in taking my interests into account. That is, you encapsulate my interests in your own interests. (Hardin 2002, p.1)

So, for example, if I have a profitable business relationship with Mary, it is in my self-interest to take Mary's interests into account. I want my relationship with Mary to continue, so I have a purely self-interested reason to be trustworthy to her given the risk of her detecting untrustworthiness and cutting off our relationship. If Mary has reason to believe that I value our relationship, then it is rational for her to trust me.[13] As with other rational choice theorists, Hardin grounds the rationality of trust in evidence that it is in the trusted's self-interest to be trustworthy. His innovation is to identify the way in which valuing ongoing relationships can make it in the trusted's self-interest to take the interests of the truster into consideration.[14]

[13] In fact, the situation is more complex than Hardin's initial presentation of trust as encapsulated interest suggests. In order for it to be rational for Mary to trust me, Mary ought to also have reason to believe that I know what her interests are. Even if Mary believes that it is in my self-interest to take her interests into account, it may not be wise of her to trust me if I am significantly mistaken about what her interests are.

[14] Hardin appears to allow for instances of trust as encapsulated interest that do not focus on ongoing relationships when he says that to say that I trust you "…is to say that you have an interest in attending to *my* interests because, **typically**, you want our relationship to continue" (Hardin 2002, p.4; my emphasis in bold). This would appear to leave room for trust in one-off interactions when individuals encapsulate each other's interests. However, when Hardin argues that incentive compatibility is not the same as trust

One way to flesh out Hardin's trust as encapsulated interest schema is to use the Prisoner's Dilemma to explain why it is rational for scientists to trust each other. Both Rescher and Blais take this approach. Rescher models trust situations as instances of an iterated Prisoner's Dilemma. He argues that "[a]s long as interagents react to cooperation with some tendency to reciprocation in future situations, cooperative behavior will yield long-run benefits" (Rescher 1989, p.37). The long-run benefit for scientists in trusting each other enough to share their information is a decreased chance of "coming up empty-handed" (Rescher 1989, p.38). Rescher thus argues that, despite the risks involved, trust can be rational in the long-run for self-interested scientists, as long as they reciprocate whenever they are trusted. In fact, the economic incentives make cooperation practically inevitable:

> If its cognitive needs and wants are strong enough, any group of mutually communicating, rational, dedicated inquirers is fated in the end to become a *community* of sorts, bound together by a shared practice of trust and cooperation, simply under the pressure of its economic advantage in the quest for knowledge. (Rescher 1989, p.43)

Blais also makes use of the Prisoner's Dilemma. He wants to show that "only cooperation, as defined … in game theory and as illustrated in the Prisoner's Dilemma, is necessary for the justification of vicarious knowledge" (Blais 1987, p.370). Relying on Axelrod's (1984) study of strategies in iterated Prisoner's Dilemma situations, Blais argues that "faith in the future behavior of another player is less necessary than durability of the relationship: as long as the 'shadow of the future' is sufficiently long, each player has less incentive to defect in order to maximize

he makes it clear that his account does not apply to one-off interactions: "Note that *our merely having the same interests with respect to some matter does not meet the condition of trust as encapsulated interest*, although it can often give me reason to expect you to do what I would want you to do or what would serve my interests (because it simultaneously serves yours). The encapsulated-interest account does entail that the truster and the trusted have compatible interest over at least some matters, but incentive compatibility, while necessary, is not sufficient for that account, which further requires that the trusted values the continuation of the relationship with the truster and has compatible interests at least in part for this reason" (Hardin 2002, pp.4-5).

payoffs in the short run; for repeated encounters favor cooperative behavior simply from an egoistic viewpoint" (Blais 1987, p.370). Blais' main point is that players in the knowledge game will find it in their interests to not defect by cheating, "by fudging, fabricating, or otherwise publishing unreliable results" (Blais 1987, p.370). What ensures that it will be in their interests to cooperate by not cheating? Here is a sketch of Blais' answer. First, scientists are engaged in long-term relationships with other members of the scientific community or with the community itself. Second, while there may be tempting immediate gains to be garnered from defecting (e.g. an enhanced reputation through publications), the scientific community can effectively retaliate in future interactions. Forms of retaliation include the destruction of a dishonest scientist's reputation, withdrawal of funding, censure and even exile from professional organizations. Finally, it is possible to detect defectors. As Blais puts it, "everyone serves as potential watchdog for everyone else" (Blais 1987, p.372). When these three conditions (long-term relationships, effective retaliation, and relatively easy detection of defectors) are in place, defection will not be an effective strategy for a self-interested scientist to pursue. Cooperation—trust in Blais' sense—is therefore rational.

Adler follows Hobbes' second solution to the problem of cooperation to show that cooperation is rational in science. On Hobbes' account of the sovereign's power, the constraints the sovereign places upon self-interested subjects make it rational for them to cooperate with each other. Similarly, Adler's account of the power of constraints in scientific practice argues that community-level constraints can make it in the best interest of a scientist to cooperate with her colleagues. According to Adler, "there are powerful constraints on informants to be truthful and reliable" (Adler 1994, p.267). These constraints include the peer review and replication process in which scientists' results are subjected to scrutiny from a variety of sources. As Adler notes, “the force of these constraints varies according to such factors as the institution or community’s sensitivity to the detection of deception or error, the costs to the informant once an error is detected, and the rapidity and extent of communication about these findings” (Adler 1994, p.267). An individual working in a community with rigorous methods of detecting defection, which it punishes with severe penalties, is more likely to find it in her own best interest to cooperate (rather than defect) than an individual whose community places weak constraints on her behavior.

These four rational choice theorists apply a sparse set of theoretical tools to the analysis of trust and trustworthiness. Their minimalist methodology leads them to understand trust in terms of self-interested rationality. This methodology makes three fundamental assumptions. These are assumptions about what is going on when person *A* trusts person *B* to do action *C*, and when *B* does do *C* as *A* trusts *B* to do.

(RC1): *A* trusts *B* to do *C* on the basis of evidence that *B* will do *C*.

(RC2): The evidence *A* relies upon is evidence that it is in *B*’s self-interest to do *C*.

(RC3): *B* acts trustworthily (by doing *C*) for self-interested reasons.

(RC1) claims that people’s reasons for trusting are evidential. (RC2) specifies what the evidence for trust is: it is evidence that the trusted has self-interested reasons for being trustworthy. According to (RC3), it is such self-interested reasons that account for trustworthy behavior. These are not presented as necessary conditions for trusting or trustworthy behavior. Instead, (RC1) and (RC2) are jointly sufficient conditions for trust to be rational; and (RC3) is a sufficient condition for an actor’s trustworthy act to be rational. *A*’s possession of good evidence that it is in *B*’s self-interest to do *C* is sufficient to make it rational for *A* to trust *B* to do *C*.

Similarly, *B*'s possession of self-interested reasons for acting trustworthily (by doing *C*) makes it rational for *B* to do *C*.

All four authors provide interesting accounts of trust that rest on these three assumptions. By focusing on the economics of cooperation, Rescher helps us understand how the need to share information binds a scientific community together. Hardin's explanation of how partners in a relationship can encapsulate each other's interests usefully explains why trust thrives in enduring relationships. In addition, Blais and Adler's emphasis on community level constraints that encourage trustworthy behavior by scientists usefully highlights the communal aspects of scientific trust. The rational choice approach certainly provides an important tool for understanding trust in science. However, it cannot explain all of the interesting features of trust between scientists or trust in general. Having described the rational choice approach, I will now explain what is missing in it and why I adopt an alternative approach to understanding trust and trustworthiness.

1.2 LIMITATIONS OF THE RATIONAL CHOICE APPROACH

Rational choice assumptions about trust, (RC1-RC3), provide useful insights into trust and trustworthiness, but they also overlook many other important aspects of trusting behavior. In particular, the rational choice approach is limited in four ways. First, it only recognizes one type of evidence used by trusters to determine whether the trusted will be trustworthy. Second, this approach only accounts for self-interested reasons for being trustworthy. Third, it does not leave room for non-evidential reasons for trusting. Fourth, it fails to mark the distinction between trust and mere reliance. It may be possible for the rational choice approach to deal with the first three limitations by adding to its account of both the evidence available to trusters and the reasons for trusting and being trustworthy. However, a simple extension of the rational choice approach will not enable it to mark the trust/mere reliance distinction that I shall outline. In what follows, I will explain the first three limitations in turn. Then I shall illustrate how these limitations combine to prevent the rational choice approach from accounting for cases of trust on the part of relatively powerless individuals. Finally, I shall address the fourth limitation, which would require more revision of the rational choice account to handle.

1.2.1 Other reasons for trust

One limitation of the rational choice approach is that (RC2) omits reasons for trust based on evidence about the trusted other than evidence that it is in the trusted's self-interest to act trustworthily. While the rational choice approach is certainly right that we often trust because we have evidence that it is in the trusted's self-interest to be trustworthy, it is also certain that we trust based on other types of evidence. For example, I may trust my colleague because I have evidence that she is strongly motivated by a sense of professional ethics. I may have heard her speak disparagingly of other colleagues who act only to further their self-interest, and I may have seen her sacrifice her own interests for the sake of adhering to professional codes of conduct. Such evidence need not be limited to trusting someone's professional ethics; one may trust based on evidence that someone is strongly motivated by a general sense of duty or honor.[15] Such

[15] One might wonder how one can distinguish evidence that someone is motivated by a sense of duty or honor from evidence that someone is motivated by a desire to be regarded as dutiful or honorable. In fact, a cynic might even ask whether a sense of duty can be distinguished from a desire to be regarded as dutiful or honorable. I am not such a cynic. I think people are in many ways motivated by a general sense of duty or honor or a particular sense of duty or honor in a certain sphere of their lives. This sense

evidence is not accounted for in the rational choice approach, which, as part of its minimalist strategy, takes only evidence of self-interested reasons to act trustworthily into account. Similarly, this approach cannot account for my trust based on evidence that the trusted cares for me. In a developing romantic relationship, people often look for cues that their partner cares deeply for them. That one's partner loves one is generally acknowledged to be a good reason to trust him or her. The only one of the rational choice approaches that perhaps has a chance of giving a plausible account of this type of trust is Hardin's view of trust as encapsulated interest. Hardin would argue that when I learn that my partner loves me, I take this as evidence that it is in his self-interest to maintain our relationship by acting trustworthily. This may account for many cases of trusting someone with whom one has a relationship, but it is an implausible account of all of them. It is not unheard of for people to trust their ex-boyfriends, ex-girlfriends, ex-spouses, or ex-friends even though the relationship has ended. Love or even just affection and caring can persist even though the relationship does not. Thus, I may trust my ex-boyfriend to return my books to me simply because I know he still cares about me and would not want to hurt me even though we may never speak again. Again, I am not arguing that the rational choice approach is wrong because it cannot deal with all types of trust. I merely wish to point out that (RC2) quite significantly restricts the type of evidence that leads people to trust.

of duty can be distinguished from a desire to appear dutiful or honorable because it leads people to act dutifully or honorably even when the desired audience is absent. I also think that we are fairly good at assessing whether someone is genuinely dutiful or honorable rather than merely acting so to enhance their reputation.

1.2.2 Other reasons for trustworthiness

Not only does (RC2) limit the rational choice approach's account of the range of reasons for trusting, but (RC3) limits its account of reasons for being trustworthy. This causes a second limitation of the approach: there are more reasons for being trustworthy other than merely finding it in one's self-interest to be trustworthy. I may act to live up to someone's trust in me out of a sense of honor or duty, or because I care about the person and have their interests at heart.[16] An account of trust and trustworthiness that ignores these reasons leaves out something significant about the rationality of trust.

It is a fact that one often trusts someone because one believes that she is a dutiful person or because she cares for one. We will have trouble accounting for the rationality of such trust if we do not acknowledge that love for someone or a sense of duty are reasons for trustworthy behavior. Acknowledging that these are reasons for trustworthiness forces us to ask important questions that are critical to assessing the rationality of trusting someone who one believes possesses such reasons. For example, suppose I trust my colleague to work diligently because I have good evidence to believe that she is motivated by a sense of professional honor. In judging whether it is rational for me to trust her to perform a particular part of her job, we would want to have answers to some of the following questions: Do I know whether she considers herself obliged to perform this part of her job diligently, or might she think that a dutiful professional could do sloppy work in this area? Do I know whether she has self-interested reasons for not performing that part of the job? If so, do I know how she tends to weigh self-interested reasons

[16] In general, the rational choice theorists' preference for a sparse set of theoretical tools leads them to adopt a limited notion of self-interested rationality according to which individuals are not motivated by altruistic motives. In addition, their desire to account for trust without needing to use moral concepts like a sense of duty means that one cannot interpret the self-interested individuals they discuss as having an interest in acting out of a sense of duty.

against reasons of duty? In other words, does she only posses a weak sense of duty, or are considerations of duty highly motivating for her? Taking reasons for trustworthiness other than mere self-interest seriously requires us to ask certain questions in assessing the rationality of trust.[17] Given that people do in fact trust based on evidence of these other reasons for trustworthiness, those who want to know what makes trust rational ought to ask these questions. Unfortunately, the rational choice view of why people act trustworthily does not provide the resources to do so.

1.2.3 Non-evidential reasons for trust

Third, as a result of (RC1), the rational choice approach is limited to explaining reasons for trust that are evidential; however, reasons for trust can also be non-evidential. I may trust someone on pragmatic grounds, e.g., I trust in order to achieve some goal. Richard Holton gives an example of trust for pragmatic reasons:

> Suppose you run a small shop. And suppose you discover that the person you have recently employed has just been convicted of petty theft. Should you trust him with the till? It appears that you can really decide whether or not to do so. And again it appears that you can do so without believing that he is trustworthy. Perhaps you think trust is the best way to draw him back into the moral

[17] To give another example, in assessing the rationality of trusting one's romantic partner, important questions are raised if we acknowledge that loving someone is a reason for being trustworthy to her. Take the issue of assurances of love; does the fact that one's partner says she loves you provide good reason for trusting her? Well, it depends. As Nancy Nyquist Potter notes, part of being a trustworthy person is giving assurances of one's trustworthiness (Potter 2002, p.26). Since loving someone is a reason for being trustworthy, telling one's partner that one loves them is a good means of assuring them of one's trustworthiness. Often, we consider the relationship to have taken on a new level of intimacy when verbal assurances of love have been made. However, we also tend to consider it unreasonable to take seriously someone who says "I love you" too early in the relationship. Verbal expressions of love can be used to try to manipulate one's partner. Therefore, a complete account of the rationality of trust would need to account for how reasonable people distinguish sincere expressions of love from insincere, potentially manipulative expressions.

> community; perhaps you simply think it is the way you ought to treat one of your employees. (Holton 2002, p.63)

In this case the shopkeeper decides to trust not because there is evidence of the employee's trustworthiness, but in order to achieve the goal of bringing the employee back into the moral community.[18] Trust can also be a leap of faith in which I trust even though I lack evidence that the trusted will act as I expect. The rational choice approach cannot explain cases of non-evidential trust. At best, the rational choice approach can explain cases of non-evidential trust as irrational trust. However, for reasons that will be discussed in detail in chapter two, I prefer to adopt an account of trust that does not preclude the rationality of some cases of non-evidential trust. Therefore, I will argue for an account of reasons for trust that is not limited to evidential reasons.

1.2.4 Powerlessness and trust

As a result of these three limitations, the rational choice approach has trouble accounting for trust on the part of someone in a position of relative powerlessness. The rational choice approach can nicely explain why I would trust someone when I have the ability to detect untrustworthy behavior and prevent it by placing retaliatory constraints on the trusted to make it in her interest to be trustworthy. Game theoretic models, like iterated Prisoner's Dilemmas, can also provide insights into continued cooperation between equals who are equally able to detect and punish defection. However, not everyone who trusts is in a position to detect potential defection and punish it with effective retaliation. There are two forms of powerlessness which

[18] Holton's example raises several pressing questions about whether we can choose to trust and whether trust requires belief. These questions will be dealt with in chapter two.

can make it difficult for someone to conform to the model of trust proposed by the rational choice approach: lack of power to influence and epistemic powerlessness.

Some people lack the power to influence the behavior of the trusted by making it in the trusted's self-interest to be trustworthy. All four rational choice theorists explain why it is rational to trust by pointing to ways in which the truster can attempt to constrain the behavior of the trusted.[19] For example, according to the iterated Prisoner's Dilemma accounts, there are two ways that the truster can make it in the trusted's self-interest to cooperate. Either the truster is able to punish uncooperative behavior on a future interaction, or the truster has something that the trusted hopes to gain from future cooperative interactions. Trusters who are unable to punish defection effectively or who have nothing to offer as future reciprocation lack the power to make it worthwhile to the trusted to act trustworthily. Similarly, according to Hardin's trust as encapsulated interest account, the truster can influence the trusted's behavior through the threat of cutting off the relationship if the trusted acts untrustworthily. Individuals who are caught in a relationship and have no viable opportunity to end it are unable to exercise this type of power.

The trust that graduate students, post-doctoral trainees and other scientists in apprenticeship positions place in their superiors illustrates this type of trust by relatively powerless individuals. Let us consider two examples of ways that graduate students trust despite having no viable opportunity to cut off their relationship with their advisors. Consider Lauri, a graduate student working in developmental genetics in a department with only one member of the faculty specializing in this subfield. Suppose that Lauri has slim chances of being able to

[19] In some cases it will not be the truster herself who enforces the constraints, but she will still have a role in constraining the trusted's behavior by setting up constraining mechanisms or by alerting the community to untrustworthy behavior so that the community can enforce its own constraining mechanisms. For example, individual scientists can alert the community to forgery so that the community can punish fraudulent scientists.

transfer to another program or change her specialization without prohibitive costs. If Lauri wants to pursue a scientific career, she has no viable opportunity to cut off her relationship with her advisor if the advisor acts untrustworthily. Nonetheless, Lauri assumes that her advisor will treat her well. She takes the risk that her advisor will turn out to be untrustworthy because it is the only viable way to achieve her goal of attaining her Ph.D., and she has no evidence that her advisor is an untrustworthy person. Further, Lauri trusts her advisor even though she has evidence that her department and university administration exercise poor oversight over the work of its professors. She has seen other graduate students severely sanctioned for blowing the whistle on fraudulent professors. She trusts even though she is in a position where neither she nor the surrounding scientific community exercise constraints that make it in the professor's self-interest to be a trustworthy mentor for her. Like the shopkeeper in Holton's example, Lauri trusts for pragmatic, non-evidential reasons.

Now consider Monica, who is in the same position as Lauri; she has no viable opportunity to cut off her advisor if the advisor proves untrustworthy, and neither Monica nor the surrounding scientific community exercise constraints that make it in the advisor's self-interest to be trustworthy. Monica, like Lauri, trusts her advisor. However, she does not do so for pragmatic reasons. Instead Monica trusts because she believes her advisor to be a morally upstanding professional. She trusts because she has evidence of her advisor's moral character. Both Lauri and Monica illustrate trust that is beyond the limits of the rational choice theory's explanatory power. Trust on the part of individuals who lack the power to influence the trusted's actions is an anomaly if we accept the rational choice model, because it provides no tools for explaining why it is rational for them to trust.

The second type of powerlessness is inability to detect untrustworthy behavior. This is a form of epistemic powerlessness that would undermine one's ability to influence the trusted's behavior. One cannot take one's ability to sanction untrustworthy behavior as a reason to expect trustworthiness if one is not even in a position to detect untrustworthiness that requires sanction. Scientists who collaborate with colleagues from a different field are in this position because they lack the expertise to understand their colleagues' work. As Paul Thagard notes, "Peer-different collaborators are exceptionally epistemically dependent on their coworkers, since they typically lack the skill to validate work done in a different field" (Thagard 1997, p.254). Scientists engaged in cross-disciplinary work are, therefore, epistemically powerless in this sense because they would not be able to detect fraud or sloppy work by their colleagues. Thus, any case in which one trusts someone but lacks the ability to detect untrustworthy behavior will present an anomaly for those rational choice theorists who argue that it is rational to trust when one has evidence that constraints like retaliation make it in the trusted's self-interest to be trustworthy. This kind of powerless has an epistemic dimension because it involves a lack of knowledge on the part of the truster.

There are also cases of what we might call epistemic powerlessness in which the truster lacks access to certain information about the trusted's interests and circumstances. Since rational choice theorists, who adhere to (RC1), explain rational trust in terms of the truster's having evidence that the trusted will be trustworthy, anyone who lacks information about the trusted's interests and circumstances will present a case of trust that the rational choice theorist cannot easily justify. If the rational choice approach is the only explanation of trust we have, then it is a mystery why people trust without this kind of information. Again, graduate students can experience this type of relative powerlessness. Graduate students need to be able to work in a particular laboratory or experimental environment for an extended period of time in order to complete the research for their dissertations. This environment is provided to them by their advisors in whose laboratories they do their research and on whose experiments they work. By beginning her work in a particular laboratory, a student runs the risk that her advisor will cut her off from the laboratory prematurely by firing her or moving to another institution. Being cut off from the lab of one's advisor disrupts and potentially halts altogether a student's research. However, graduate students are not usually privy to relevant information about their advisors' circumstances that could result in their being cut off from their advisors' laboratory. Advisors' decisions are often influenced by funding restrictions and opportunities as old grants are cancelled or new ones offered. Funding considerations may cause an advisor to shift the focus of her work or even move to another institution. Thus, any graduate student who is not privy to all the details of her advisor's funding situation is in a position of epistemic powerlessness. She lacks information that would aid her in assessing whether it is in her advisor's self-interest to remain her advisor. Nonetheless, one could still argue that a student in such a position may still reasonably trust her advisor not to abandon her prematurely. If the advisor has promised the student that she will not leave the institution, if at all possible, before the student finishes, and the student knows that her advisor refused to abandon a previous student when under funding pressures, then it seems reasonable for the student to trust her advisor not to leave. However one might make this argument, one would not be able to make it along the lines suggested by the rational choice account, which requires the truster to have enough information about the trusted's circumstances to be able to judge whether trustworthiness is in the trusted's self-interest. Therefore, the rational choice account is limited in its ability to account for trust on the part of the epistemically powerless.

At this point, a rational choice theorist might argue that these examples of powerlessness do not present cases of reasonable trust that fall outside the scope of rational choice explanations. There are two arguments such a theorist might make. First, she might argue that the powerless simply do not engage in trust. This is Hardin's approach. He addresses issues of powerlessness as follows:

> There are inherent problems in trusting another who has great power over one's prospects. In an iterated exchange between two relatively equal partners, both stand to lose more or less equally from the default of the other. If a much more powerful partner defaults, however, she might be able to exact benefits without reciprocating. Moreover, she might be able to dump partners willy-nilly and replace them with others, while they cannot dump her with such blissful unconcern because there may be few or no others who can play her role. Hence as in the discussion of endgame effects, the weaker party to an unequal relationship is at threat of seeing the interaction terminated at any time but is most likely not in a position actually to terminate it. (Hardin 2002, p.101)

Hardin does not take these problems facing the powerless as evidence that his approach cannot account for their trusting behavior. Instead, he raises these issues to explain why "[i]n general, therefore, the weaker party cannot trust the more powerful much at all. Inequalities of power therefore commonly block the possibility of trust" (Hardin 2002, p.101). While Hardin is quite right that the powerless do often distrust the powerful, this is only half the story. The important fact Hardin misses is that the powerless *do* often trust the powerful. Graduate students do trust their advisors and collaborators do trust their partners from different disciplines. Wherever we find cases of those who lack the power to detect and sanction the untrustworthy behavior of those they depend on, we have found a domain of trusting relationships which cannot be accounted for in the rational choice model.

That said, a rational choice theorist might have a second argument against this claim: she might concede that there are cases of trust by the powerless but argue that these are cases of irrational trust. She might argue that perhaps many individuals do trust when they lack evidence that it is in the trusted’s self-interest to be trustworthy, but these individuals are pursuing an irrationally risky strategy. Similarly, it is simply not wise to place one’s trust in someone who may find it in her best interest to be untrustworthy. However, while there may be many cases of powerless individuals unwisely placing their trust in untrustworthy people, this argument cannot explain away all cases of trust by the powerless. The fact that many cases of interdisciplinary collaboration in science, cases in which each partner lacks the expertise to detect untrustworthy behavior, are successful makes it implausible to discount such trust as irrational. Similarly, I do not think it is irrational of Lauri and Monica, the trusting graduate students, to trust even though they lack the power to make it in their advisors’ self-interest to be trustworthy. The rational choice approach is limited because it provides no theoretical tools to understand such perfectly reasonable, and not unusual, instances of trust by the relatively powerless.

Note that Lauri and Monica’s cases illustrate different limitations of the rational choice approach. Lauri, who has non-evidential reasons for trusting her advisor, illustrates the third limitation of the rational choice account that is preventing it from justifying trust by the powerless. Since it does not account for pragmatic reasons for trust, the rational choice approach cannot explain why it is rational for Lauri to trust when she has no other viable advisor with whom to work. Monica, who trusts because she has evidence of her advisor’s moral character, trusts based on evidence not acknowledged by the rational choice account. This is a type of evidence that frequently gives people in powerless positions reason to trust. Such trusters cannot easily be dismissed as trusting irrationally. If we are to follow the rational choice approach, we will have to take the unpalatable position of labeling many common and pervasive interactions irrational, including many employer/employee, parent/child, doctor/patient, professional/client, and cross-disciplinary relationships. In fact, many relationships between people who trust each other despite differences and difficult epistemic locations will have to be dismissed as irrational. Taking this position simply to avoid having to formulate an account of trust to supplement the rational choice approach is an extreme move.

Issues of powerlessness also illustrate the limitations of (RC3)—the assumption that the trusted acts trustworthily for self-interested reasons. This assumption about the reasons for trustworthy behavior limits our ability to account for trustworthiness rooted in recognition of one's moral responsibility to be trustworthy, especially when one is in a position of power. In general, rational choice theorists avoid delving into issues of moral responsibility. This is partly because their primary interest is in questions about the rationality of trust rather than questions about the morality of trust. However, this silence on questions of moral responsibility is also caused by their minimalist desire to provide the simple explanations of rational trust behavior. Blais, Hardin and Rescher all express the hope that their accounts of trust can succeed without needing to bring in moral terms that they consider vague or overly complex. Blais' attempt to explain the collective knowledge acquired by cooperation between scientists illustrates this evasion of moral language. He states one of his goals as "to suggest that the kind of trust that is needed in any such collective system of beliefs can be illuminated by these results stemming from the Prisoner's Dilemma, *provided that the concept of trust be taken not in the moral, but in a strategic sense*" (Blais 1987, p. 363 my emphasis). Later, he explains what he means by taking trust in a strategic rather than moral sense:

> In science, cooperation is of the essence. This type of cooperation does not require trust in the moral sense. It is not necessary to assume that trustworthiness is a moral virtue of the trustee, or that trust be construed as a "confident expectation of something; hope" (*Oxford English Dictionary*)…. I should like to model trust as a *strategy*, rather than as some state of the truster perhaps related to some other state of the trustee…. The idea is to see how far we can go in assuming that cooperation in the knowledge game is justified, even if the players of the game have no such moral virtues. (Blais 1987, p.370)

Blais does not deny the existence of moral virtues, but he is trying to see how much of scientific cooperation he can explain without referring to them. Other rational choice writers agree that there may be more to trust and trustworthiness than just self-interested rational behavior, but, like Blais, they see themselves as providing a useful, if not essential, initial explanation of the central cases of trust. If, after an acceptable rational choice account of trust behavior has been provided, more needs to be said about some peripheral instances of trusting and trustworthiness, then perhaps the language of virtues, character and moral responsibilities can be brought in to fill the gaps. This is Hardin's approach:

> Many writings on trust convey a vague sense that trust always requires more than rational expectations grounded in the likely interests of the trusted. If this sense is correct, then we are at a very early stage in the development of any theory to account for trust or even to characterize it in many contexts. If an account from interests is largely correct for a large and important fraction of our trusting relationships, however, we already have the elements of a theory of trust that merely wants careful articulation and application. (Hardin 2002, p.6)

Rescher also argues that cooperation “need not ensue from a moral dedication to the good of others and care for their interests” (Rescher 1989, p.43). The rational choice approach, therefore, attempts to explain why individuals trust and act trustworthily in different contexts by showing why it would be in the self-interest of a rational agent to do so. Explicitly moral language is avoided in these explanations, and the test of a successful rational choice theory of trusting behavior is its ability to provide this type of explanation for a wide range of trusting relationships.

Unfortunately for the rational choice theorists, we do need to bring in moral concepts like responsibility to account for some important aspects of people’s reasons for being trustworthy. While Blais claims that “it is not necessary to assume that trustworthiness is a moral virtue of the trustee…,” Nancy Nyquist Potter argues persuasively that it *is* necessary to assume that trustworthiness is a moral virtue in order to account for the responsibility that the powerful bear towards the relatively powerless. Citing Annette Baier's work on power and trust, Potter maintains:

> Trust itself alters power positions (Baier 1986, 240): trusting others involves depending on them, being vulnerable to the possibility of disappointment or betrayal, and risking harm to self. This further feature of trust, in turn, indicates a moral requirement of the one *being trusted*: being worthy of another's trust requires that one takes care to ensure that one does not exploit the potential power that one has to do harm to the trusting person. (Potter 2002, pp. 9-10)

Potter is arguing that the power that the trusted has over the truster implies a moral responsibility on the part of the trusted. Throughout her book, she argues that the powerful have a moral obligation to cultivate a trustworthy character, one that recognizes "the importance of being trustworthy to the disenfranchised and oppressed" (Potter 2002, p.29). Thus, her interest in power relations spurs her to adopt a virtue theory of trust and trustworthiness. She justifies this focus on character as follows:

> The locus of trust is on character because, when differences in privilege and power exist between us, we may be uneasy about what each other cares about: each sees that the other values some things which she or he sees as either incompatible with or hostile to the things *she* or *he* values. Hence, **the emphasis is on how willing and able one is to care for the goods others value even when those are not, or do not appear to be, entirely harmonious with the goods one values oneself**. However, differences in power and privilege make it more difficult to assess the trustworthiness of others, so it is important to give and receive assurances of our trustworthiness. (Potter 2002, p.12 my emphasis in bold)

Those who are trusted have the power to hurt the interests of trusters, and in cases of differences in privilege, the trusters may fear that the trusted will use that power to pursue goods that are incompatible with those things they themselves value. This is easily illustrated in employer/employee relationships as seen in debates about the labor practices of large retail chains. The employees of a large retail chain trust the company to provide safe and healthy conditions for their work; they place their valued safety and health in the hands of their employer. The employer thus has the power to hurt the interests of its employees by failing to provide adequate working conditions. Additionally, there is a huge difference in power between the employer and its non-union employees. Critics of certain large retail chains worry that the company's interest in its profit margins is in tension with the employees' interest in good working conditions. Of course, this tension is present in most employer/employee relations. However, some corporations are motivated by a sense of corporate ethics. These companies believe that they have a moral responsibility to care for their employees' health even when doing so is not entirely harmonious with their own interest in profits. When critics argue that certain employers are lacking something important that these other companies possess, Potter's analysis suggests that they are arguing that some employers are not as trustworthy as others. In this case, using the language of moral responsibility allows us to say something important about the trustworthiness of corporations; and the rational choice approach does not provide us with these kinds of conceptual tools.

1.2.5 The trust/mere reliance distinction

Lastly, the rational choice approach fails to mark the conceptual distinction between trust and reliance (and correlatively between trustworthiness and reliability). Baier makes this distinction when she says "We can still rely where we no longer trust" (Baier 1994, p.98). Baier illustrates the distinction with the following example:

> Once we have ceased to trust our fellows, we may rely on their fear of the newly appointed security guards in shops to deter them from injecting poison into the food on the shelves. We may rely on the shopkeeper's concern for his profits to motivate him to take effective precautions against poisoners and also trust him to *want* his customers not to be harmed by his products, at least as long as this want

> can be satisfied without frustrating his wish to increase his profits. (Baier 1994, p.99)

This example suggests that there may *be* a difference between relying on someone and trusting them, but what *marks* the difference in these two ways of depending on someone? Baier continues:

> We all depend on one another's psychology in countless ways, but this is not yet to trust them. The trusting can be betrayed, or at least let down, and not just disappointed. Kant's neighbors who counted on his regular habits as a clock for their own less automatically regular ones might be disappointed with him if he slept in one day, but not let down by him, let alone had their trust betrayed. (Baier 1994, p.99)

When one merely relies on a person, one takes a similar attitude towards him or her as one takes towards a clock. A good clock can be counted on to tell the time accurately, and one may reasonably alter one's behavior based on its behavior. But one does not feel that the clock has betrayed one or failed to meet a responsibility when it is inaccurate. In short, one's relationship with the clock's activities does not carry moral weight. This is because clocks are not part of our moral community, so norms do not apply to our interactions with them. One cannot make a contract with a clock or enter into a norm-governed relationship with it. One element of our education is to learn which norms apply to different kinds of interactions, and we learn that inanimate objects are not subject to the same norms as those governing interactions with other people. In learning this, we learn when certain emotional responses are appropriate (for example, we learn that it is inappropriate to feel betrayed by a clock). Thus, part of what it means to be a mature moral being is to have the emotional responses that are appropriate to the situation at hand. This is why we generally do not feel betrayed by our clocks (or if we do feel betrayed we are likely to acknowledge the feeling as silly). I will say that the kind of relationship we have with a clock is one of ‘mere reliance’, as opposed to one of trust. In general, our relationships with inanimate objects are relationships of mere reliance.

Our relationships with people are more complex. Sometimes our interactions with people resemble our relationships to inanimate objects. People can merely rely on other people. Baier’s example of Kant and his neighbors illustrates this nicely. The neighbors who used Kant’s regular walks about town as an indicator of the time merely relied on him. They were not betrayed by his failure to act as a good indicator of the time on the day he slept in. Their relationship with him (if it can even be called a relationship) did not carry that kind of moral weight.[20] However, most of our interactions with people are governed by norms. When one counts on someone in a context governed by norms, one has a different relationship with that person than one has with an inanimate object. These types of relationships are the kind of relationships that Baier called trusting relationships. In contrast to relationships of mere reliance, trusting relationships do possess the possibility of betrayal.[21] When a child’s trust is broken, the child can be said to have been betrayed. Infidelity on the part of a trusted spouse is a betrayal, and the disloyal spouse is subject to moral blame.

Even though Baier’s main interest is in answering a different question than the one of interest to the rational choice theorists, her distinction between reliance and trust suggests one line of argument against their approach. Baier is trying to determine what distinguishes trust

[20] It is worth noting that whether one is inclined to grant moral weight to Kant’s relationship with his neighbors depends on how one imagines several details of the story. If we imagine that Kant entered into an agreement with them to try to stick to a timely walk so that they could keep the time, then I think we no longer want to say that Kant cannot betray his neighbors. In fact, I think intuitions could change if we just imagine that Kant is aware that his neighbors rely upon him. When I say that the relationship does not carry this kind of moral weight, I am taking Baier’s example to be that the neighbors have decided to rely upon him without his consent or knowledge.

[21] Baier relies heavily on the concept of betrayal to mark the trust/reliance distinction. Unfortunately, she says little about what exactly betrayal is. In chapter two, I will provide an account of betrayal that will fill this gap in her account.

from other social phenomena; the rational choice theorists instead focus on what makes trust rational. However, notice how Baier's argument that trust involves vulnerability to betrayal has implications for the question of when trust is rational. If trust involves vulnerability to betrayal, then *rational* trust will involve *appropriate*[22] vulnerability to betrayal. In assessing the rationality of someone's trust, one must therefore ask questions like whether it is appropriate to feel betrayed by the trusted when the trusted fails to fulfill one's expectations. Thus, an answer to the question *What makes trust rational?* that is based on a recognition of the trust/reliance distinction would include something like the following claim: If *A* is trusting rationally, then it is appropriate for *A* to feel betrayal when the trusted fails to fulfill the truster's expectations. This is not part of the rational choice account of trust. Thus, the rational choice approach presupposes an account of trust that does not mark the trust/reliance distinction. Once one appreciates the distinction, one might be tempted to dismiss rational choice accounts on the grounds that they are accounts of the rationality of reliance that do not cast much light on trusting behavior.

To elaborate, Baier's distinction suggests the following argument against the rational choice approach. If trust is "essentially rational expectations about the self-interested behavior of the trusted" as Hardin and the rational choice theorists claim, then when one trusts one is vulnerable to being disappointed by one's expectations (because one has miscalculated the effects of external circumstances on what is in the trusted's self-interest to do), but one is not vulnerable to betrayal. However, when one trusts one is vulnerable to betrayal. So, trust must not be what the rational choice account tells us it is. That is, if I rationally expect that someone will do something because I think it is in her self-interest to do so, then I am justified in feeling

[22] As will be shown in chapter two, it is possible for one's trust to be problematic not only because one trusts without good reason to do so, but also because the nature of one's relationship with the trusted makes feelings of betrayal inappropriate.

disappointed when she acts contrary to my expectations. However, I would not be justified in feeling betrayed, because I simply did not see the external circumstances clearly enough. Were I to have recognized the situation for what it was, I would have predicted that the self-interested person would not act as expected. The failure here is largely my own. I am the one who expected the other to do whatever was in her self-interest. I was disappointed by my own evaluation of what was in the self-interest of the other. The proper response to such a disappointment is to think "Oh, I should have known better." So if Baier is correct that if one reasonably trusts someone, one is entitled to feel betrayed by her untrustworthy behavior, then we are led to the conclusion that the rational choice approach does not deliver an account of trust.[23]

The following thought experiment adds to the attractiveness of this criticism of the rational choice account. Consider two scientists from different fields. Alice needs Betty to conduct an experiment that Alice lacks the expertise to perform or supervise. Alice is concerned because Alice knows that Betty is a self-serving individual who has no moral compunction about giving her colleagues sloppy data if she can get away with it. But luckily for Alice, there is another scientist, Claire, who works in Betty's lab. Alice and Claire are good friends, and Claire offers to check over Betty's data for Alice because Claire knows, and cares about, how important this experiment is for Alice. Betty knows that Claire will be checking the data. Alice now expects that Betty will not give her sloppy data because she thinks Betty knows that Claire will expose Betty's carelessness and damage her professional reputation if Betty does so. Unfortunately for Alice, when the time comes, Claire is overwhelmed with work. Betty notices this and takes the opportunity to slip by some sloppy work. Claire gives Betty's data only a

[23] Potter provides a similar argument (Potter 2005, p.5).

passing glance before falsely telling Alice that everything is fine with it. As a result, Alice's own project grinds to a halt because of flaws in Betty's data. Now, how is Alice to respond when both Betty's and Claire's carelessness is uncovered? The key question for our purposes is whether Alice would justifiably feel betrayed by either Betty or Claire. Intuitively, I think we want to say that Alice would be entitled to feel betrayed by Claire but not by Betty. We might explain this intuition by saying that it was Claire on whom Alice was really counting. Alice fully expected that Betty would disappoint her if Betty could get away with it. So, Alice's attitude towards Betty will not likely change as a result of this disappointment. The same is not true of Alice's attitude towards Claire. Alice probably believed that Claire cared enough about Alice to look out for her even when it was inconvenient. Alice now sees that this is not the case, and Alice's attitude will change. Alice will feel betrayed by someone she thought was a trustworthy friend. If Baier is right that we justifiably feel betrayed when those we trust let us down, then it is clear that Alice did not trust Betty. Alice did rely on Betty for the data, but theirs was not a trusting relationship. This is problematic for the rational choice theorists because Alice's expectation that Betty would produce good data would be a paradigmatic case of trust for them. Alice had reason to expect that Betty's self-interest would, under these circumstances, lead Betty to act in a certain manner. One might, therefore, conclude that Baier's distinction between reliance and trust suggests that the rational choice approach gives us interesting accounts of reliance and reliability but not trust and trustworthiness.

But is this the right conclusion to draw? Is this enough of an argument to support abandoning the rational choice approach to trust and trustworthiness? While this argument and thought experiment reveal something important about the group of behaviors and dispositions we call trust and trustworthiness, I do not think we should dispense with the rational choice approach so quickly. The main reason for this is that I find it unproductive to argue about what we really mean by the English verb 'to trust'. Even a cursory survey of common usage of 'trust' shows that we use the term to refer to both what Baier calls reliance *and* what she calls trust. We say that we trust objects of which we do not make moral judgments. Witness the perfectly everyday assertion: "I trust my car to get me to work each day," and these more cultured examples from the Oxford English Dictionary's entry on trust: "He trusts much more to the Sun, for his Guide, than to the Creator of it" and "The mushrooms, that grow in meadows, are of the best kind: all others are dangerously trusted." We can, and do, use 'trust' to mean 'reliance' in Baier's sense. Therefore, we should not take her account of the difference between trust and reliance as arguing that vulnerability to betrayal is a necessary condition for the competent application of the English word 'trust'. The actual situation is that our use of 'trust' covers two very different ways in which we count on objects and people (including, in the case of self-trust, oneself). Her distinction, when applied to interpersonal interactions, helps us see that sometimes we count on people in ways that make us vulnerable to betrayal and sometimes we count on them in the way we count on a clock or a car. Depending on the social context and the norms governing the interactions, sometimes our counting on them means we hold them to a particular kind of moral standard and sometimes our counting on them has no such moral weight attached to it. We should not, therefore, argue that the rational choice theorists have no right to describe their project as an account of trust and trustworthiness.

Indeed, there are lessons to be learned from the earlier argument and its attendant thought experiment. First, we learn that we can recognize the distinction Baier makes in concrete cases. In thinking about Alice's differing attitudes towards Betty and Claire, we can see that Alice counts on Claire in a different way than Alice counts on Betty. This example suggests that there is something important about Alice and Claire’s relationship that is not accounted for by the rational choice account. As we have seen, some rational choice theorists claim that we can reduce trust to reliance on self-interested behavior without losing sight of the interesting phenomena. But we should not be so hasty to rush to this conclusion. There are clearly ways of depending on people that do not fit the rational choice model, and my slightly contrived thought experiment suggests that some of these may be found in interactions between scientists. Therefore, my analysis of trust between scientists departs from the rational choice approach and instead provides an account of trust and trustworthiness, in Baier’s sense. The fact that trust involves vulnerability to betrayal is an essential part of my account of trust.

In conclusion, the rational choice approach to trust and trustworthiness has much to offer. It provides a useful starting point to understand some of the dynamics of trusting relationships, both between individuals and within a community as a whole. However, like most theoretical tools, the rational choice approach has limited scope. It fails to recognize that we often act trustworthily for reasons other than self-interest, and it also overlooks that we frequently trust people because we see those other motivations in them. In addition, the rational choice approach fails to address non-evidential reasons for trust. By providing such a limited view of reasons that motivate people’s actions in trusting relationships, the rational choice approach cannot adequately explain trusting relationships where the parties have unequal power. Finally, the rational choice approach cannot mark the distinction between trust and reliance. Having outlined some of the limitations of the dominant approach to trust, the next chapter provides an alternative account that lacks these limitations.

2.0 AN ACCOUNT OF TRUST

Before I introduce and defend my account of trust, it will be helpful to have on the table a set of criteria of adequacy by which to judge it. Some of these goals must be met by all legitimate accounts of trust, while others need only be met by accounts aimed at capturing a certain range of trust phenomena. The criteria of adequacy are as follows. 1) An account must show how trust involves vulnerability and risk. Reliance and trust are *similar* in that by trusting or relying on someone we are vulnerable to being let down. In showing where the vulnerability lies in trusting, my account will clarify the close relationship between trust and reliance. 2) An account must mark the *distinction* between reliance and trust by showing how trust involves vulnerability to betrayal. As we saw in chapter one, one can, following Baier, make a distinction between reliance and trust by pointing to the particular kind of vulnerability to which the trusting are susceptible; while those who merely rely can be vulnerable to being disappointed, they are not, unlike the trusting, vulnerable to betrayal. An account of trust phenomena needs to explain what betrayal is and how it can be used to delineate the trust/reliance boundary. 3) Any account of trust must apply to, and unify, many different cases of what common usage would call trust. An account which is overly narrow and covers only an idiosyncratic group of cases of trust cannot be legitimately called an account of trust. A successful account of trust also needs to show what cases of the relevant type of trust have in common; it should unify the phenomena under study without falling prey to familiar counterexamples. After explaining my account of trust, I return to these criteria and argue that the account meets all three of them.

My account of trust is as follows:

(T) *A* trusts *B* to do *C iff (1) A* takes the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background, *(2)* in part, because *A* believes that their relationship morally obliges *B* to do *C.*[24]

Clearly, both the nature of the cognitive attitude found in (1) and the kind of relationship and obligations at issue in (2) need explication and justification. I address each in turn.

2.1 THE COGNITIVE ATTITUDE OF TRUST

2.1.1 Does trust require belief?

There is debate in the trust literature about whether trust requires belief that the trusted will act as expected. When I sincerely say, "I trust my friend to keep my secret," do I have the belief that she will keep my secret to herself? If I trust my spouse to remain faithful, do I believe that my spouse will not cheat on me? Several commonly shared, and potentially contradictory, intuitions about trust make it difficult to answer these questions. First, many authors share Baier's intuition that we do not ordinarily choose to trust (e.g. Baier 1994, Hieronymi 2008).[25] At first glance, it seems that trust is not under voluntary control. If we do not trust someone, we cannot decide to change our position and trust her. This intuition potentially explains why the

[24] This account shares many features in common with Holton's account of trust as reliance from the participant stance (Holton 1994).

[25] As Holton notes, Baier's views on this issue appear to evolve. In "Trust and its Vulnerabilities" she changes her statement to "trusting is rarely something we *decide* to do" (Baier 1994, p.141) from the stronger claim in "Trust and Antitrust" that "[t]he child, of course, cannot trust at will any more than experienced adults can" (Baier 1994, p.110).

demand “Trust me!” is so impotent (Baier 1994, p.110)**.** The parents of a teenager may fervently want to trust their child but feel that the child’s past irresponsible behavior makes it impossible for them to choose to trust, even though the child asks the parents to “just trust me.” This intuition that trust is involuntary is taken as evidence for the view that trust requires belief that the trusted will act as expected, because it is commonly argued that belief is involuntary (Williams 1973). Thus some conclude that since we cannot will ourselves to believe whatever we like, we cannot choose to trust whomever we like.

This view is undermined when we reflect on some of the reasons we cite for our trust in others. Some of these reasons do not appear to be reasons one could cite to support a belief that someone will do something. One can trust because one thinks one ought to trust people with whom one has a certain kind of relationship, or because one thinks that one’s interactions with someone will go more smoothly or simply if one trusts. For example, recall Holton’s example of the shopkeeper who decides to trust the convicted thief with the till. The shopkeeper, who chooses to trust because that is how she feels she ought to treat her employees or because she hopes to draw the thief back into the moral community, is not trusting for reasons that constitute evidence for the belief that the thief will not steal. Similarly, Lauri, the graduate student who trusts her advisor because she has no other alternative, may be choosing to trust because she thinks the relationship will be simpler and smoother if she takes a trustful rather than distrustful attitude. So here, as in the shopkeeper example, we have a case where one has reasons for trusting which could not be reasons for belief that the trusted will act as expected. Non-evidential reasons for trusting create problems for the view that trust requires belief that the trusted will act as expected. If one can have reasons adequate to justify trust but not adequate to justify belief, then it seems that trust does not require belief that the trusted will act as expected.

On my account, trust does not *require* belief that the trusted will act as expected. Often, one *does* trust on the basis of one's belief that the trusted will live up to one's expectations, but this is not always the case. There are cases of trust when the truster cannot sensibly be said to have the belief that the trusted will act as expected. Even though this account denies that trust requires belief, it does not claim that trust has no cognitive element. Trusters do have some cognitive attitude towards the proposition that the trusted will come through for them. On my account, *A* either *believes* or *accepts* that *B* will do *C* and this belief or acceptance is the basis of *A*'s practical reasoning. This is what it means to say that the proposition that *B* will do *C* is part of *A*'s adjusted cognitive background. Having situated this account of trust in the context of the debate about the role of belief in trust, I now explicate the key notions of belief, acceptance and adjusted cognitive background.

2.1.2 The distinction between trust and acceptance

In "Practical Reasoning and Acceptance in a Context," Michael Bratman usefully outlines a distinction between belief and acceptance. In general, reasonable belief has four features that acceptance lacks (Bratman 1992, pp.3-4). First, reasonable belief is context-independent. My beliefs do not change as I move from one intellectual or practical context to another. Second, reasonable belief is "shaped primarily by evidence for what is believed and concern for the truth of what is believed" (Bratman 1992, p.3). In other words, belief aims at truth. Third, we do not normally have direct voluntary control over our beliefs. Fourth, an agent's beliefs are subject to demands for consistency and coherence. We aim at having a coherent belief system. In contrast, acceptance is context-dependent, shaped by factors other than evidence, voluntary, and exempt from demands for overall consistency across contexts. These four features distinguish belief from acceptance.

Bratman's argument for the belief/acceptance distinction proceeds by presentation of several examples of context-dependent acceptance. For instance:

> The three of us need jointly to decide whether to build a house together. We agree to base our deliberations on the assumption that the total cost of the project will include the top of the estimated range offered by each of the sub-contractors. We facilitate our group deliberations and decisions by agreeing on a common framework of assumptions. We each accept these assumptions in this context, the context of our group's deliberations, even though it may well be that none of us believes these assumptions or accepts them in other, more individualistic contexts. (Bratman 1992, p.7)

In this case, our building group has decided to use the highest estimated prices for materials and labor in our practical reasoning about the cost because it will make our work smoother. This is a case where I can legitimately accept a set of assumptions in this one context that I would not accept in another context; were I asked to place a bet on the cost of the house, I would not take the highest sub-contractor estimates for granted in my calculations. This example also illustrates how reasonable acceptance, unlike belief, does not necessarily aim at truth. One can have pragmatic reasons for accepting a proposition in a given context; in this case, we have a pragmatic interest in simplifying our group deliberations. The building group's acceptance of the cost framework is voluntary. Thus, it does not have the third feature of belief. Finally, we would find it strange were someone to criticize the group for accepting the high cost estimate on the grounds that it is inconsistent with the set of propositions the group accepted when they were trying to figure out what the cheapest price for the house might be. Thus, sometimes we adopt a cognitive attitude of acceptance which is not subject to the ideals of consistency and coherence across contexts.[26]

Bratman argues for a model of practical planning and deliberation, in which an agent's beliefs create a "*default cognitive background*" that can be adjusted to suit practical reasoning about what to do in a specific context (Bratman 1992, p.10). We bring to all contexts a set of involuntary beliefs that are subject to demands for evidence and consistency. However, depending on the nature of the particular context at hand, we can bracket a belief that *p,* which is part of our cognitive background, or we can posit that *p* despite not maintaining a belief that *p* in the default background. We thus engage in practical reasoning in a specific context based on our "*context-relative adjusted cognitive background*" (Bratman 1992, p.11). This adjusted cognitive background includes all the un-bracketed propositions that we believe as well as all the propositions we have accepted for this particular context.

There are a number of types of practical pressures that can cause us to adjust our cognitive background by accepting that *p* when we do not believe that *p,* or by bracketing our belief that *p*. These pressures include the need to simplify one's reasoning, to take into account asymmetries in the costs of errors, to facilitate social cooperation, and to satisfy the pre-conditions for any practical reasoning at all. The building example illustrates how both the demands of social cooperation and need to simplify our reasoning can lead us to choose to posit a cost for the building that we would not accept in individualistic contexts. A cautious driver who operates on the assumption that all the drivers around her are driving drunk when she drives at night on New Year's Eve is someone who accepts a proposition (that the other drivers are

[26] Acceptance is, however, subject to a demand for consistency within a given context. The premises one uses in one's practical reasoning about a particular course of action ought to be consistent. In fact, as Holton notes, in observing someone's plans, we take evidence that their plans are based on a particular premise as evidence that the person does not accept premises inconsistent with that premise (Holton 1994, p.72).

drunk) in a particular context (on New Year's Eve) because of the asymmetries in the costs of errors (it will cost her little to make this assumption and could cost her a lot not to make it and drive less cautiously). Finally, a soldier who doubts that she will survive the day of battle ahead illustrates the need to accept certain propositions as a precondition for practical reasoning (Bratman 1992, p.8). Even though the soldier doubts that she will survive the day, she nonetheless needs to accept that she will so that she can engage in necessary practical reasoning about tomorrow's schedule. In all these examples, individuals are led by pragmatic considerations to accept certain propositions as the basis of their practical reasoning.

To accept that *p* is to choose to take *p* as a premise in one's practical reasoning. It is to make *p* part of one's adjusted cognitive background for practical reasoning in the particular context at issue. Accepting that *p* is not the same as believing that *p* with a low degree of confidence. Bratman uses the following example to illustrate this point:

> I have a chair and a two-story ladder. In each case I think it equally and highly likely that it is in good condition. Indeed, if you offered me a monetary bet about whether the chair/ladder was in good condition I would accept exactly the same odds for each object. But when I think about using the chair/ladder things change. When I consider using the chair I simply take it for granted that it is in working order; but when I am about to use the ladder I do not take this for granted. (Bratman 1992, p.7)

In this case, the asymmetries in the costs of errors provide reason to accept one proposition and not the other, even though one has the same degree of confidence in each. Accepting that *p* is also different from supposing that *p* and pretending that *p*. In short, "Context-relative acceptance is tied more directly to action than is mere supposition; and it is tied more directly to practical reasoning than is mere pretence" (Bratman 1992, p.9). These distinctions are particularly important to understanding trust, so I discuss supposition and pretence further after having explained how trust involves taking the proposition that the trusted will act as expected as part of one's adjusted cognitive background.

2.1.3 *A* takes the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background

When *A* trusts *B* to do *C*, *A* has the proposition that *B* will do *C* as part of her adjusted cognitive background. When we trust someone to do something, we make plans based on the assumption that she will come through for us. This assumption is part of the background of our deliberation. Sometimes we may make plans based on the assumption that she will come through for us even when we do not have good evidence to support the belief that she will do so. So sometimes our trust is based on acceptance, rather than belief, that the trusted will act as expected. In the example of Lauri the graduate student, she may not have enough evidence to support a belief that her advisor will treat her well, but she may still choose to make plans for her graduate program based on that assumption. Therefore, one way that the proposition that *B* will do *C* can be part of *A*'s adjusted cognitive background is by *A* accepting the proposition for pragmatic, non-evidential reasons.

Another way that the proposition that *B* will do *C* can be part of *A*'s adjusted cognitive background is by *A* believing it. If *A* has a context-independent belief that *B* will do *C*, *A* may find herself in a specific context in which there is no reason to bracket this belief. In this case, *A*'s trust in *B* to do *C* is based on the belief that *B* will do *C*. In this situation, *A* uses the premise that *B* will do *C* in her practical reasoning—*A* makes plans based on the assumption that *B* will do *C*. Thus, this account is broader than accounts of trust which require that trust involves belief (e.g. Hieronymi 2008). Trust may involve belief that someone will do something, but it may instead just involve acceptance that someone will do something. In either case, trust involves taking the premise that someone will do something as the basis of one's practical reasoning about what to do in a particular context, which is to say that trust involves having the proposition that someone will do something as part of one's adjusted cognitive background.

One of the primary virtues of this account is that it recognizes that trust is context-dependent. Whether or not we trust someone depends crucially on the details of the context of trust. I may trust my friend Leslie not to tell my secret when exposure would do me little harm, but I may also not trust Leslie with same secret in an environment where it would seriously damage my reputation. Situations like this are easily accounted for on my account. As Bratman points out, the difference between what we reasonably accept in one context and do not accept in another can be strongly influenced by asymmetries in the costs of errors (Bratman 1992, p.7). Thus, I have one of two attitudes towards Leslie. Either I believe that she will keep my secret and I choose to bracket this belief when the costs of error are high, or I doubt that she will keep my secret but I nonetheless choose to accept that she will keep it because the costs of errors are low, and I think trusting her will have pragmatic benefits (for example, the benefit of cementing my relationship with her). Thus my trust in Leslie to keep my secret in the less risky context may be based on either acceptance or belief. Our ability to adjust our cognitive background for practical reasoning in light of the details of the particular context of deliberation nicely explains why trust is context-dependent. Having explicated and justified part (1) of my account of trust, I turn to objections.

2.1.4 Objections and replies

What about the initial intuition that trusting is not subject to voluntary control? If some cases of trust involve acceptance, rather than belief, that someone will do something, then some cases of trust *are* under voluntary control. While my account denies that trusting can never be done at will, it can nonetheless accommodate some of the considerations that I take to be behind this intuition that trusting is not voluntary. The key is to recognize that *acceptance* is not *supposition*. Bratman makes a few brief remarks about this distinction, and elaboration of it can allay concerns about the voluntariness of trusting. According to Bratman, "Context-relative acceptance is tied more directly to action than is mere supposition" (Bratman 1992, p.9). He uses the following example to illustrate supposition:

> "Suppose I had a million dollars", I ask myself. "What should I do with it?" Such a question may trigger contingency planning based on the mere supposition that I have such wealth. But this planning will not directly shape my action. If I conclude, for example, that with such wealth I should invest in General Motors my conclusion will not lead directly to my calling up my broker. (Bratman 1992, p.9)

So, according to Bratman, supposition does not lead directly to action, but acceptance can directly shape action. While it would be preferable to have a more precise account of how a cognitive attitude leads directly or indirectly to action, it is relatively clear what Bratman has in mind here. Supposition can be used in hypothetical reasoning about what one *would* do were one to have good reason to act on the supposition. In contrast, reasoning on the basis of what one accepts is not hypothetical in this way.

If one accepts that *p* in a context, one takes oneself to have good reason *ceteris paribus* to act on *p* in that context. The reasons one has for accepting that *p* can, as Bratman points out, be non-evidential, but that does not mean that evidence has no place to play in acceptance. There are some epistemic constraints on acceptance. One might wonder why there should be epistemic constraints on the premises we can use in our practical reasoning. It seems that we should be able to use any premise as part of our planning process. However, we need to remember that practical reasoning is reasoning about what *to do*. As such, it is distinct from daydreaming or hypothetical reasoning about what one would do in another faraway possible world. In our practical reasoning we make plans about what to do in this world, and, therefore, our planning must be guided by premises that have at least some relevant semblance to the actual world. Otherwise our plans become daydreams or hypotheses about very distant possible worlds. Thus, massive amounts of evidence that *p* is not true (or not even approximately true) can make it unreasonable to accept *p* in virtually all contexts. The abundant evidence that I do not have a million dollars makes it unreasonable to accept that I do and call my broker. A reasonable person could very well make hypothetical plans based on the supposition that she is a millionaire, but only an irrational person would plan her life around that attitude in the face of glaring evidence to the contrary. A reasonable person would find herself incapable of accepting that proposition. Were we to ask her to try to plan her life around it, she might tell us that she just cannot, no matter how hard she tries. I submit that the same can be true for trust.

In the vast majority of contexts, given a substantial amount of evidence against the trustworthiness of a person, we cannot trust her no matter how hard we try. Suppose I have seen my friend to be a terrible keeper of secrets. If I have an important secret, I may find that no matter how much I want to show confidence in my friend, I cannot accept that she will keep my secret. I cannot choose to trust her. Trust is, therefore, inconsistent with holding that one has overwhelming evidence that the trusted will not come through for one. This sketch of the epistemic constraints on acceptance nicely accounts for this intuition; just as one cannot accept that *p* if one takes oneself to have massive evidence that *p* is not even approximately true, one cannot trust that *B* will do *C* if one takes oneself to have massive evidence that *B* will not do *C*. I cannot trust my friend to keep my secret if I take myself to have good evidence to believe that she will divulge it. However, this epistemic constraint on trust only rules out trusting someone to do something one has overwhelming evidence to believe they will not do; it does not rule out trusting someone when one is agnostic about the likelihood that they will act as expected. Holton makes this same point about reliance: “I do not need to have the belief that you will do what I rely [on] you to do, but I do need to lack the belief that you will fail” (Holton 1994, p.71). In summary, unlike belief, acceptance can be voluntary, but that does not mean that we are always free to choose to accept whatever we wish. Since acceptance, unlike supposition, is subject to some epistemic constraints, we find that we are not entirely free when it comes to trusting. Even though this account of trust as sometimes involving acceptance can, as I have shown, accommodate intuitions which initially seem to support the view that trust requires belief, there are other objections to this account, which now need to be tackled.

One objection, proposed by Pamela Hieronymi, argues that trust without belief is merely a poor cousin of the type of trust of which we should want to give an account. She thinks that trusting without believing that the trusted will come through for you shows a lack of confidence in the trusted: “your lack of confidence betrays a lack of trust” (Hieronymi 2008, p.6). Hieronymi uses the following example to support this claim:

> Suppose that, in the morning, you and I agree to meet for dinner at a certain time at a certain restaurant to plan an upcoming event. Later in the day you learn that all my friends have decided to go to my favourite restaurant to celebrate a surprise promotion bestowed on one of them. You now doubt whether I will keep my engagement with you. You are not certain I will not, but then you are not certain I will either. You are in a state of doubt. In the face of your doubt, you decide to go to the restaurant and wait for me. (Hieronymi 2008, p.6)

Hieronymi imagines that when I arrive at the restaurant, you tell me about your doubts that I would show up and explain that you decided to go to the restaurant in spite of the doubts. She thinks that in this scenario, I will be concerned that your doubts express a lack of trust in me that I will keep my agreement (Hieronymi 2008, p.6). The idea here is that I will be concerned that you did not believe that I would keep our dinner date. I will think that this means that your trust in me is less than could be desired. Hieronymi calls trust accompanied by belief "full-fledged" trust. Full-fledged trust is a sort of ideal trust: "…even if one thinks the full-fledged sort of trust would be positively inappropriate in the circumstances, one can still imagine what it would be to have it, and its inappropriateness is typically explained by features of the situation seen as regrettable" (Hieronymi 2008, pp.6-7). On this objection, insofar as my account of trust allows room for trust based on acceptance rather than belief, it is an account of a less trusting sort of trust. What we really want is an account of the full-fledged, fully trusting, sort of trust that requires belief.

I not only fail to share Hieronymi's intuitions about the restaurant example, but I also fail to see why an account of trust need be an account of ideal or perfect trust. First, it seems to me that when you tell me about your doubts that I would arrive for dinner it would be just as natural for me to respond, "I'm sorry I gave you reason to doubt me, but thanks for trusting me anyway." I do not think that we always take doubt and the absence of complete confidence to suggest a lack of trust. Second, even if we agree with Hieronymi that trust with belief is a more trusting sort of trust, I do not see why an account of trust need only focus on this narrow class of trust phenomena. Hieronymi acknowledges this response when she notes that some may dismiss her account of full-fledged trust as a mere 'purist's' notion of trust. She argues that given the problems with the alternative accounts of trust, we ought to adopt the purist's notion of trust as "a natural refinement of our ordinary notion" (Hieronymi 2008, p.2). Therefore, in order to fully respond to her objection, I must show that the problems she sees with the alternatives do not arise with my account.

One such objection to the view that trust does not require belief rests on the sensible demand that trust be distinguished from mere pretence of trust. Many authors worry that trusting without belief is too similar to acting as if one trusts without actually trusting (Baker 1987, Hardin 2002, Hieronymi 2008, Holton 1994). This worry is thought to be particularly pressing when it comes to trusting what one is told. Suppose that my friend tells me that she is innocent of the crime of which she has been accused. It might seem right to say that what my friend wants of me is to believe that she is innocent, and she might charge me with failing to trust her if I do not believe in her innocence. Judith Baker uses this scenario to make the following argument that trust must require belief:

> Someone might try to distinguish trust from genuine or full belief. Trust, on such a view, would be a watered down variant of belief, something more like pretence or acting-as-if something were true. But this is to view trust as a non-serious form of belief. Whereas what one demands from one's friends is belief, not pretence, that one is innocent. And what some outsiders find amazing is just the fact that serious belief continues in the face of rising evidence against it. As a strategy then, this response amounts to an arbitrary denial of the phenomena which raise the problem in the first place. (Baker 1987, p.6)

Richard Holton, who agrees with my view that trust does not require belief, takes this objection to be problematic for his view. He says, "It is surely right that when we trust a friend, we do not simply act as if we believe what they say; we really believe them" (Holton 1994, p.73).

Now I agree that when we trust a friend we do not merely act as if we believe her, but I do not think this means that trusting others when they speak to us necessarily involves belief in what they assert. Bratman's account of acceptance allows us to identify a cognitive attitude that is neither belief nor pretence. Bratman focuses on the link to action to draw the acceptance/pretence distinction:

> Is such context-relative acceptance mere pretence? I do not think it is. In accepting that *p* I do not simply behave as if I think that *p*: I also reason on the assumption that *p*. So there is not the kind of indirect, circuitous connection between reasoning and action that is characteristic of pretence. (Bratman 1992, p.9)

Suppose that I say I trust my friend when she tells me that she is innocent, but I am unwilling to say that I believe she is innocent. On my account of trust, it is unfair to charge me with merely pretending to trust my friend. If I accept what she says in this context, I decide to work her innocence into my plans. Were I to be merely pretending to believe that she is innocent, then I would make plans with the goal of leading her (and perhaps others) to think that I believe she is innocent. My practical reasoning about how to keep up this pretence would be more convoluted and indirect than would be my reasoning based on the acceptance that she is innocent. Consider an example Bratman attributes to Michael Dummett:

> My close friend has been accused of a terrible crime, the evidence of his guilt is strong, but my friend insists on his innocence. Despite the evidence of guilt, my close friendship may argue for assuming, in my ordinary practical reasoning and action, that he is innocent of the charge. In making plans for a dinner party, for example, such considerations of loyalty might make it reasonable for me to take his innocence for granted and so not use this issue to preclude inviting him. Yet if I find myself on the jury I may well think that I should not take his innocence for granted in that context for reasons of friendship. (Bratman 1992, p.8)

The reasoning involved in the decision to invite the friend to dinner is relatively simple. By accepting the friend’s innocence, one decides to base one’s reasoning on the premise that he is innocent, and thus, as Bratman says, there is no reason not to invite the friend as one normally would. In contrast, the reasoning involved in merely pretending to trust the friend’s profession of innocence must involve considerations about how to keep up the appearance of trust. Therefore, instead of just deciding to do as one normally would and inviting the friend, one decides to act as usual because otherwise the friend might detect the pretence. This example illustrates how acceptance of testimony need not be merely pretence.[27] In this way, the worry that my account of trust is incapable of marking the distinction between trusting and pretending to trust is alleviated.[28]

It is on this question of the distinction between trusting and acting as if one trusts, that my account differs most strongly from its closest relative—Holton's account of trust as reliance from the participant stance. As I have noted, Holton agrees with my view that trust does not require belief. He also agrees that we can sometimes choose to trust. His account of trust also has two parts which closely mirror my own. The first part of his account is, like mine, meant to capture what trust and reliance have in common.[29] When I trust or rely on someone to do something, "I plan on the supposition that they will do it" (Holton 1994, p.72).[30] This is very similar to saying that I make the assumption that they will do it part of my adjusted cognitive background.

Despite these similarities, Holton does not clearly mark the distinction between trusting and acting as if one trusts. Holton characterizes trust as "a kind of acting-as-if" (Holton 1994, p.73). He gives the following example of trust circle games to illustrate his account of trust. In these games, the participant stands in a circle of people who are supposed to catch her. The participant closes her eyes and lets herself fall backwards, and the people in the circle catch her. Holton says that at the moment before one falls, one can choose to fall despite having some

[27] Bratman's example also provides a nice counterexample to the assumption shared by Holton and Hieronymi that trusting the testimony of others is one type of trust that requires belief in what is said (Hieronymi 2008, p. 8; Holton 1994, p. 73). I find it plausible to say that one does trust what the friend says about her innocence even though one does not thereby acquire a context-independent belief in her innocence.

[28] Holton also provides a nice example that illustrates the difference between pretending to believe that *p* and working *p* into one's plans: "When I feign belief in God, I do not work my plans around the supposition of God's existence: I do not, for instance, plan with an eye to the Day of Judgement. All that I work into my plans is my pretence itself, not the truth of that which I am pretending" (Holton 1994, p.72).

[29] I will discuss the second part of his account, which depends on the notion of the participant stance, in section 2.

[30] Holton occasionally also describes the truster as working it into her plans that someone will do something (Holton 1994, p.72).

doubts about whether one will be caught. This is his example of choosing to trust. To explain what is going on when one makes that choice, Holton draws the following analogy: “Just as the non-believer in the [religiously] strict society can decide to act as a believer would, so I can decide to act on the supposition that you will catch me. That is to decide to rely on you” (Holton 1994, p.69). This analogy reveals that Holton’s account does not adequately distinguish between trusting and acting as if one trusts. Deciding to trust the people in the trust circle is here being compared to acting as if one believes the religious doctrines of a strict religious society. Thus, Holton maintains that trusting is a kind of acting-as-if.

The problem with Holton’s account is that trusting and acting as if one trusts are clearly different. One can act as if one takes *p* as a premise in one’s practical reasoning without actually doing so. Any act of pretense will fit this description. Both parts of Holton’s analogy illustrate this. The non-believer who acts as a believer may do so simply to avoid shunning without taking any of the believer’s religious doctrines as premises in her practical reasoning. Similarly, I can decide to fall into the arms of the catchers in the trust circle without in any way taking the premise that I will be caught as a premise in my practical reasoning. I may fall, thus acting-as-if I trust, without believing or accepting that I will be caught. For example, I may reason that I ought to fall backwards because I think that I will not be caught and that this will make the others feel guilty and treat me better in the future. The problem with Holton’s account of trust is not that he says trusting involves acting on the supposition that the trusted will act as expected.[31] The problem is that Holton maintains that acting on such a supposition is a case of acting-as-if. My account does not confuse these two notions.

[31] Although for the sake of clarity, I would replace his use of ‘supposition’ with ‘assumption’ to keep in line with Bratman’s distinction between supposition and acceptance.

Having defended this part of my account of trust from various objections, I should explain how it meets the first of the criteria of adequacy. The first criterion demanded that an account of trust be able to show how trust, like reliance, involves vulnerability and risk. Part (1) of **(T)** is able to do this because when one takes the premise that someone will do something as part of one's adjusted cognitive background, one works it into one's plans that she will do it—one uses the assumption that she will do it in one's practical reasoning. Therefore, when one counts on someone in this way (by either trusting or relying), one is vulnerable to having one's plans undermined. If she fails to act as one expects, then the success of one's practical reasoning is threatened. One has made plans and reasoned about what to do based on a false premise. This is the risk that one takes when one takes the premise that someone will do something as part of one's adjusted cognitive background. This risk is present both when one relies upon and when one trusts someone. Therefore, this first part of **(T)** [*A* trusts *B* to do *C iff (1) A* has the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background] explains what reliance and trust have in common. Having explained what it means for an agent to take the premise that someone will do something as part of her adjusted cognitive background, I will now turn to the second part of my account of trust.

2.2 REASONS FOR TRUST AND VULNERABILITY TO BETRAYAL

One of Baier's insights into the usefulness of the ethical notion of trust is that it contains both cognitive and emotional aspects (Baier 1994, p.10).[32] The first part of **(T)** covers the cognitive aspect of trust by defining trust in terms of one's adjusted cognitive background. The second part of **(T)** captures the emotional element by providing a framework through which to understand the connection between trust and betrayal, which is the emotion that one usually feels when one is let down by someone one trusts. Before I explain this second part of **(T)**, I should first make explicit some goals I hope to achieve with this second part of my account of trust.

First, it needs to mark the distinction between trust and reliance. The notion of taking the proposition that someone will do something as part of one's adjusted cognitive background captures what trusting and relying have in common. As explained in section 1, when one trusts or relies on someone to do something, one works it into one's plans that the trusted/relied upon will do that thing. However, trusting is, on my account, importantly different from relying. By explaining *A*'s trust in *B* to do *C* in terms of particular reasons why *A* takes this cognitive attitude, I provide an account of trust which, unlike the rational choice accounts, does not reduce trust to mere reliance. Some reasons for taking the proposition that someone will do something as part of one's adjusted cognitive background make one susceptible to feeling betrayed instead of merely disappointed when one's expectations are frustrated. In particular, taking the cognitive attitude described in (1) *because* one believes that one has a relationship with someone which obliges that person to do something, makes one susceptible to feeling betrayed. Thus, a

[32] Baier says, "[T]o trust is neither quite to believe something about the trusted nor necessarily to feel any emotion toward them—but to have a belief-informed and action-influencing attitude" (1994, p.10). This is a nice characterization of the account of trust I present.

discussion of the nature of betrayal and its connection to beliefs about relationships and reasons for adopting cognitive attitudes will reveal how this second part of **(T)** marks the distinction between trust and mere reliance.

Second, this part of **(T)** needs to mark the trust/reliance distinction without falling prey to some counterexamples that plague other accounts of trust. Holton's critique of Baier's account of trust provides a useful illustration of the type of pitfalls I hope to avoid. Baier's account of trust is similar to mine in that she thinks trusting is a particular way of relying on someone. She thinks trusting is relying on the trusted's goodwill towards one (Baier 1994, p.99). This account of trust does mark a distinction between trust and reliance. For example, it can explain why, to use Baier's example, it does not make sense to say that Kant's neighbors trusted him when they used his regular walks as a time keeping system. They did rely upon him, but they were not relying on his goodwill. Unfortunately, as Holton points out, Baier's account still allows for some cases of mere reliance to count as cases of trust. The confidence trickster is one counterexample that proves this point. A con artist may befriend her victim and then rely on the goodwill of the victim as part of her plan to steal the money. But clearly the con artist does not trust the victim to give her the money. Thus, Baier's view that trust requires reliance on the goodwill of the trusted is insufficient to mark the distinction between trust and reliance where we intuitively want it to be. I provide an account of trust that distinguishes between cases of mere reliance, including the reliance of a confidence trickster, and instances of trust. Having made clear what is to be achieved by this part of **(T)**, I can now turn to the details of **(T)**.

Recall that my account of trust is as follows:

(T) *A* trusts *B* to do *C iff (1) A* takes the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background, (2) in part, because *A* believes that their relationship morally obliges *B* to do *C*.

The first thing to notice about (2) is that it makes trusting a matter of taking a particular cognitive attitude *partly on the basis* of a particular type of reason. The nature of one's reasons for taking the cognitive attitude described in (1) determine whether one trusts or merely relies upon someone when one works it into one's plans that she will do something. To explain why this is the case, I must take a detour to discuss the nature of betrayal.

2.2.1 An account of betrayal

Suppose that I am counting on my friend to keep a secret that I have shared with her. Now if I come across some evidence that makes me think that she has told my secret to others, then I will feel betrayed because she did not live up to the obligations that I expect of my friends. That feeling of betrayal is the typical reaction to the violation of trust. Following Baier, many authors correctly identify vulnerability to betrayal as the essential difference between trust and reliance (e.g. Baier 1994, Hieronymi 2008, Holton 1994). Since betrayal is so central to understanding trust, it is necessary to have an account of this emotion. In this section, I provide an account of betrayal as a reactive emotion that connects it to beliefs about moral relationships and the obligations involved in such relationships.

The reactive attitudes (or 'reactive emotions') were outlined by Strawson in "Freedom and Resentment" (Strawson 1962). While my conception of the reactive attitudes differs from Strawson's, we agree on the following key observations. The emotions of resentment, indignation and guilt are reactive attitudes. They are "non-detached attitudes and reactions of people directly involved in transactions with each other" (Strawson 1962, p.4). The reactive attitudes are antithetical to the *objective attitude* adopted by the scientist towards her objects of research, the psychiatrist towards her patient or the public policy maker towards the public. The objective attitude is the stance of treating the other person as a subject of treatment, “something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided” (Strawson 1962, p.9). The therapist may have some emotions towards her patient, but a therapist who feels resentful towards a patient who fails to respond to treatment has formed an inappropriate relationship with her patient. The reactive emotions are those emotions to which humans are naturally subject as a result of their inevitable involvement, or participation, in interpersonal relationships (Wallace 1994, pp.31-2). Thus, normally we adopt a *participant attitude* (or, as Holton calls it, ‘participant stance’) towards people with whom we have relationships. It is only when we are engaged in certain kinds of unusual relationships (e.g. patient/psychiatrist) or when we see someone as “warped or deranged or compulsive in behaviour or peculiarly unfortunate in his formative circumstances” that we tend to adopt the objective attitude towards that person (Strawson 1962, p.9). In adopting the objective attitude towards someone, we “set him apart from normal participant reactive attitudes” (Strawson, p.9). Strawson also seems to think that when we adopt the objective attitude towards someone, we fail to view her as someone with whom one has a moral relationship. Speaking of someone whom we deem insane, Strawson says,

> We may say: to the extent to which the agent is seen [as insane], he is not seen as one on whom demands and expectations lie in that particular way in which we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community. (Strawson 1962, p.17)

Thus, sometimes we adopt an attitude towards people that denies that we have a moral relationship with them. When we do so, we are not subject to certain normal human emotions towards such people. However, normally we view people as members of the moral community, as individuals with whom we have moral relationships, and when we do so, we tend to experience the reactive emotions in the course of our interactions with them.

There are three basic reactive emotions: resentment, indignation and guilt.[33] These different emotions are distinguished by the types of explanations we give when we experience them. As Rawls says, "In general, it is a necessary feature of moral feelings, and part of what distinguishes them from the natural attitudes, that the person's explanation of his experience invokes a moral concept and its associated principles" (Rawls 1971/1999, §73). The differences between resentment, indignation and guilt are marked by the concepts and principles which we invoke in our explanations. In other words, it is an explanation's invocation of the belief that a certain kind of obligation has been violated which specifies which reactive emotion is felt. When I explain my emotion by appealing to my belief that someone has wronged me, I am feeling resentment. When my explanation invokes my belief that another person has wronged someone else, I am feeling indignation. And when my emotion is explained in terms of my belief I have wronged someone, then the emotion I am explaining is guilt.

One can be susceptible to both moral and nonmoral versions of each of these reactive emotions (Wallace 1994, p.34). The moral reactive attitudes are responses to perceived violations of moral obligations. Nonmoral reactive attitudes are responses to perceived violations of nonmoral obligations. Thus, one can feel moral resentment when one believes that someone has breached a moral code, for example, by lying. This resentment is moral resentment because the expectation that the person tell the truth is justifiable in terms of moral obligations.

[33] Strawson's original account of the reactive attitudes is much more inclusive. His list includes "such things as gratitude, resentment, forgiveness, love, and hurt feelings" (Strawson 1962, p.4). Wallace argues that an inclusive account of the reactive attitudes undermines Strawson's key claims that 1) the reactive attitudes are practically inevitable for us and 2) they are opposed to the objective stance (Wallace 1994, pp 29-32). I will not rehearse the details of these arguments, but I do find them persuasive and use Wallace's shorter list of the reactive attitudes in what follows.

In contrast, one could feel nonmoral resentment about another's breach of etiquette (Wallace 1994, pp.36-7). If the code of etiquette is not justifiable on moral grounds, then one feels nonmoral resentment when it is breached.[34] In sum, the key to classifying the type of reactive attitude a person is feeling is to identify the type of belief that explains why she is feeling the emotion in question. To determine if it is resentment, indignation or guilt, one needs to know whether she believes herself or another to have violated an obligation towards her or others. To determine whether she feels a moral or nonmoral reactive emotion, one should know whether she believes a moral or nonmoral obligation was violated.

As Holton has observed, betrayal is a reactive attitude (Holton 1994, p.66). My account of betrayal is as follows: betrayal is a feeling of resentment explained by invoking the belief that someone has violated a moral, relational obligation to perform particular behavior towards oneself that one was counting on her to perform on the basis of one's belief in the existence of the relational obligation to do so.[35] The elements of this account of betrayal are as follows: 1) betrayal is one form of the reactive emotion of *resentment*, 2) it is a *moral* reactive emotion, 3) the explanation of the emotion invokes the belief that someone failed to perform some action that one had *believed or accepted* that she would perform, and 4) one believed or accepted that the trusted would perform this action partly *because* one believed that one had a *relationship* with the trusted that *obliged* her to perform that action. I will argue for each of these elements in turn.

First, betrayal is a species of *resentment* because it is an emotional response to a perceived wrong to oneself. One might object here that just as I will feel betrayed if my friend

[34] However, one might argue, following Sarah Buss, that matters of etiquette are matters of morality (Buss 1999).

[35] To simplify terminology, I describe betrayal in terms of a reaction to someone failing to perform an action, but we can also, and frequently do, feel betrayed when someone fails to adopt an attitude towards us that we think they are obliged to take.

reveals my secret that I expected her to keep to herself, I can feel betrayed if my mother's friend divulges my mother's secret. However, cases of feelings of betrayal in response to someone else being wronged are best explained either as instances of sympathetic feelings of betrayal when one joins another in feeling the betrayal they feel, or as instances of feeling betrayed because of a wrong committed toward oneself, such as the wrong of acting as an untrustworthy family friend. Thus to feel betrayed is to feel a type of the reactive emotion of resentment.

Second, betrayal is a *moral* reactive attitude because moral justifications are given for the obligations that we believe have been violated when we feel betrayed. The particular kind of obligations that, when violated, make one susceptible to feeling betrayed are *relational obligations*. Relational obligations are moral obligations to which one is subject when one participates in a relationship.[36] For example, sometimes we explicitly agree to take on a certain type of relationship by agreeing to undertake certain obligations (e.g. wedding vows). Other times we explicitly agree to do something in order to reassure someone that we are indeed in a particular kind of relationship (e.g. agreeing to keep a secret as a sign of friendship). This explicit agreement creates a relational obligation. However, explicit agreement is not the only

[36] When one has a relational obligation to do something, one has a duty *to the other party* in that relationship to do that thing. This is important to recognize because sometimes our relationships involve obligations to do things to or for third parties. As H.L.A. Hart notes, it can be easy to get confused about the nature of the obligations and corresponding rights in these situations (Hart 1955). Consider the relationship between an at-home nurse, X, and employer, Y, who has hired X to care for his ailing mother. X has a relational obligation to care for Y's mother. Now suppose that X is derelict in her duty. X has not lived up to an obligation to someone, but to whom? Hart explains, "Rights arise out of this transaction [between X and Y], but it is surely Y to whom the promise has been made and not his mother who has or possesses these rights. Certainly Y's mother is a person concerning whom X has an obligation and a person who will benefit by its performance, but the person to whom he has an obligation to look after her is Y. This is something due to or owed to Y, so it is Y, not his mother, whose right X will disregard and to whom X will have done wrong if he fails to keep his promise, though the mother may be physically injured. And it is Y who has a moral claim upon X, is entitled to have his mother looked after, and who can waive the claim and release Y from the obligation" (Hart 1955, p.180). Thus, when one has a relational obligation to do something, one has an obligation to the person with whom one has that relationship. It is this person who is liable to feel betrayed if one does not live up to the obligation. Third parties who failed to benefit from one's expected living up to that obligation are not individuals to whom this duty has been breached. Thus, they are not liable to feeling betrayed.

way that one can become subject to a relational obligation to perform some action. Moral justifications for obligations can also derive from the trusted's implicit agreement to behave in a particular way given their participation in a certain kind of relationship with the truster. One might say "My therapist ought to maintain confidentiality because it's an essential part of the therapist-patient relationship" or "That's what friends do—they keep each other's secrets." By participating in a therapist-patient relationship, the therapist implicitly agrees to maintain confidentiality, and is thus subject to a relational obligation to do so. We normally feel betrayed when someone breaches a relational obligation. However, we do not feel betrayed when someone breaches a code of etiquette, even if that code of etiquette applies to a particular type of relationship. Feelings of betrayal are normally associated with a sense that the betrayer failed to do something she ought, morally speaking, to have done. Thus betrayal is a moral form of resentment because it is a response to a perceived violation of a moral obligation.

Third, betrayal is felt only if one believed or accepted that the trusted would live up to her relational obligation. One might feel some other type of resentment when someone has failed to live up to a moral obligation of behavior towards oneself, but one will not feel betrayal unless one adopted the assumption that the trusted would live up to the obligation as part of one's adjusted cognitive background. Suppose I have two roommates, Pam and Angela. Pam regularly eats my food, while Angela never does. If I buy some of my favorite cereal, I would likely feel resentful when Pam eats it, but I would not feel betrayed. In contrast, I would feel betrayed if I bought the cereal while Pam was out of town and Angela ate it. I would feel betrayed because I counted on Angela not to eat my food. I assumed that Angela would not eat my food, but experience had long ago taught me not to assume that Pam would not eat my food. One does not feel betrayed when one neither believes nor accepts that someone will live up to her relational obligations to one. Betrayal, then, is explained in terms of a certain kind of perceived wrong, the wrong of having one's practical reasoning undermined.[37]

Finally, in order to feel betrayed rather than just wronged by someone, one must also have counted on her *because* one believed her to be under a moral obligation to perform the action in question. This move is intended to rule out feelings of betrayal as a response to someone failing to be merely reliable. Suppose I have hired a house sitter to take care of my home while I am on vacation. I am not sure that the house sitter will not steal my belongings, so I put a noticeable surveillance system in place to make it in her interest not to steal. Now if the house sitter steals from me despite these precautions, I might feel disappointed, but I would not feel betrayed. However, suppose that instead of trying to make it in her interest not to steal, I had believed her to be a morally upstanding person who would take her duties as my house sitter seriously. In this situation, I would feel betrayed, and not merely disappointed, when I discover her theft. My feeling in these two scenarios thus fits my account of betrayal as a type of moral resentment explained in terms of a belief that someone failed to live up to an obligation that I was counting on them to live up to *because* I believed that our relationship created a moral obligation to do so. I feel resentful because someone wronged me, and it is moral resentment because the house sitter had a moral obligation not to steal from her employer. I was counting on the house sitter in a way that I would not be counting on a random passerby not to jump in an open window and steal my belongings. And my reason for counting on her was not that I

[37] In this, I disagree with Hieronymi, who thinks that "one can be betrayed even in cases in which one does not trust at all (by, say, a dastardly politician) and that one can risk betrayal without trusting" (Hieronymi 2008, p.17). I classify cases like being let down by a dastardly politician as cases of being wronged, but not cases of being betrayed. If one does not trustingly take the proposition that a dastardly politician will do something as part of one's adjusted cognitive background, then one cannot be betrayed by when the dastardly politician fails to do it; instead one will have merely been wronged.

thought I had been clever enough to put effective deterrents in place. My reason for counting on her was that I believed her to be motivated to live up to the obligation not to steal.

This account also helps us explain our intuitions about the scientists, Alice, Betty and Claire, from chapter one. Recall that Alice works it into her plans that Betty, her colleague from another discipline, will perform her part of their joint project carefully. Alice's reason for doing so is that she thinks she has made it in Betty's self-interest to work carefully because Alice has asked Claire to check Betty's work. So Alice has also worked it into her plans that Claire will carefully check Betty's work. However, Alice does so for very different reasons than those which prompt her to count on Betty. Alice counts on Claire because she thinks that Claire is her friend and will carefully check Betty's work because her friend asked her to do so. I argued that, intuitively, we want to say that Alice will feel betrayed when Claire lets her down; however, when Betty lets her down, Alice will not feel betrayed, but instead merely disappointed. We can now explain why we have this intuition. One feels betrayed in response to having one's plans frustrated by someone who fails to do something one assumed she would do only when one's reasons for making those plans include the belief that one was in a relationship with that person that obliged her to perform that action. Alice believes that she has a friendship with Claire that obliges her to keep her promise to Alice, and that is part of the reason why Alice assumes Claire will do so. In contrast, Alice's reasons for basing her plans on the assumption that Betty would do her work carefully are completely different. They are not reasons that take into account a moral relationship and relational obligations. Thus, in adopting the assumption that Betty would do her work carefully, Alice did not take the participant stance towards Betty; she did not make herself susceptible to feeling betrayed. These examples illustrate how betrayal is a response to being let down by someone towards whom one adopted the participant stance. Therefore, identifying the place betrayal occupies in the reactive attitudes can provide a detailed account of the emotional aspect of trust.

2.2.2 *A* adopts the cognitive attitude of trust, in part, because *A* believes that their relationship obliges *B* to do *C*.

Having provided an account of betrayal as a reactive attitude, I now use this account to explain part (2) of my account of trust. Recall that my account is as follows:

(T) *A* trusts *B* to do *C iff (1) A* takes the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background, *(2)* in part, because *A* believes that their relationship morally obliges *B* to do *C*.

Trust is not simply a matter of taking a particular cognitive attitude towards the proposition that someone will do something. In order for one to trust someone to do something, one must adopt that cognitive attitude in a way that makes one susceptible to the reactive attitude of betrayal. As I have shown, betrayal is a feeling of resentment explained by invoking the belief that someone has violated a relational obligation to perform particular behavior towards oneself that one was counting on her to perform on the basis of one's belief in the existence of the relational obligation to do so. Thus, in order to adopt the relevant cognitive attitude in a way that makes one susceptible to betrayal, one's reasons for adopting that cognitive attitude must include belief that one's relationship with that person obliges her to do the thing one trusts her to do. In other words, in order for *A* to count as trusting *B* to do *C*, it is not enough for *A* to take the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background. If *A*'s reasons for doing so do not include the belief that their relationship obliges *B* to do *C*, then *A* is not susceptible to feeling betrayed. *A* would then be like Alice who does not trust Betty or like the employer with the surveillance equipment who does not trust her house sitter. So if *A* takes the proposition that *B* will do *C* as part of *A*'s adjusted cognitive background, in part, because *A* believes that their relationship obliges *B* to do *C*, *A* meets a necessary condition for trusting *B* to do *C*.

Baier's example of Kant's daily walks helps to show that (1) and (2) provide jointly sufficient conditions for trust. First consider the example as Baier originally presents it. Kant takes regular daily walks, and unbeknownst to him, his neighbors use his walks as a reliable indicator of time. They begin to make plans for their schedules on the assumption that he will take his walk at the same time every day. This assumption has been adopted as part of their cognitive background. Kant does not know that they make plans based on his walks, and they do not believe that he has any obligation to them to be punctual. They have no relationship with him. He is just a man with peculiarly punctual habits who happens to live in their neighborhood. Given this description of the situation, we do not want to say that they trust him to walk at the same time every day. They are just using him like they would a clock, and we tend not to think of people as trusting their clocks. Now suppose we change the scenario simply by creating a relationship between Kant and his neighbors. As a sign of friendship, Kant agrees to keep his walks punctual so that his neighbors can tell the time. In this scenario, his neighbors make plans on the basis of Kant's assumed punctuality because they believe that, as a friend, Kant has an obligation to keep his promise to walk on time. When we change the details of the situation in this way, our intuitions change, and we are now likely to say that the neighbors trust Kant to walk punctually. Thus, adopting the premise that someone will do something as part of one's cognitive background on the basis of one's belief that there is a relationship that obliges that person to do that thing, is a sufficient condition for trusting that person to do that thing.

Before addressing some objections to the second part of **(T)**, I will show that it meets the two criteria for adequacy that were laid out at the beginning of section 2.2. First, this account marks the distinction between trust and reliance. As I showed in section one, trust and reliance are similar in that they both involve vulnerability on the part of the truster. What distinguishes them is that trust involves vulnerability to feeling betrayed, and reliance does not. I have shown that meeting conditions (1) and (2) of **(T)** makes one vulnerable to feeling betrayed. Any cognitive attitude that meets (1) but does not meet (2) will be an instance of reliance rather than trust.

From this it follows that the rational choice approach provides an account of reliance instead of trust. In chapter one, I outlined the rational choice account of trust as involving the following two assumptions about the reasons for trusting:

(RC1): *A* trusts *B* to do *C* on the basis of evidence that *B* will do *C*.

(RC2): The evidence *A* relies upon is evidence that it is in *B*'s self-interest to do *C*.

According to the rational choice account of trust, *A* trusts on the basis of evidence that it is in *B*'s self-interest to do *C*. On my view, the rational choice approach provides an account of reliance rather than trust because the reasons cited for believing that *B* will do *C* do not include *A*'s belief that *A* and *B* are in a relationship that obliges *B* to do *C*. The rational choice account eschews reference to moral obligations. In doing so it generates an account of reliance rather than trust because susceptibility to betrayal, the hallmark of trust, comes from working it into one's plans that someone will do something *because* one believes that there is a relationship that obliges that person to act that way. The graduate student/advisor relationship illustrates this. Whether a graduate student trusts or merely relies on her advisor to give her appropriate credit for her work depends on the reasons why she works it into her plans that the advisor will do so. If the graduate student makes plans on this assumption because she believes that it is in her advisor's self-interest to give her appropriate credit (say because she thinks she has enough power to damage the advisor’s reputation for plagiarism), then the student merely relies on the advisor. However, if the student makes plans on the assumption that the advisor will not steal her work partially because she believes that the student/advisor relationship carries with it a moral obligation for advisors to give their students appropriate credit, then the student trusts her advisor. When the student adopts this cognitive attitude towards her advisor’s future actions on the basis of this kind of reason, she makes herself susceptible to feeling betrayed when the advisor plagiarizes her work. This example shows that one can use my account of trust to distinguish between cases of reliance and cases of trust.

The second criterion of adequacy for (2), outlined at the beginning of section two, is that it not make **(T)** susceptible to the confidence artist counterexample that plagues Baier’s account. One problem with Baier’s account of trust as reliance upon the goodwill of the trusted is that it implies that a confidence artist trusts her victim to pay her money. My account does not have this counterintuitive result. A confidence artist’s assumption that her victim will pay her the money may meet condition (1) of **(T)**, but it will not meet (2). The confidence artist’s reasons for making plans on the basis of the assumption that her victim will give her money do not include the belief that she has a relationship with the victim that obliges the victim to give her the money. The confidence artist assumes that the victim will pay because the con artist believes that her confidence tricks will be successful. She believes that she will be able to manipulate the victim into doing something that the victim has no obligation to do. She relies on the victim to be easily manipulated. It would be a strange confidence artist who believed that the victim had an obligation to be victimized. Thus, this example does not cause problems for my account.

2.2.3 Objections and replies

There are two types of objection that one might press against this second part of **(T)**. First, one might want to object to my account of betrayal. Second, one might want to raise various counterexamples to **(T)**. I present and reply to some instances of both types of objection.

One might object that my account of betrayal provides no room for irrational feelings of betrayal. R.J. Wallace has leveled a similar charge against Rawls and Butler's accounts of the moral sentiments, which share relevant similarities with the account of the reactive attitudes on which my account of betrayal depends. Wallace frames his discussion in terms of the irrational guilt that one feels when one does not believe that one has done anything wrong:

> Very often when one feels guilt inappropriately [Wallace uses 'inappropriate' where I use 'irrational'], what makes the guilt seem inappropriate is precisely the fact that one does not believe oneself to have done anything morally wrong at all; there are no moral obligations one accepts that one believes oneself to have violated. If this is correct, however, then genuine guilt cannot always be explained by distinctively moral beliefs, as Butler and Rawls propose. (Wallace 1994, pp.46-7)

Along these lines, one might argue that my account of betrayal does not allow for irrational feelings of betrayal, since on my account explanations of feelings of betrayal invoke the belief that the trusted has not lived up to her obligations as one assumed she would. Irrational feelings of betrayal might not appear to fit this account because one cannot explain one's feeling of betrayal in terms of a belief that the trusted has violated an obligation. Thus one might, following Wallace, object that this account of trust provides no room for irrational feelings of betrayal.

However, it is possible to explain how we can trust irrationally and experience irrational feelings of betrayal. Irrational betrayal might result from a conflict amongst one's beliefs. While it is outside the scope of this project to fully explore how conflicting beliefs could lead to irrational emotions, there are two ways one might begin to flesh out this idea. First, one might have a general moral belief that conflicts with one's beliefs about what a particular person ought to do in a particular situation. For example, the lapsed Catholic may believe that, in general, it is permissible to have sex outside of marriage; however, when it comes to herself, she cannot apply this principle to her actions because she believes that it is wrong for *her* to do so. Similarly, irrational feelings of betrayal could be explained by a conflict between one's belief that one has been betrayed by someone one was counting upon and one's other belief that, in general, it would not be wrong of someone to let one down in this way.

Wallace may object that this explanation sidesteps the real issue because it does not explain how one might feel guilt despite not believing oneself to have done something morally wrong. In other words, Wallace may point out that the lapsed Catholic would say that she *does* believe herself to have done something wrong by having premarital sex. In response, there may be another way to explain the irrational guilt (and betrayal) in terms of a conflict between beliefs. If it makes sense to talk of unconscious beliefs, then there could be a conflict between one's conscious and unconscious beliefs. For example, the lapsed Catholic's guilt could be explained by the unconscious belief that it is wrong for her to have sex outside of marriage. It is true that, if asked, she is likely to say that she does not believe that it is wrong to have sex outside of marriage; but nonetheless, at some level, she does still believe that it is wrong for her to do so. Again, the irrationality would be caused by the inconsistency of her beliefs. In the same way, one might have irrational feelings of betrayal were one to consciously believe that, in general, it would not be wrong of someone to let one down in this way and also unconsciously believe that so-and-so betrayed one when she acted in that way. Trust could, therefore, be irrational when one does not have consistent beliefs about the kinds of obligation to which the trusted is subject.

There are other objections to **(T)** that are based on proposed counterexamples. One might worry that this account of trust is too inclusive. While some authors (e.g. (Baier 1994) and (Holton 1994)) have attempted to provide a broad account of trust, one might question the need for an account of trust that covers phenomena as dissimilar as counting on one's spouse to remain faithful and counting on a stranger to give good directions to the nearest library. These are certainly very different types of interaction, and it would be a mark against an account of trust if it were unable to provide any conceptual machinery to differentiate between them. However, it is also true that there is something that connects trust in intimate relationships and trust in strangers. Therefore, I think the best approach is to provide an account that both shows what these types of interaction have in common and also explains their key differences. On my account, trust in one's spouse and trust in a stranger are similar in that both can lead to feelings of betrayal when the trust is disappointed. In most cases, our interactions with a stranger when we stop to ask for directions are brief, and when we find that they have led us astray, we feel only annoyance and would not deem it appropriate to feel betrayed. However, humans are capable of rapidly forming bonds with other people, and in some cases one might feel that one has formed enough of a relationship with the helpful stranger to feel betrayed if one is let down. In those cases where one feels betrayed by the stranger, my account will acknowledge you as having trusted the stranger. In cases where enough of a relationship has been established so that one becomes susceptible to feelings of betrayal, a similar phenomenon is at play as the trust between spouses. However, the account can explain the difference between stranger-trust and spouse-trust by pointing to the strength of the relationship between the truster and the trusted. In those rare cases when one forms a bond with a helpful stranger, one does develop a relationship with the stranger, but it is a very weak relationship, and as such it will normally only give rise to very mild feelings of betrayal. In contrast, the relationship between spouses is much more significant and carries with it the potential for strong feelings of betrayal. Someone who thinks that my account of trust is too inclusive most likely takes the central cases of trust to be trust between spouses or friends or other close relationships. While I agree that it is most natural to talk about trust between close relations, I do think some sense can be made of cases of very minimal trust such as that between strangers.

The worry that this account of trust is too inclusive can also be allayed by recognizing that the account also allows for inappropriate trust. Some of the cases that this account classifies as instances of trust will be cases of inappropriate trust. So many apparent counterexamples will turn out to be instances of inappropriate trust. Some of Baier's examples of trust in strangers are, I think, better described as instances of reliance. For example, she says "We trust those we encounter in lonely library stacks to be searching for books, not victims. We sometimes let ourselves fall asleep on train or planes, trusting neighboring strangers not to take advantage of our defenselessness" (Baier 1994, p.98). Suppose I am attacked in the library stacks or on a train. I think the normal emotional reaction would be to feel egregiously wronged but not betrayed. Betrayal seems out of place in this situation. We can accommodate this intuition by noting that it is part of normal psychological development that we learn which types of emotions are socially appropriate in given situations. Feelings of betrayal are usually inappropriate in our dealings with total strangers.[38] Thus, when someone does feel betrayed by a stranger, we are tempted to say that their trust in that person was inappropriate. My account of trust will count

[38] Notice that inappropriate trust is not the same as the irrational trust I described in response to Wallace's imagined objection. These are two different ways that trust can go wrong. In irrational trust, one's trust is based on a belief that contradicts other beliefs held in one's adjusted cognitive background. The problem with inappropriate trust is not one of holding or failing to hold particular beliefs. Inappropriate trust happens when our emotions misfire; the emotion of betrayal is triggered in the wrong situation.

any instance of feelings of betrayal as a response to violated trust, but this does not mean that every instance of trust will be a case of appropriate trust. Therefore, in many cases where one worries that my account is counting too many phenomena as cases of trust, the worry can be somewhat alleviated by noting that the account does at least allow one to call many such phenomena instances of inappropriate trust.

One might object that **(T)** cannot explain the natural phenomenon of a child's trust in her parents. According to this objection, my account requires the truster to have cognitive machinery and moral concepts that children do not have, since it requires the truster to have the belief that there is a relationship that obliges the trusted to perform some action. One might claim that children do not have sophisticated beliefs about relationships and obligations, but they nonetheless trust. Thus, my account inappropriately denies that children can trust. My response to this objection is to bite the bullet to some extent and concede that very young children do not trust. In fact, very young children cannot even rely on their parents. A very young child, who lacks the cognitive ability to weigh the evidence and practical considerations that bear on her cognitive attitudes towards the future actions of her parents, cannot adopt the cognitive attitude described in (1) of **(T)**. At some point, a child develops these abilities, and it seems possible that this happens at a stage before she develops moral concepts like obligation. If this is the case, then a child at this stage is able to rely on, but not trust her parents. It is not until a child is able to form beliefs about the kinds of roles that people play and the kinds of things that people playing the role of parent should do, that we can say the child trusts her parents. So to the extent that an infant lacks these cognitive abilities, that infant cannot be said to trust her parents. However, children may develop these cognitive abilities and moral concepts relatively early on, since one does not need to have a sophisticated theory of moral obligation to be able to form the belief that a parent ought to take care of her child. Therefore, biting the bullet on this objection does not require me to give up very much. Throughout most of their childhood, children do have the ability to trust. Having responded to several objections to this account, I finish by comparing my account of trust to the rational choice account criticized in chapter one.

2.3 COMPARISON TO THE RATIONAL CHOICE ACCOUNT

In chapter one, I argued that while the rational choice approach provides useful insights into trust, it also overlooks many other important aspects of trusting behavior. In particular, I argued that the rational choice approach is limited in four ways. First, it omits reasons for trusting based on evidence about the trusted other than evidence that it is in the trusted's self-interest to act trustworthily. Second, its explanation of the reasons for acting trustworthily only accounts for self-interested reasons for being trustworthy. Third, it does not leave room for non-evidential reasons for trusting. Fourth, it fails to mark the distinction between trust and mere reliance. I have argued that **(T)** is able to mark the trust/reliance distinction, so it is clear that my account does not share the fourth limitation. Thus, I will now argue that my account of trust does not share the first three limitations facing the rational choice account.

In contrast to the first and second limitations of the rational choice approach, **(T)** can account for different types of evidence for trusting because it recognizes non-self-interested reasons for being trustworthy. Recall that **(T)** allows for trust based on either belief or acceptance. Trust may involve belief that someone will do something, or it may involve acceptance that someone will do something. In cases when *A*'s trust that *B* will do *C* is grounded in belief that *B* will do *C*, *A*'s trust, if it is reasonable, is "shaped primarily by evidence for what is believed and concern for the truth of what is believed" (Bratman 1992, p.3). In these cases, *A* trusts on the basis of evidence that *B* will do *C*. But what kind of evidence grounds such trust? In order for *A*'s attitude towards *B* to be one of trust rather than reliance, part of *A*'s set of reasons for trusting must include a belief that their relationship obliges *B* to do *C*. Thus, evidence that *B* is motivated to live up to the obligation to do *C* will be good evidence for *A*'s trust in *B*.

Like the rational choice approach, **(T)** takes evidence of *B*'s motivations as evidence that counts in favor of *A* trusting *B* to do *C*. The difference is the kind of motivations that each emphasizes. The rational choice approach highlights evidence that *B* has self-interested motivations for doing *C*. **(T)** recognizes that people often have *other-interested* motivations; in other words, they are often motivated to act for the good of others even when it is not in their self-interest to do so. There are many types of motivations for acting for the good of others that are not self-interested reasons, including caring[39], a sense of honor[40] and a sense of duty. All of these can motivate *B* to do *C*, but a sense of duty is the motivation that most directly motivates trustworthy behavior. Often we do something because we believe we have an obligation to do it, and that sense of obligation is motivating for us. Thus, when *B* recognizes a relational obligation to do *C* for *A*, *B* may find the existence of that obligation a motivating reason to do *C*. This can be the case even when doing *C* is not in *B*'s self-interest. This provides room for a different kind of evidence to justify trust than is recognized by the rational choice approach. *A*'s possession of evidence that *B* will be motivated by a sense of duty to do *C* can justify *A*'s trust in *B* to do *C*.

[39] Even Hardin, a rational choice theorist, recognizes that *B*'s love for *A* provides motivation for *B* to be trustworthy, and thus evidence of *B*'s love for *A* can be evidence that justifies *A*'s trust in *B*: "When I fall in love with you, I partially take your interests as mine. I actually want you to be happy—*and not only because I will benefit from your happiness*…. My love of you grounds your trust because it makes me a trustworthy agent for your welfare and happiness" (Hardin, p.142 my emphasis).

[40] Steven Shapin (1994) provides a fascinating account of the role that a sense of honor played in motivating trustworthiness in seventeenth-century gentlemen.

What might count as such evidence for a motivating sense of duty?[41] The following are a few types of evidence that *A* might have for believing that *B* will do *C*: evidence that *B* has a general disposition to find obligations motivating, evidence that *B* has a localized disposition to be motivated by the type of obligation at issue in this situation, evidence that *B* recognizes that *A* and *B* have a relationship, evidence that *B* recognizes that the relationship generates an obligation to do *C*, and evidence that *B* is not subject to conflicting obligations to abstain from doing *C*. That people trust on the basis of such evidence is familiar from everyday experience. In order to know whether trusting someone is justified, we often try to learn something about her character. In particular, we want to know whether that person tends to live up to her obligations. We ask for character references both formally, by checking up on someone's job references, and informally, by asking friends and colleagues. The rational choice approach has a difficult time in accounting for the value of these types of inquiries. Why would the fact that *B* is thought by others to have a virtuous moral character provide evidence that it will be in *B*'s self-interest to do *C* now? Similarly, in evaluating the trustworthiness of others, we look for evidence that the trusted shares our values and recognizes the same norms as we do. For example, one might choose not to become close friends with an acquaintance because one learns that the acquaintance has a very different view of what friendship means (say, the acquaintance does not believe that one should keep one's friends' secrets). A professional may be reassured to learn that her colleague is a member of the same professional society that she belongs to because she takes this as evidence that her colleague has committed herself to the same code of ethics. The value of this type of evidence is that it provides reason to believe that the trusted will recognize

[41] The question of how one can live up to the epistemic responsibilities of justified trust is discussed further in chapter four.

an obligation to do the thing one trusts her to do.[42] A full account of the types of evidence that justify the kind of trust outlined in **(T)** is outside the scope of this project. However, it should be clear that **(T)** provides theoretical machinery for analyzing types of evidence for trust and motivations for trustworthiness that are omitted from the rational choice account of trust.

Another limitation of the rational choice approach is that it does not account for non-evidential reasons for trust. **(T)** does not share this limitation. Recall that *A*'s cognitive attitude towards the proposition that *B* will do *C* can be one of either belief or acceptance. If *A* accepts that *B* will do *C* in part because *A* believes that their relationship obliges *B* to do *C,* then *A* trusts *B* to do *C*. *A*'s reasons for accepting that *B* will do *C* must include this belief, but they may also include non-evidential reasons for accepting this proposition. Take Holton's example of the shopkeeper who trusts an ex-thief employee with the till. The shopkeeper's reasoning could be as follows: "This young convict is my employee, so she shouldn't steal from me. I'm not sure that she won't. But I am her boss, and maybe if I assume she won't steal from me, it will show her that people still think she is a responsible person despite her past. That might awaken her conscience and draw her back into the moral community." **(T)** recognizes the shopkeeper's attitude as one of trust. The shopkeeper has a non-evidential reason for accepting that the thief will not steal—the shopkeeper thinks that accepting this proposition will help bring the thief back into the moral community. The shopkeeper's belief that she has an employer/employee relationship that obliges the thief not to steal is part of the reason for this acceptance, since the shopkeeper is hoping that by holding up her end of the relationship (by trusting the thief), the pragmatic goal will be attained. Similarly, Lauri, the trusting graduate student, may trust her

[42] The rational choice approach can account for the value of this evidence only indirectly: we want to know that the other shares our norms because that gives us reason to believe that the trusted will fear punishment for breaking those norms.

advisor in the absence of evidence that the advisor will be trustworthy. Lauri may recognize that the advisor/advisee relationship obliges her advisor to treat her well and because she wants to make her interactions with her advisor go smoothly, she decides to base her planning on the assumption that the advisor will be trustworthy. In general, when one has a relationship that generates relational obligations, one often values that relationship. Thus, one can find oneself in situations when there are pragmatic reasons (often to do with facilitating or growing that relationship) for assuming that the partner in your relationship will live up to her obligations. These non-evidential reasons for trusting can be accommodated by **(T)**.

In chapter one, I argued that the rational choice approach's limitations prevent it from accounting for trust in unequal relationships. It is a puzzle, if one follows the rational choice account, why people in positions of powerlessness trust people in positions of power. Two types of powerlessness present particular problems for the rational choice approach: inability to detect untrustworthy behavior and inability to sanction untrustworthy behavior. Trust under these conditions of powerlessness is not puzzling on my account of trust. There are two ways that **(T)** can account for such trust. First, powerless individuals may reasonably choose to trust for non-evidential, pragmatic reasons. This is the case with Lauri, who despite having no (or very little) way of sanctioning her advisor for untrustworthy behavior, nonetheless chooses to trust her advisor for pragmatic reasons. Second, and I think more commonly, powerless individuals trust on the basis of evidence that the trusted is motivated to live up to her obligations. Scientists who collaborate with colleagues from different fields may be powerless to check up on the work of their partners in research, but even so, they may trust their colleagues on the basis of evidence that their partners are motivated by a sense of professional duty. Graduate students may recognize that they have little power to sanction the behavior of their advisors, but they still trust because they see that their advisors are morally upstanding mentors.

Such considerations can generate trust on the part of people in the most powerless of positions. Refugees living in camps around the world are caught in positions of extreme powerlessness. Routinely prevented from leaving the camps to work, they are almost completely dependent on non-governmental organizations (NGOs) for food and basic necessities. Refugees routinely have almost no voice in the way that the camps are run and have very limited communication with the outside world. For these reasons, refugees are often epistemically powerless and powerless to sanction the individuals who work for NGOs. Despite this stark power imbalance, refugees do sometimes trust NGO workers. This can happen when a refugee builds a personal relationship with the worker and comes to believe that the worker is motivated to do this work for moral reasons.[43] It seems to me that a refugee who sees that a particular worker is motivated by a sense of moral obligation and thereby comes to trust that worker is doing so with good reason. This kind of trust cannot be easily explained by the rational choice approach.

In this chapter, I have presented an account of trust that provides an alternative to the dominant rational choice approach to trust. While my account lacks some of the theoretical simplicity of the rational choice approach, it provides tools for analyzing features of trust and trustworthiness that are left out of the rational choice picture. Humans are not motivated by self-interest alone. People act trustworthily even when it is not in their self-interest to do so. People find moral obligations motivating, and we recognize this fact about each other. That is why we frequently look for evidence of a sense of duty in those we trust. Such evidence is particularly

[43] I saw this firsthand while working with refugees in Africa.

valuable to people in positions of powerlessness, who are not in a position to use the types of strategies outlined by the rational choice theorists. Powerless individuals also trust for pragmatic reasons. This illustrates another feature of trust left out of the rational choice picture. Trust is not simply an involuntary belief that someone will do something. Trust can be chosen in order to achieve particular goals. My account of trust provides room for all these features of trust. In the next chapter, I use this account to analyze trust between scientists. In doing so, I expose features of the social epistemology of science that have been left out of the dominant accounts, which have been driven by the same kinds of assumptions behind the rational choice approach to trust.

3.0 THE EPISTEMOLOGY OF TRUST IN SCIENCE

As discussed in the introduction, the past few decades of socio-historical study of science have challenged the traditional view of the field. The notion that science owes its privileged status as the paradigm of objective inquiry to scientists' disinterested dedication to the pursuit of truth has been perceived to be undermined by sociological and historical research showing that scientists are influenced by biasing factors. One response to this challenge has been to move towards a social epistemology of science according to which the objectivity of science appears at the community level. One common form of this move to social epistemology investigates how self-interested scientists can interact in ways that are communally objective (e.g. Goldman 1992, Hull 1988, Kitcher 1993, and Railton 1994). These social epistemologies of self-interest are interesting and useful analyses of the social aspect of the epistemology of science, especially since, as a matter of fact, scientists can be motivated by considerations of self-interest. However, self-interest is not the only biasing factor that can influence scientists. Scientists can also be swayed by their concern for others. Thus there is conceptual space for a social epistemology that examines how communities of other-interested scientists can be well-organized to act in rational ways that produce objective knowledge.

I provide such an epistemology in this chapter. I address the epistemological consequences of personal relationships between scientists by focusing on trusting relationships. I argue that scientific communities can garner several significant epistemic benefits from groups of other-interested scientists engaged in trusting relationships. Science as a whole benefits when scientists are trustworthy and when they trust each other to act trustworthily ('trust' and 'trustworthy' are used in the sense defined in chapter two). This analysis is analogous to the approach of the social epistemology of self-interest, for it shows how what were traditionally considered biasing factors can actually generate positive epistemic effects at the community level.

Like other social epistemologies, this approach does not pretend that the apparently biasing factor always plays a positive role in science. The goal of social epistemologies of self-interest is not to show that self-interest never has negative epistemic consequences for a scientific community. Similarly, a social epistemology of other-interest will not maintain that other-interest never leads to problematic particularism or secrecy. The goal of this chapter is simply to show that other-interested scientists who trust each other can create epistemic benefits for the community, just as Kitcher, Railton, and others have argued that self-interested scientists can create epistemic benefits for the community. After having read this chapter, one might be convinced of this point, but nonetheless remain concerned that trust may also have negative epistemic consequences. Chapter four addresses these concerns.

3.1 EPISTEMIC APPRAISAL AND SOCIAL EPISTEMOLOGY

Any social epistemology needs to present clear standards for epistemic appraisal. Of course, any epistemology needs to do so, but social epistemologists should be particularly careful about this task in order to avoid blurring the line between sociology and social epistemology. Sociology of science is not philosophy of science. Sociology of science is primarily a descriptive enterprise. It attempts to describe accurately the social practices of scientists. Philosophy of science in general, and epistemology of science in particular, are not and should not be primarily descriptive enterprises. While it is the job of the sociologist to describe what a particular practice or structure in science is like, it is the task of the epistemologist to specify epistemic standards and evaluate whether that practice or structure conforms to those standards. As Alvin Goldman puts it: "What interests the epistemologist are the epistemic consequences of such structures" (Goldman 1992, p.183). As a normative enterprise, social epistemology needs standards to evaluate the epistemic consequences of social practices. For social epistemology of science, these standards must be grounded in an account of what the epistemic aims of a scientific community ought to be.

Goldman argues that *veritism* is a fitting approach for social epistemology. Veritism "rates true belief as the ultimate epistemic aim" (Goldman 1992, p.190). A 'verific' social epistemology assesses the effects that social practices and structures have on various measures of true belief in a community. Goldman's version of veritism lays out five standards for evaluating social practices or structures in science. Thagard summarizes them as follows:

> 1. The *reliability* of a practice is measured by the ratio of truths to total number of beliefs fostered by the practice;
> 2. The *power* of a practice is measured by its ability to help cognizers find true answers to the questions that interest them;
> 3. The *fecundity* of a practice is its ability to lead to large numbers of true beliefs for *many* practitioners;
> 4. The *speed* of a practice is how quickly it leads to true answers;
> 5. The *efficiency* of a practice is how well it limits the cost of getting true answers. (Thagard 1997, p.247)

Put otherwise, a practice that prevents scientists from believing errors is a reliable practice; but a practice that enables scientists to believe truths is a powerful practice. As I understand it, Goldman's standard of fecundity is actually a species of the genus power since it is one measure of how well a practice enables larger groups of cognizers to believe interesting truths. To see that there are other possible ways that a practice could be powerful other than being fecund, consider the difference between a practice that enables a small elite to believe a large number of truths (a powerful practice that is not fecund) and a practice that enables a large number of scientists to believe a few truths (a powerful and fecund practice). The standard of speed measures how quickly practices and structures generate true beliefs, which is important for cognizers whose time for investigation is finite. Finally, the efficiency standard measures the costliness of a practice.

This chapter evaluates the epistemic significance of trust in science using a modified version of Goldman's standards. However, I shall not measure the reliability, power, fecundity, speed and efficiency of a practice in terms of how that practice contributes to the *true* beliefs of a community. There are two reasons why evaluating social practices in terms of their ability to produce true beliefs will not serve present purposes. First, how are we to know which beliefs are true in order to evaluate how reliable, powerful, fecund, fast and efficient our social practices are? While the truth of some of our beliefs about observations can be checked more or less directly (if anything can be), many others, those traditionally called 'theoretical,' cannot be. Our beliefs about whether a theory's predictions are borne out, or whether an attempted intervention was successful, can be checked. However, our beliefs about whether the higher reaches of that theory are true cannot be. This means that a social epistemologist who wants to measure the contribution that a social practice makes to reliability could not determine whether the practice increases reliability, because we do not know which of the beliefs are true. This kind of evaluation is necessary if Goldman's standards are to be applicable to the evaluation of science, but taking truth to be the ultimate epistemic aim makes such evaluation impossible. Second, since it is the goal of this project to show the benefits of an analysis of the epistemic significance of trust, I hope to provide an account that can be used by epistemologists of many stripes. This includes those who disagree about the nature of the epistemic aims of science. Axiological anti-realists, such as van Fraassen (1980), do not take truth to be the ultimate epistemic aim of science. Goldman's verific social epistemology makes his standards, measured in terms of a practice's contributions to true beliefs, unacceptable for such anti-realists. There is no need to create standards for social epistemology that take a position on the realism debate. Both realists and anti-realists can agree that increasing the empirical adequacy of the claims believed in a community is a good thing. Therefore, the standards used in this analysis do not follow Goldman's veritism.

Instead, my standards measure the epistemic benefits of a social practice in terms of their contribution to predictive power and utility for intervention. Epistemologists of any stripe ought to find these qualities valuable since practices that lead to better prediction and intervention are practices that produce more knowledge; prediction being a form of knowledge-that and intervention being a type of knowledge-how. Taking predictive power and utility for intervention as epistemically valuable is something upon which realists and anti-realists can agree. Anti-realists consider predictive power and utility for intervention to be the ultimate epistemic aim of science. Realists value empirical adequacy (in part) because they take it to be an indicator of truth. Therefore, I will measure the epistemic benefits of social practices by their ability to produce accurate predictions and successful interventions.[44]

[44] This is not to suggest that one could not add more epistemic values like explanatory power and simplicity to the list. Realists take explanatory power to be another mark of truth, so a realist may also want to evaluate social practices in terms of their explanatory power. However, following (van Fraassen 1980), many anti-realists take explanatory power to be merely a pragmatic value rather than an epistemic value. Thus, in order to avoid taking a

The social epistemology of science presented here takes accurate predictions and successful interventions (PI) as epistemically valuable and evaluates social practices in terms of how they contribute to various measures of PI. The measures used for evaluating how well social practices contribute to PI are adapted from Goldman's five standards for social epistemology (Goldman 1992, pp.195-6). Thus, the following standards will be used to measure the epistemic significance of trust:

1. The *reliability* of a practice is measured by the ratio of accurate predictions and effective interventions to the total number of checkable predictions and interventions enabled in a community by the practice.
2. The *power* of a practice is measured by its ability to help a community of scientists make accurate predictions regarding the questions that interest them and to intervene effectively on the aspects of the world that interest them.
3. The *speed* of a practice is how quickly it leads to accurate predictions and effective interventions.
4. The *efficiency* of a practice is how well it limits the cost of making accurate predictions and effective interventions.
5. The *fertility* of a practice is how well it enables a community of scientists to pursue accurate predictions and effective interventions for questions that they were previously unable to ask or investigate.

I have replaced Goldman's standard of fecundity, which is not particularly useful for present purposes, with an additional standard: fertility. As a species of power, fertility measures the ability of a practice to enable scientists to predict and intervene in areas that were *previously inaccessible*. This standard is based loosely on the value Kuhn called "fruitfulness", which he describes as the ability to "disclose new phenomena or previously unnoted relationships among those already known" (Kuhn 1977, p.322). Just as fruitful scientific theories can allow scientists to see new phenomena or recognize relationships that were previously hidden, fertile social practices and structures can permit scientists to pursue avenues of research that were previously inaccessible. Scientists value practices which enable them to pursue PI about a previously

position in the realism debate, I do not consider explanatory power in this chapter. This does not prevent a realist reader from doing so herself.

inaccessible part of the world. Epistemologists should recognize the epistemic praiseworthiness of communities that are able to tackle a wide variety of research questions.

Having now outlined the standards by which the epistemic consequences of social practices may be measured, I now turn to the analysis of the epistemic benefits of trust in science.

3.2 EPISTEMIC BENEFITS OF TRUST BETWEEN SCIENTISTS

Cooperation has a variety of epistemic advantages (Hardwig 1985 and 1991, Thagard 1997). In this section, I argue that trusting relationships confer epistemic benefits on a community in virtue of their role in fostering cooperation in the form of sharing. Trust motivates and sustains cooperation. Scientists are more likely to cooperate with people they trust. Thus, trust can encourage cooperation. However, as instances of scientific fraud show, trust by itself does not guarantee success. As Bernard Barber says about one case of fraud, “There was too much trust but not enough trustworthiness” (Barber 1987, p.130). Thus, successful collaboration requires not just trust, but also trustworthiness.[45] Fortunately, people may be more likely to be trustworthy towards people who they know trust them (Baier 1994, p.197). Of course, this is not always the case, so one cannot simply advertise one’s trust in someone and rest assured that she will act trustworthily. Cooperation in science can also succeed when trusted scientists are

[45] Cooperation is most stable when trusted individuals are trustworthy and trusters trust on the basis of good evidence for the trusted’s trustworthiness. However, in order for cooperation to get off the ground, it is not necessary that trusters have such good evidence. Trusters who trust for pragmatic reasons and who are fortunate enough to trust trustworthy individuals can also find success in cooperative activities.

motivated by a sense of duty to live up to their relational obligations to those who trust them. Thus, trust and trustworthiness can foster cooperation between scientists.

The cooperative practice analyzed in this chapter is sharing. Scientists engage in two main types of sharing: 1) sharing materials and technology, and 2) sharing information, ideas and critical scrutiny though the practices of shop talk and technical gossip. I analyze each type of sharing along the following lines. First, I show that scientists share in this particular way. Then I demonstrate that this type of sharing has epistemic benefits by evaluating its epistemic consequences according to the standards outlined in section 3.1. Third, I argue that this type of sharing is often grounded in trust and trustworthiness. Finally, I respond to the objection that these types of sharing, and the epistemic benefits they confer on communities, can be explained in terms of mere reliance instead of trust. In responding to this objection, I expand my argument to show that key features of these sharing practices cannot be explained without recognizing that trust is necessary for the formation and maintenance of sharing networks.

While numerous examples of sharing practices will be provided, one case study will be used to illustrate, in detail, the role of sharing in a scientific community. This case study focuses on the community of Drosophilists whose primary research tool were fruit flies of the *Drosophila* genus. The community of Drosophilists that emerged out of Thomas Hunt Morgan's lab at Columbia University at the beginning of the 20th Century acquired epistemic benefits from a cooperative style that was fostered by personal trusting relationships. This community has its origins in Morgan's decision, around 1906, to turn *Drosophila* into a laboratory animal. Morgan's lab and his students became the center of a growing community of Drosophilists as biologists realized what a powerful tool the fly could be. Numerous accounts of the community of Drosophilists reveal a highly cooperative group of scientists who shared ideas, materials, time and even credit (e.g. Allen 1978, Kohler 1994, and Sturtevant 1965). There are many possible explanations for the development and evolution of this cooperation. Personal relations of trust and trustworthiness, motivated by a shared code of morality, feature prominently in historical accounts as well as the scientists' own descriptions of their behavior. That trusting relationships were so prominent in the history of the so-called 'fly people' is particularly interesting given the productive and groundbreaking nature of their work. Thus, the work of the fly people will be used to illustrate both the forms that scientific sharing can take as well as the epistemic benefits that sharing can confer on a scientific community.

3.2.1 Sharing materials and technology

3.2.1.1 Scientists share materials and technology

Sharing materials and technology is one of the most concrete ways in which scientists cooperate with each other. For instance, biologists working on a model organism share varieties of stocks, and high energy physicists share technology when they collaborate on an experiment (Knorr Cetina 1999). In the case of biologists' stocks, scientists are sharing both concrete, material objects and pieces of technology. Stocks of flies bred by Drosophilists to possess certain combinations of useful mutants are technological innovations (Kohler 1994). Shared technologies can also be less concrete, such as computer programs for creating simulation models or for analyzing data. Materials or technology take different forms within a particular scientific field, but I suspect that almost all fields exhibit some degree of sharing of materials and technology. Even scientists whom we think of as relying mostly on simple observation, such as researchers into the behavior of meerkats in the wild, use materials and technology that can be shared across the field, such as special cameras for observing the meerkats underground.

Sharing materials and technology became a central feature of the Drosophilists' research practice. Their work depended on having access to a variety of mutant populations of *Drosophila*. Any new mutants that were discovered needed to be kept alive and populations of them maintained so that the mutant could be used as a research tool. Their stock-keeping work involved not only maintaining populations of new mutants but also populations of flies that the fly people had constructed through cross-breeding to possess multiple useful mutations. Keeping one's flies alive and in useable condition was no small task as any number of disasters could befall them: the flies could succumb to temperature shock, mite infestations, starve, escape, and be invaded by wild flies entering the lab or other mutants flying around the lab (Kohler 1994, p.82-3). The cost in terms of money, space and time of having a particular variety of fly available as a research tool thus involved not simply generating the mutant but also keeping it alive and in good condition. Since different mutants were useful for different experiments, Drosophilists were constantly asking each other to share their stocks. In fact, the amount of stock swapping was immense, because the Drosophilists were constantly changing their evaluation of the research value of particular mutants. As more was learned about each mutant, its relative usefulness often rose and fell (Kohler 1994, p.81). All of these aspects of the fly people's work meant that they simply could not make progress without devoting a significant amount of time to stock upkeep and preparing stocks for shipment to a colleague who had requested them. Thus, the Drosophilist community shows that sharing materials can become a significant and time-consuming part of a community's work.

3.2.1.2 The epistemic benefits of sharing materials and technology

A scientific community in which scientists share materials and technology can reap epistemic benefits that a stingy community of non-sharers forgoes. First, sharing is a process that increases *reliability*. In his discussion of testimony, Jonathan Adler (1994) has noted the power of *implicit replication* to uncover fraud and error in scientific results. A result will be implicitly replicated when it is incorporated into the investigations of others who attempt to build on the result. Adler argues that implicit replication can be a powerful force for uncovering fraud and error: "Once it is allowed that fraud and error are rapidly identifiable in intense areas of research, it follows that, over time, the uncovering of fraud is highly likely so long as an informant's results are, as is standard, an integrated part of a larger research project" (Adler 1994, p.267). Thus, implicit replication is a reliability-enhancing process, one that reduces the number of inaccurate predictions and failed interventions made by members of the community. This means that sharing is also reliable because a community in which sharing of materials and technology occurs is one with increased implicit replication. For example, when one scientist shares her mutant strain of *Drosophila* with other Drosophilists, the recipients of the mutant stock can detect problems with the strain (or problems with claims about the strain) that the generous scientist may have overlooked. For example, one of the problems with creating new mutants by cross-breeding is that one can create a mutant that possesses multiple mutations which have effects on each other (Kohler 1994). Not knowing that these other mutations are present can skew one's results. If multiple researchers are working with a particular stock, there is more opportunity for someone to detect a complicating mutation that decreases the reliability of results. Likewise, bugs in a computer program are more likely to be discovered when multiple researchers are using it in their work. Therefore, the reliability of results can be increased when materials and technology are widely shared within a scientific community.[46]

[46] One might object that sharing may decrease the reliability of results if researchers abandon a reliable material or technology in favor of a shared unreliable piece of material or technology. All objections of this kind (that sharing has negative as well as positive epistemic consequences), are dealt with in chapter four.

Second, sharing can increase the *power* of a community because more hypotheses are tested, i.e. more research questions can be tackled if materials and technology are not hoarded. Sharing is not valuable simply because it can help a community weed out errors; it can also help scientists attain accurate predictions and effective interventions. For example, the sharing of a technology for growing a model organism under extreme temperatures can increase the power of a community, since more scientists will be able to address different aspects of the model organism under this range of conditions. Additionally, since there are limits to the time and other resources individuals have, they are usually not capable of exploiting all the possibilities for research inherent in a material or piece of technology. Finally, sharing can be *fertility*-enhancing. Even if one individual were not limited by time or other resources, she is unlikely to be able to think of all the possible research questions that could be answered with her material or piece of technology. There is significant cognitive variation between individual scientists. As several philosophers of science have noted, different facts, or aspects of data, will be salient to different individuals (Kitcher 1993, Kuhn 1962/1996 and Solomon 1992). Thus stingy, unsharing behavior can deprive a scientific community of PI it may have discovered were more scientists able to use the resource to pursue the different questions that are salient to them.

Third, the practice of sharing can increase the *speed* of scientific research. Sometimes having just the right materials or technology can quicken the process of one's experiments. Also, when more scientists are working in a field, they can build on each other's work and use others' results instead of having to do everything themselves.[47] Finally, sharing materials and technology can increase the *efficiency* of research. The costs of producing materials are greatly decreased when scientists can borrow materials from others rather than having to produce them

[47] This is not to say that replication and checking of others' results is not needed and valuable.

by themselves. For example, the overall costs of developing particular *Drosophila* mutants are decreased when development only needs to be done once and then shared throughout the community. In summary, the practice of sharing materials can have the epistemic benefits of reliability, power, fertility, speed and efficiency.

3.2.1.3 The practice of sharing materials and technology is grounded in trust

Sharing of materials and technology is often grounded in trust and trustworthiness. Scientists tend to share their materials and technology with people they trust. In order to establish that any particular instance of sharing is grounded in trust, I need to show that both parts of **(T)** are met. In other words, I need to show that (*i*) the truster takes the cognitive risk described in part (1) of **(T)**, and (*ii*) that the truster takes this risk, in part, because she believes that her relationship with the trusted obliges the trusted to act as expected. I address each in turn.

Recall that trust, like reliance, involves risk. In chapter two, I argued that when *A* trusts *B* to do *C*, *A* takes the proposition that *B* will do *C* as part of her adjusted cognitive background. *A* makes plans on the basis of her adjusted cognitive background. Thus, trust involves the risk that one's practical reasoning will be undermined. When it comes to sharing, scientists trust each not to use shared materials and technology to scoop, steal or otherwise undermine the donor's research. Some of the risks involved in giving materials or technology to another include: that the receiver will plagiarize and take credit for the materials or technology, that the receiver will use the materials to complete one's own research project faster thereby rendering one's own work useless, that the receiver will use the materials to complete other research projects faster and thereby gain a better reputation than oneself, and that one will waste time preparing the materials for sharing instead of making progress on one's own research project. These are significant risks for the donor.

We can see these risks in the situation faced by the Drosophilists. Giving another Drosophilist one’s stock opened up the possibility that the other could try to scoop one’s research. Drosophilists often put years of work into a particular experiment that involved constructing just the right stocks for the task. By allowing another scientist to use stocks that one had taken years to perfect, one risked the possibility that the other could use those stocks to finish the experiment faster and publish the results first. When the Drosophilists shared with each other, they made plans on the assumption that the recipient scientist would not undermine their work in this way. Kohler outlines several particular things that the donors of stocks expected of the recipients. First, donors routinely worked on the assumption that the recipients would reciprocate by sharing their own stocks with the donor (Kohler 1994, p.143). Second, it was assumed that the recipient would fully disclose what they planned to do with the stock and keep the donor informed of the results of any experiments (Kohler 1994, p.144). Third, donors assumed that the recipients would not use the stocks to try to scoop the donors by completing their experiments faster (Kohler 1994, p.145). Thus, when the Drosophilists shared stocks, they risked having their plans undermined.

An unfortunate historical incident involving betrayed trust illustrates the importance of these plans. Milislav Demerec, a Drosophilist, was working with Franz Schrader, a cytologist, on the particularly promising project of producing a systematic genetic and cytological map of *D. virilis*. Demerec had spent years working on constructing a genetic map of *D. virilis*, so he had useful stocks of the organism. Demerec shared stocks with a Japanese group at the University of Kyoto under the understanding that they were only using them for a study of some translocations and inversions. But in 1936 Demerec and Schrader were “taken aback” when the Japanese team announced that they had completed cytological maps of *virilis*—the very goal that

Demerec and Schrader were working towards (Kohler 1994, p.155). Kohler reports that what upset Demerec was that the Japanese had not kept them informed about their progress (Kohler 1994, p.156). Thus, in this instance, Demerec was let down because he assumed that the Japanese would keep him updated about what they were doing with his stock. He was making plans for his own research based on the assumption that the Japanese team was not already working on the same problem at a faster pace. The secrecy on the part of the other team caused Demerec and Schrader a great deal of wasted effort. This is just one example that shows scientists sharing on the basis of their assumption that the recipient will do something. In cases like this one, scientists' cognitive attitude meets condition (1) of **(T)**.

So far, what I have said is consistent with an understanding of sharing networks completely in terms of mere reliance. As argued in chapter two, both trust and mere reliance involve vulnerability to having one's practical reasoning undermined. To show that sharing materials and resources is often grounded in trust, rather than always mere reliance, I need to show that donor scientists make the assumptions described above *because* they believe that their relationship with the recipient scientists obligates the recipients to act as expected. Focusing on the Drosophilists, I need to show that (a) they believed that the recipients were obligated to reciprocate, disclose, and not scoop, and (b) this belief was part of the reason why they assumed that the recipients would, in fact, live up to these expectations.

The Drosophilists believed that they were obligated to reciprocate and keep the donor updated on how the stocks were being used. The Drosophilists developed a moral code, which Kohler calls the rules of their moral economy, that supported public ownership and free sharing of stocks. Kohler explains the concept of a moral economy of the laboratory as follows: "…unstated moral rules define the mutual expectations and obligations of the various participants in the production process. … Moral conventions regulate access to tools of the trade and the distribution of credit and rewards for achievement" (Kohler 1994, p.12). One of the main theses of his account of the fly people's work is that they operated with a set of moral rules that guided appropriate interactions between them all. Reciprocity, disclosure and not scooping were three of these rules: "Reciprocity was one [rule]: the privilege of receiving stocks entailed the obligations to reciprocate" (Kohler 1994, p.143). Second, receiving stocks entailed the obligation to keep the producer of the stock updated on one's plans for using the stocks and what results one found from them; "[f]ailures to disclose were taken as serious reasons to worry about the borrowers' intentions" (Kohler 1994, p.144). Finally, Drosophilists embraced the rule that one ought not to use someone else's stocks to try to scoop them by completing their experiment faster (Kohler 1994, p.145). This is the obligation that the Japanese group failed to live up to when they scooped Demerec and Schrader. Thus, the Drosophilists recognized relational obligations; they believed that recipients had obligations to the donors of stocks.

As noted in chapter two, in most cases of trust, one trusts because one has evidence that the trusted is trustworthy. Since scientists tend to share materials and technology with people they trust, this means that scientists are watching out for signs of trustworthiness when deciding whether to share with someone. Recall that being trustworthy, in my sense, is different from being merely reliable in that we do not consider someone trustworthy if she is acting solely out of self-interest. Therefore, when scientists are determining whether to share materials or technology with someone, they often look for evidence of other-interest in the potential receiver. One thing scientists look for as evidence of trustworthiness is a sense of duty to follow the professional codes of conduct (or 'rules of the moral economy', in Kohler's terminology). Scientists will attempt to find out whether a potential receiver is trustworthy by making informal investigations into the character of the receiver by, for instance, contacting people who have worked with or shared materials with the receiver in the past.

The Drosophilists illustrate this practice of assuming that someone will act in a particular way on the basis of one’s belief that she is motivated to live up to her obligation to perform that action. It was widely known that anyone who passed through Morgan’s lab was taught the importance of playing by the rules of the moral economy. Allen and Kohler cite pieces of Morgan’s correspondence in which he would write to ex-members of the lab reminding them of these moral rules (Allen 1978, p.252-4; Kohler 1994, p.152). Drosophilists tended to assume that anyone who had spent time in the lab, or who had been properly acculturated to the rules of the Drosophila economy, was someone who could be trusted to act as expected. When, on occasion, doubts surfaced about whether it was justified to assume that someone would reciprocate, disclose, or not scoop, Drosophilists looked for evidence that the recipient had a sense of duty before they shared their stock with them. An interaction between Curt Stern and Gert Bonnier illustrates this point. At an earlier point in time, Bonnier flouted the rule of disclosure. Otto Mohr had given Bonnier one of Calvin Bridges’ stocks. Bonnier failed to tell either Mohr or Bridges what he was doing with the stock. As Kohler reports, “This lapse made [Bonnier’s] intentions suspect. Stern hesitated when Bonnier later asked him for stocks, and wrote Mohr first to ask if Bonnier could be trusted” (Kohler 1994, p. 145). Mohr responded that Bonnier’s failure to disclose previously was “due to lack of acquaintance with the scientific practice (coûtume) and not to bad will” (Mohr qtd. in Kohler 1994, p.146). Mohr told Stern that now that Bonnier had spent time in the Morgan lab, Mohr was confident that Bonnier had been educated to recognize a duty to disclose. He therefore told Stern that as long as he made it clear who the donor of the stocks was, that Stern could justifiably assume that Bonnier would live up to his obligation to disclose:

> Just denote the *special* stocks with the name of their owner…I cannot doubt that Bonnier, who has been at Columbia, now must be fully aware of the "étiquette scientifique." At the time he was just a beginner in scientific work. (Mohr qtd. in Kohler 1994, p.146)

Mohr does not reassure Stern by explaining why it would be in Bonnier's self-interest to disclose in the future. Instead, Mohr reassures Stern by pointing to the moral education that scientists received in Morgan's lab. Mohr provides Stern with reason to believe that Bonnier is motivated by a sense of obligation to disclose. Thus, were Stern to follow Mohr's advice in working it into his plans that Bonnier would disclose, he would be trusting rather than relying on Bonnier. Thus, trust that the recipient will not misuse the stocks is sometimes the foundation of sharing materials and technology.

3.2.1.4 The rational choice objection

In response to this evidence that scientists' sharing of materials and technology can be grounded in trust, a rational choice theorist might object that networks of sharing can be explained in terms of mere reliance on the self-interest of scientists. One might argue that sharing is grounded in the kind of reciprocity that motivates cooperation in iterated prisoners' dilemmas. One scientist might share with another in the hopes that the receiver will later return the favor with some of her materials or technology. On this account, scientists are reliable stewards of materials and technology that a colleague has shared with them because it is in their interest to maintain a sharing relationship with the donor for future sharing.[48] Similarly, donors can rely upon the self-interested motives of their colleagues if the donor is in a position to make a credible threat of punitive measures, such as refusing to share in the future or shunning by the community at large. Thus, one might wonder whether the epistemic benefits of sharing should be attributed to trust in science or instead to mere reliance.

To respond to this objection, I must first point out that my account does not pretend to show that trust is the *only* source of the epistemic benefits of sharing networks. There is no doubt that individuals operate from both self-interested and other-interested motives, and scientists recognize this fact. Scientists both trust and rely on each other. So considerations of self-interest do play a role in sharing of materials and technology. As will be discussed further in chapter four, a well-designed scientific community will have mechanisms for increasing sharing by increasing both reliance and trust between scientists. However, this rational choice objection is wrong to claim that all sharing behavior can be accounted for in terms of mere reliance. As I will now argue, there are key features of sharing networks that cannot be explained by mere reliance alone.

Not only do scientists trust their colleagues to use shared materials and technology properly, but scientists act trustworthily out of other-interest. I have shown that scientists share because they believe their colleagues to be other-interested. They look for evidence that their colleagues are motivated by a sense of duty to live up to the obligations inherent in the donor-recipient relationship. Showing that scientists look for such evidence, and make decisions on the basis of it, is not quite to show that scientists actually are other-interested. However, the most

[48] This is the explanation Rescher provides for sharing of information in science (Rescher 1989). Michel Blais' (1987) account of cooperation in science as a kind of epistemic tit for tat is another example of this rational choice explanation.

convincing evidence of the role of other-interest in supporting networks of sharing materials and technology is an example of a scientist living up to an obligation out of a sense of duty even when it is not in her interest to do so. The Drosophilist community provides such examples of other-interested altruism.

In addition to recognizing duties that recipients of stocks had to donors, the Drosophilists recognized duties on the part of potential donors. The Drosophilists were expected to share stocks upon request: “It was customary to get permission from the inventors of these valuable and versatile tools before using them, but it was taken for granted that permission would not be refused” (Kohler 1994, p.145). The Drosophilists believed that they were obligated to share their materials and technology. Morgan articulates such a sense of obligation in the following passage from a letter:

> We make a point of supplying any individual or group of individuals with any material in stock, not only material that has been studied by ourselves but also material as yet unpublished if it can be utilized. The method of locking up your stuff until you have published about it, or of keeping secret your ideas and progress has never appealed to me personally, and I think as a simple matter of policy that such a procedure is as injurious to the student as it is to the progress of science, which we profess to have most at heart. (qtd. in Kohler 1994, p.134)

This sense of obligation is the reason why stocks were even shared with individuals from whom one could expect little in terms of reciprocation. The elite members of the fly people at research institutions like Columbia and the University of Texas regularly put significant amounts of time and resources into providing stocks for teachers at small institutions to use for their students. For example, the Universities of Chattanooga, Kansas, Oklahoma, New Mexico, and Pittsburgh, as well as colleges in Texas, Allegheny and Procopious were given teaching stocks (Kohler 1994, p.143). It is interesting that the elite Drosophilists did this work even though it required quite a sacrifice in terms of time and effort:

> Supplying stocks for teaching was no small burden. Sometimes [Charles] Zeleny [who was one of Morgan's graduate students] was asked to supply directions for culturing, breeding, and handling socks, and colleges had to be continually resupplied, since there were no researchers there to maintain stocks, and cultures died out every summer and had to be "ordered" again each fall. (Kohler 1994, p.143)

These institutions could not be counted on to produce their own new stocks that the elite groups could hope to get in return. Since the teachers were not top-level researchers, the Drosophilists could also not hope to get much in terms of news of promising new research in return for their stocks. Thus, it is hard to explain the donation of stocks to these universities in terms of a desire on the part of the fly people to build profitable reciprocating relationships with other Drosophilists. Thus, trustworthiness, in the sense of living up to one's obligations out of other-interest, was a feature of the Drosophilist sharing network.

This example of other-interested altruism reveals the key feature of sharing networks that cannot be accounted for in terms of mere reliance and self-interested reliability. Recall from chapter one that the rational choice approach has trouble accounting for trust on the part of the powerless. This is because there is little incentive for the trusted to act trustworthily towards the powerless. To apply this to the case of the Drosophilists, the elite Drosophilists shared stocks with small institutions at significant cost despite receiving little to no benefit from it. This point will apply generally in many different cases of powerlessness on the part of scientists. For example, if a scientist is marginalized in the community or if she is unlikely to have anything that the receiver will want in the future, she could not rely on self-interest alone to motivate her receiver to use the materials and technology responsibly. Thus, mere reliance and reliability cannot explain sharing in scientific communities in which members differ in their value to each other.

Another key feature of sharing networks that is difficult to explain in terms of mere reliance is how new sharing networks get off the ground. Recall that the rational choice theorist's objection to my account of the role of trust in sharing is that scientists need not trust each other in order for sharing to occur; sharing networks can function if scientists rely on threats of punishment for uncooperative behavior. These punishments can take place at the level of individuals, as when one party refuses to share with another who previously failed to act in the cooperative manner expected. Punishment can also happen at the community level, as when a whole community shuns or expels a member who failed to live up to the norms governing sharing. As Niklas Luhmann puts it, these threats of sanctions make the interests of each party interdependent (Luhmann 1979, p.36). I, the truster (in the rational choice sense) have an interest in you acting as expected, and you, the trusted have an interest in acting as expected so that you can avoid the punishment. In situations in which these interdependencies exist, the rational choice theorist can provide a nice account of the rationality of sharing. However, the situation is very different when the threat of punishment is not yet present. Before a sharing network exists, there is no threat of being excluded from such a network if one acts uncooperatively. So it seems that the rational choice theorist will encounter difficulties explaining why someone would take that first step to share.

Before there is a network of sharing, there can be no network-level norms that individuals are punished for not obeying. Thus, the first individual to share cannot count on the threat of shunning by or expulsion from the network to motivate cooperative behavior by people with whom she shares. In response, rational choice theorists like Blais and Rescher point to iterated Prisoners' Dilemmas to explain the development of cooperation, but even here there are problems. The kind of punishment on the rational choice view is the cutting off of the sharing relationship. The idea is that the person contemplating whether to share can rehearse to herself the following argument: “If I share with this person, then she will probably act cooperatively with me, because she will calculate that it is in her self-interest to cooperate, for she knows I will cut her off from future sharing if she doesn’t.” The problem is that, as Luhmann puts it, “The argument does not directly rest upon this calculation [of the potential trusted person], but upon the truster’s ability to anticipate it” (Luhmann 1979, p.36). The first sharer anticipates that this will be the calculation that the recipient will make, but what grounds does she have for anticipating this? How is the recipient to know that the sharer will punish her by cutting her off? In the first round of interaction, the recipient has no guide to predict how the sharer will respond to certain behaviors. While many donors would not continue to share with recipients who do not reciprocate, not everyone will retaliate to uncooperative behavior. Some donors may continue to share with uncooperative recipients because these donors feel that sharing is the morally right thing to do. Human nature is sufficiently diverse on this matter of retaliation that the recipient cannot automatically assume that if she fails to cooperate, the donor will cut off the relationship with her. So the sharer cannot assume that the recipient will calculate that it is in her best interest to cooperate with the sharer.

One way to solve this problem would be for the first sharer to make the threat of punishment explicit. But Luhmann points out that this will not work:

> The structure of the trust relationship requires that such calculation should remain latent, evolving in its generalizing fashion covertly, purely as a reassuring consideration. In his overt behaviour the truster must show himself ‘utterly trusting’, lest he himself sow the seed from which later reciprocal distrust may grow, thereby producing exactly the results which he is trying to avoid. (Luhmann 1979, p.36)

Explicitly threatening someone shows a lack of trust that risks "poisoning the relationship" with the person whom one threatens (Luhmann 1979, p.36). People do not react well to being treated with suspicion. Someone treated with distrust tends to respond in the following ways:

> In so far as he continues the relationship at all, he will respond at first perhaps with explanations, with forbearance, then with caution and finally with distrust himself. He feels himself relieved of moral obligation by the distrust which is brought against him and given the freedom to act in his own interests, or indeed actually feels the need to revenge himself for this unearned treatment. And thus he gives distrust additional justification and further nourishment. (Luhmann 1979, p.74)

Thus explicitly threatening the potential recipient with punishment for failing to act in a cooperative manner is unlikely to encourage that recipient to act cooperatively. It is more likely to poison any sharing relationship before it starts. Thus, the first sharer's behavior cannot be explained in rational choice terms. Either she is not warranted in anticipating that the recipient will act cooperatively because the recipient may not recognize the threat of punishment, or she explicitly threatens the recipient, which dooms the sharing relationship to mutual distrust and retaliation from the start.[49]

In response, a rational choice theorist may respond that the first step in a sharing relationship is justified because it is part of a long-term successful strategy. Both Rescher's economic model of trust and Blais' account of trust as a form of epistemic Tit-for-Tat, present the initial cooperative move as justified because it is part of a cooperative but punitive strategy that yields benefits over the long-term (Blais 1987, Rescher 1989).[50] The idea is that a long-term strategy of cooperating with cooperators and punishing uncooperative defectors is successful (Axelrod 1984, Rescher 1989, p.37). In order for one to get the benefits of this long-term

[49] Luhmann puts this point as follows: "Because contrived interchanges of this [punitive] nature are not expected, to bring them about requires open communication about them between the people concerned; and this, as we have suggested, may introduce an atmosphere unfavourable to trust into the relationship (Luhmann 1979, p.36).
[50] For more on these types of explanations of cooperation, see Axelrod (1984) and Woods (1989).

strategy, one must take an initial step of “blind trust” (Rescher 1989, p.38). In other words, these rational choice theorists bite the bullet to some extent and respond to my objection that the first step in a reliance-based cooperative practice is not justified at the moment it happens. It only becomes justified later as it is shown to have been part of a successful strategy for stimulating a profitable cooperative relationship (Rescher 1989, p.40). Rescher compares this process to the process by which we justify our reliance on cognitive instruments like telescopes or microscopes:

> In inquiring, we cannot investigate everything; we have to start somewhere and invest credence in something. But of course our trust need not be blind. Initially bestowed on a basis of mere hunch or inclination, it can eventually be tested, and can come to be justified with the wisdom of hindsight. (Rescher 1989, p.40)

Therefore, a rational choice theorist might argue that I have been asking for the wrong type of justification for the first step in a sharing network. I have been arguing that the rational choice theorist cannot provide a justification at the moment of the first act of sharing, but the rational choice theorist responds that justification by hindsight, is instead all that is required—or possible.

The problem with this response is that it only makes sense for someone to take this initial act of blind trust if the costs for being let down are low. If I stand to gain large benefits in the long-run by taking a small initial risk, it makes sense to take that risk. However, if the potential costs of being wrong on that first risk are high enough, they may not be outweighed by the long-term benefits. Rescher says, “When others tend to respond in kind to one’s present cooperativeness or uncooperativeness, then *no matter how small* one deems the chances of their cooperation in the present case, one is nevertheless well advised to act cooperatively” (Rescher 1989, p.37 my emphasis). But this is too strong. It is not rational to take the first sharing step, if the costs in the present case are high enough. By the rational choice theorists’ own lights, if the costs of your uncooperativeness are high enough, then I ought not to cooperate with you if the chances of your cooperation are low enough.

This is why this rational choice account provides a poor explanation of the origin of sharing networks in science. Recall how significant the costs are for the scientist who has her research program scooped. These costs are high enough that it would not be rational for the initial sharer to trust without first making a determination of how likely it is that the first recipient will act cooperatively. This means that some justification of that first decision to share is required. The costs of having one's whole research program undermined require a scientist like Morgan to have some justification for sharing. Thus, the rational choice theorist finds herself back in the position of the dilemma presented by Luhmann: either the first sharer makes the threat of punishment explicit, or the first sharer takes an unjustified risk.

In contrast, my account of trust can explain how sharing networks get off the ground. The first sharer may reasonably decide to share if she has evidence that the recipient is a morally upstanding person. For example, if I know that you are motivated by such moral principles as the Golden Rule, I may have reason to assume that you will not act uncooperatively when I share with you. Another possible explanation within my account for how sharing gets off the ground is that the first sharer decides to share for pragmatic reasons. Just as the shopkeeper from chapter two decided to trust the ex-thief with the till in order to bring the thief into the moral community, I might decide to share with you in order to inspire you to share with others (including myself). This seems to provide a plausible interpretation of Morgan's role in getting the Drosophila sharing network off the ground. Morgan clearly believed that biologists ought to share their materials and technology with each other. When he first brought *Drosophila melanogaster* into the laboratory and found it to be a useful research tool, there was no

Drosophila community with whom to share his stocks. It was quite a significant risk that he took in sharing his stocks with others. It was entirely possible that he could have shared his stock with someone in another laboratory who would have been able to make more progress with the organism than he did. One way to explain why he took this risk by sharing his stocks with other researchers who were interested in starting work on this organism is that he wanted to inspire those who wanted to work on the organism to create a sharing community. These are reasons for sharing that, unlike the threat of sanctions, can be explicitly communicated to recipients without undermining the sharing relationship.

In conclusion, my account of the epistemic benefits of sharing materials and technology as grounded in trust explains two key features of sharing that cannot be explained in terms of mere reliance. Not only does the rational choice approach have trouble accounting for sharing between parties of unequal power, but it also faces difficulties explaining how sharing networks initially get off the ground. Having defended my account from the rational choice objection, I now turn to another form of sharing which is also epistemically valuable and grounded in trust.

3.2.2 Sharing shop talk and technical gossip

3.2.2.1 Scientists share shop talk and technical gossip

Science flourishes when there is free sharing of ideas, information and critical scrutiny. Scientists share these things both formally, in publications and presentations, and informally, in conversations and correspondence. Much of the philosophical work on trust in science has emphasized scientists’ dependence on testimony (e.g. Hardwig 1985 and 1991, Kitcher 1993, and Shapin 1994). As students, scientists learn by absorbing facts from textbooks. As they progress to mature scientists, they shift from learning from textbooks to learning about their field from journal articles and presentations at conferences. As philosophers point out, most of what scientists believe about the history of their field (e.g. their beliefs about how the structure of DNA was discovered) and what they believe about the current state of their field (e.g. their beliefs about what we currently know about the genetic basis of behavior) is grounded in testimony rather than direct experience or evidence. Thus, much of what scientists believe is grounded in these modes of formal communication—textbooks, journals and presentations. However, there is also another significant mode of communication for scientists: informal communication (Crane 1972). In this section, I argue that two forms of informal communication—shop talk and technical gossip—rest on a foundation of trust and trustworthiness.

Anyone who has spent much time with a group of scientists will have noticed them engaging in what Knorr Cetina calls ‘technical gossip’ (Knorr Cetina 1999). Technical gossip takes place at the laboratory bench, during lunch breaks and even during social occasions when scientists are not at work. Examples of technical gossip include discussion about how so-and-so’s experiment is progressing, how so-and-so’s working relationship with her partner is deteriorating, how so-and-so’s work habits are changing. A paradigmatic example of technical gossip is: “I just talked to someone in Greg’s lab and he can’t get his flies to grow on the new medium he’s trying out.” Technical gossip is gossip insofar as it is informal conversation about a third party or parties; it is technical because it refers to the third party’s scientific work rather than purely personal affairs. It is important to recognize that while in ordinary English ‘gossip’ has negative connotations, ‘technical gossip’ does not necessarily have negative valence. Technical gossip includes two parties innocently discussing the well-known successes of a third party’s research. But sometimes technical gossip can have negative valence, such as when one person tells another something about a third party that the third party would not want discussed. Knorr Cetina, following Bergmann's (1993) analysis of gossip in general, characterizes gossip as a form of communication with the following features:

> …certain relational characteristics (it features A gossiping to B about C) and a sequential structure (there may be an opening sequence, then the gossip, then a closing sequence). It is situationally embedded (there is gossip within purely sociable interactions and within work), and the subject of gossip is established and managed in specific ways (for example, the persons who become the subject of gossip are initially named). There exist a number of policies of information presentation (for example, the information is presented as worthy of communication, as believable, and passively acquired), and the gossipers often appear to indulge in the joy of speculating and are interested in generalization. (Bergmann 1993: 71ff) (Knorr Cetina 1999, p.205)

Technical gossip, Knorr Cetina says, always includes a reference to a knowledge object (e.g. fruit flies, subdetectors, PCR technique). Thus, by engaging in technical gossip, scientists use an everyday practice to communicate about the objects of study, theories, methodologies, technologies, results and collaborations that compose their field. To summarize, technical gossip is informal conversation about a third party or parties that refers to the third party's scientific work.

Shop talk is another type of informal communication that has epistemic benefits and is fostered by trusting relationships. I define 'shop talk' as informal communication about one's own work. Thus, it is distinguished from technical gossip in that it is not about a third party. Scientific shop talk includes discussion about the progress of one's experiments, one's ideas for research, one's methodology, results, partnerships and other features of one's research. Scientists frequently share the details of their research and ideas with each other in informal conversations. They do so in order to let others know what they are doing and also to get their help in dealing with problems they are having. Shop talk, therefore, includes critical scrutiny. For example, one scientist might ask for her laboratory benchmate's advice about some new puzzling data. Much scientific communication happens through informal shop talk rather than formal publications or presentations.

One of the widely discussed features of the Drosophila community, and Morgan's lab in particular, has been its collaborative style of inquiry. Allen describes the collaborative environment as follows:

> A unique feature of the Morgan group was the atmosphere in which the major ideas were developed and worked out. There was a constant give and take and sense of equality among the four central members, but this sprit rubbed off to varying degrees on all those who spend any time in the fly room. (Allen 1978, p.188)

Kohler describes life in the early years of Morgan's lab as communal:

> It is crowded, crammed with eight desks and the paraphernalia of *Drosophila* production: odd-looking homemade incubators, shelves loaded with milk bottles, tables for preparing fly food…. There were no personal, private spaces in the fly room, except for Morgan's small adjoining office, the door of which was always open. Space was arranged so that everyone knew what everyone else was up to. (Kohler 1994, p.98)

The inhabitants of the lab took advantage of their close quarters to work closely together: "The fly room was a noisy place, too, with the clinking of bottles and an unceasing flow of banter, gossip, and shoptalk. Every new result or technical problem, or letter from another lab was openly and vigorously discussed by everyone present" (Kohler 1994, p.98). Alfred Sturtevant and Calvin Bridges, the two main centers of activity in the lab, were known for stimulating this shop talk. Dobzhansky remembered Sturtevant as "talking much of the time" (qtd. in Kohler 1994, p.99), and Muller recalled Bridges as follows:

> We'd keep up a more-or-less running, often desultory, conversation as things came up. We'd speak freely to one another. 'Oh,' Bridges would say, 'Here is a strange case I just got the results from.' We'd all discuss it, and I might make a suggestion about what to do next… and so, we simply went along in an informal way, talking things over with one another as they came along; very seldom on what you might call a philosophical plane. (qtd. in Allen 1978, pp. 190-1)

This sharing of ideas and collective problem solving may have started in the cramped Columbia lab, but it spread throughout the national, and international, Drosophilist community alongside the work and tools of the fly group. Drosophilists shared news, ideas and craft lore through extensive letterwriting. Letters between a scientist in one lab and another somewhere else were semi-public communications, since they were read aloud in the lab and shared with colleagues (Kohler 1994, p.99 and p.165). Technical gossip and shop talk were conspicuous features of the Drosophilist community.

3.2.2.2 The epistemic benefits of sharing shop talk and technical gossip

Communities of scientists who, like the Drosophilists, engage in much collaborative shop talk and technical gossip can garner many epistemic benefits. Technical gossip can have a positive impact on the *power* of a scientific community. As philosophers working on the epistemology of testimony have noted, much of science today is done in collaborations (Blais 1987; Hardwig 1985 and 1991, Shapin 1994). Collaboration can have significant epistemic benefits for a community. A recent study of 19.9 million papers over five decades has shown that teams produce more “high-impact” research in the sciences than do individuals (Wuchty et al. 2007). In order for teams to be able to produce this high-impact work, it is important that scientists be able to find good collaborators with whom to work. Informal communication like technical gossip can provide scientists with useful information that helps them locate technically competent and morally upstanding colleagues with whom to work. Reports from scientists who have worked with a potential collaborator in the past can prove invaluable. Public information about a scientist’s past work, such as a publication record or CV, which many academics now publish on their websites, provides very limited information about what that scientist is like to work with. Without technical gossip, one might be able to learn that a potential colleague has been a part of several successful teams that published high-impact work in the past, but one would not be able to determine to what extent that individual contributed to that team's work. In addition, information about the moral character of scientists is spread through technical gossip. Learning that a scientist has an upstanding character can reassure a potential collaborator who may have doubts about whether a partnership is a good idea. This would seem to be of particular value for scientists considering an interdisciplinary collaboration. As was discussed in chapter one, scientists involved in peer-different collaborations experience a kind of epistemic powerlessness since they do not have the expertise to check the work of their partners. Thus, a scientist considering an interdisciplinary collaboration would value technical gossip that shows a potential collaborator to be morally upstanding. Since interdisciplinary work often has the potential to open up whole new avenues of research that were previously inaccessible, practices like technical gossip that encourage interdisciplinary work increase the *fertility* of a research community. Having those doubts assuaged through technical gossip can clear the path to a fruitful collaboration that increases the power and fertility of the research in that scientific community.

Shop talk can also have a positive impact on the *power* of scientific research. When scientists discuss their work with each other, they can help each other come up with solutions to problems. This is amply illustrated in the case of the Drosophilists. This quotation from Muller already presented above is a classic description of the communal problem solving style of this community:

> We'd keep up a more-or-less running, often desultory, conversation as things came up. We'd speak freely to one another. 'Oh,' Bridges would say, 'Here is a strange case I just got the results from.' We'd all discuss it, and I might make a suggestion about what to do next…and so, we simply went along in an informal way, talking things over with one another as they came along. (qtd. in Allen 1978, p.190)

Thus shop talk can have significant effects on the power of scientific research within a community.

Technical gossip can have a positive effect on the *reliability* and *efficiency* of results. While collaboration can have significant epistemic benefits, it also involves risks of scientific misconduct. As scientists depend on each other more and more, there is increasing room for dishonest or careless scientists to do harm to the community if their misdeeds go unnoticed by their collaborators. When a community uncovers a case of scientific fraud, there is much anger over the errors the fraud perpetuated and the time wasted by other researchers who tried to replicate or build on the fraudulent work (Barber 1987). This anger is commonly directed not only at the wrongdoer, but also at the wrongdoer's colleagues who, it is thought, should have caught the errors and stopped the rest of the community from adopting the fraudulent work. Communities in which scientists either are able to avoid working with dishonest or careless partners, or are able to detect their misconduct, are in a better epistemic position in regards to the reliability and efficiency of their work than are communities in which scientists are unable to do so. Thus, practices and structures which enable scientists to learn about and keep an eye on their potential or actual collaborators increase reliability and efficiency.

Technical gossip can be one such practice. Suppose I am a biologist looking for a chemist to be part of my interdisciplinary experiment. If I hear some technical gossip about how chemist *A* has performed suspicious experiments on another interdisciplinary project or has showed a lack of circumspection when assessing her own abilities in my subfield of biology, then I have some prima facie reason to avoid working with *A*. When it functions as an informal source of information about potential collaborators, technical gossip can be a community-level mechanism for isolating those scientists who are prone to contribute negatively to the reliability and efficiency of the community's work. By providing individuals with information that is likely to disincline them to working with the problematic members of the community, technical gossip can confer a benefit not just on the individuals who avoid bad partners, but on the community as a whole.

Shop talk also increases *reliability* because it facilitates implicit replication as scientists learn from each other and incorporate the work of others' into their own research. In addition, shop talk provides a medium for scientists to submit each other's work to critical evaluation. When one scientist shares her work with others, the chance that problems with her work will be discovered is greater. Thus, critical scrutiny by others increases reliability by increasing the likelihood that errors will be detected. Not all results make it to publication. In particular, null results are harder to publish in a peer-reviewed journal than are positive results. Thus, if such results are to receive critical scrutiny, it will often need to be through the informal communication method of shop talk. Therefore, a community in which shop talk is prevalent can have greater reliability of results since more results will be subjected to critical scrutiny.

Technical gossip and shop talk also increase the *speed* of research because it facilitates the spread of information. Informal means of communication are often faster than formal communication in publications that sometimes publish an article that is a year or more old.[51] A community of scientists that has access to the latest information through technical gossip and shop talk produces results at a faster pace than one which relies on purely formal means, because scientists have access to the latest results and techniques. In addition, when scientists engage in the collaborative problem solving discussed above, they not only increase the likelihood that they

[51] However, the rise of internet journals and internet-first publishing in recent years has decreased this lag time in some fields.

will attain PI, but they also arrive at PI faster. Sturtevant describes this as one of the benefits of the Drosophilists' collaborative problem solving approach:

> There was a give-and-take atmosphere in the fly room. As each new result and each new idea came along, it was discussed freely by the group. The published accounts do not always indicate the sources of ideas. It was often not only impossible to say, but was felt to be unimportant who first had an idea….I think we came out somewhere near even in this give-and-take, *and it certainly accelerated the work*. (Sturtevant 1965, pp.49-50; my emphasis)

Thus, the speed of research can be increased by informal means of communication.

3.2.2.3 The practice of sharing technical gossip is grounded in trust

Trust has indirect epistemic benefits because it helps create the atmosphere that makes technical gossip and shop talk possible. To prove this I need to show that in sharing technical gossip and shop talk, one takes a cognitive risk and that one does so, in part, because one believes that there are relevant relational obligations. In addition, I need to respond to the obvious objection that this form of sharing can be accounted for in terms of mere reliance instead. I address these points regarding technical gossip first and shop talk later.

There are risks involved in participating in technical gossip, just as there are for engaging in everyday gossip. Since it is part of the nature of gossip that one is discussing the affairs of another person, there is the potential that one will say something that the person being gossiped about would not appreciate. One is not, therefore, usually able to be sure that one is not engaging in the kind of gossip that has negative valence. This being the case, we depend on the person with whom we are gossiping to exercise enough tact and discretion not to spread the gossip further in a way that will reflect poorly upon us. It would be indiscreet for the recipient of my gossip to go to the third party and tell her that I am gossiping about her, unless the recipient is confident that the third party will not mind me discussing her affairs. Thus we often want to be able to trust our partners in gossip to be able to keep the gossip confidential. Even if one is not concerned that what one is saying would upset the person one is gossiping about, one may still have other reasons for wanting the gossip to be confidential. For example, one may be sharing some information that one does not want one's competitors to find out. Scientists who trust each other to respect the often unclear, but nonetheless important, boundaries of confidentiality are more likely to engage in technical gossip with each other. When we gossip, we normally do so, in part, because we assume that the recipient of the gossip will exercise discretion and tact. It is this assumption of discretion that one takes as part of one's adjusted cognitive background when deciding whether or not to gossip with someone. Thus, gossip involves the kind of cognitive risk that fulfills part (1) of **(T)**.

We make this assumption, in part, because we believe that the recipient has a relational obligation to exercise discretion and tact. The key to recognizing this is that gossip is something that typically takes place between friends. Recall from chapter two that relational obligations arise in a couple of ways. First, sometimes we explicitly agree to take on relational obligations as a sign that we are indeed in a particular kind of relationship. A common opening to an exchange of gossip is for the sharer of the gossip to say something along the lines of "Promise you won't spread this around…." The recipient of the gossip may reply by agreeing to exercise discretion because she wants to reassure the sharer that they are friends. Second, relational obligations can also derive from the trusted's implicit agreement to behave in a particular way given their participation in a certain kind of relationship with the truster. Thus, if I am friends with someone, I implicitly agree to exercise discretion with any information she shares with me. In both of these ways, people can incur obligations to exercise discretion and tact. We depend on our friends to live up to these relational obligations, and we feel betrayed when they fail to do so.

3.2.2.4 The rational choice objection about technical gossip

At this point, a rational choice theorist may object that moral concepts like friendship and relational obligation are not necessary to explain the existence of technical gossip—the motivation to engage in technical gossip can be explained in purely self-interested terms. According to this account, I exchange gossip with someone with the hopes that she will reciprocate and provide me with useful gossip in the future. The recipient exercises tact and discretion with the gossip I give her because she wants me to share gossip with her in the future. In other words, she can be counted upon to be discreet because it is in her interest not to be punished by having future gossip withheld from her.

In response, I do not deny that self-interested motives can account for some features of technical gossip. In particular, the reciprocity of many instances of gossip may be well accounted for in self-interested terms. However, I find it implausible that it can account for discretion on the part of the recipient of gossip. The problem with gossip, one that we are all familiar with, is that once one has told someone something one has almost no way of knowing how the information one shared is being presented to others. Thus, the truster is in a position of the kind of epistemic powerlessness discussed in chapter one. Perhaps in a small community, one could, with only casual inquiry, find out whether and how one's gossip is being spread around. But in larger communities, this is impractical. Every person knows many people with whom she could share gossip. If I tell Susan something, there are many people with whom Susan could indiscreetly share my gossip. I may not normally interact with these people. Thus, the gossip may spread through the community in a way that reflects poorly on me, and I would not know. This makes it extremely hard to determine whether one's partner in gossip has let one down. One could, I suppose, constantly check in with all the members of the community to see whether they have heard what you said about so-and-so, but the costs of such vigilance are bound to outweigh the benefits, and those people are not bound to tell you the truth. In any case, one is in a position of relative epistemic powerlessness. When one has difficulty determining whether the relied upon has let one down, one is unable to retaliate effectively against her. Thus, the reciprocity rational choice explanation provides a poor explanation of the practice of gossip in general. I see no reason why it should provide a better explanation in the particular case of technical gossip between scientists.

3.2.2.5 The practice of sharing shop talk rests on trust and is difficult to explain in terms of mere reliance

The sharing of information about one's own work through the practice of shop talk rests on trust in much the same way that sharing of one's own materials and technology does. The same risks of scooping are present for the scientist who shares information about her work. The scientist who shares information about her work assumes that the recipient will not use the information to scoop the donor. In making plans on the basis of this assumption, the scientist becomes vulnerable to having her practical reasoning undermined. Thus, by sharing, scientists adopt the cognitive attitude described in part (1) of **(T)**. The same arguments given above for why sharing materials and technology rests upon trust rather than mere reliance also apply to shop talk. First, mere reliance accounts cannot explain why powerless individuals share information about their work. Second, rational choice accounts cannot explain how networks of sharing information about one's work, which carries significant risks, get off the ground.

Therefore, shop talk, like sharing materials and technology is best explained in terms of a background of trust.

There is one additional feature of the background of trust that supports shop talk in communities that engage in extensive collaborative problem solving. Assigning individual credit for work in such a community is problematic. When everyone jointly engages in the process of criticizing and refining the research, it is difficult to give each person who contributed to the final product the appropriate credit for the work she did. Communities which work highly collaboratively sometimes resolve this problem by abandoning the private ownership of ideas. This was a widely recognized feature of Morgan's lab:

> A unique feature of the Morgan group was the atmosphere in which the major ideas were developed and worked out. There was a constant give and take and sense of equality among the four central members, but this sprit rubbed off to varying degrees on all those who spend any time in the fly room. Especially in the earlier years (between 1910 and 1915) there was little concern about priority or ownership of ideas. (Allen 1978, p.188)
>
> This group worked as a unit. Each carried on his own experiments, but each knew exactly what the others were doing, and each new result was freely discussed. There was little attention paid to priority or to the source of new ideas or new interpretations. (Sturtevant 1965, p.295).[52]

Knorr Cetina also found this to be a feature of the high energy physics communities she studied during the late 1980's (Knorr Cetina 1999). The experiments conducted in this field require large numbers of scientists to work together. In these communities

> [i]ndividuating authorship conventions have disappeared; papers reporting experimental results will have all members of the collaboration listed on the first page(s) of the paper…. The names are in alphabetical order; no clues as to who originated the research or performed large chunks of it can be derived from the list. (Knorr Cetina 1999, p.167)

[52] Although this may have been the norm, both Kohler and Allen note that there were individual members of the Drosophila community, like Hermann Muller, who retained a strong desire for individual credit in publications (Allen 1978, p.205; Kohler 1994, p.105).

By not trying to figure out which individual should get credit for each part of the research, collaborative communities save themselves a lot of hassle and potential quarrels.

That said, this system will only work if junior members of the community can depend on senior members to provide them with alternative means for advancing their careers. In a community in which a graduate student, postdoc or junior scientist cannot count on having a publicized record of her specific contributions to the field, she depends more on letters of references and other opportunities to prove her worth. In the high energy physics community, junior members of the community received 'exposure' by being selected to give the presentations on behalf of the experiment at conferences (Knorr Cetina 1999, p.169). In Morgan's fly group, the junior members were highly dependent on either positions provided by Morgan or references from him. Thus, the willingness of junior members of a community to participate in the collaborative problem solving process of shop talk rests on their assumption that they will be fairly provided for by their superiors. Given the relative powerlessness of junior members, it is hard to explain this assumption in terms of mere reliance. Instead, junior members in collaborative communities trust senior members to live up to their relational obligations as mentors and supervisors to provide for them.

In conclusion, communities which share materials and technology, technical gossip and shop talk can garner significant epistemic benefits. The reliability, power, speed, efficiency and fertility of research can be enhanced in sharing communities. Key features of these sharing practices rest on a foundation of trust. Therefore, trust helps provide epistemic benefits to scientific communities. At this point, one might be persuaded that trust can provide such benefits, but one might remain worried that trust can also undermine the reliability, power, speed, efficiency and fertility of research. I address this concern in the next chapter.

4.0 BALANCING TRUST

While the previous chapter argued that trust has positive epistemic consequences for scientific communities, this chapter considers the potential negative epistemic effects of trusting relationships. In response to the arguments of chapter three, one might object that on the whole, trust does more harm than good to the epistemic projects of science. In order to respond to this concern, this chapter presents and responds to three worries one might have about the role of trust in producing good science. First, one might worry that too much sharing will lead to harmful conformism within science. Second, concerns could be raised that trust makes science inappropriately value-laden. Third, one might be concerned that trust between scientists leads to epistemic laziness. Part of my response to these worries is that the epistemic success of science results, in part, from science's ability to balance competition and cooperation, trust and distrust, self-interest and other-interest. This lays the groundwork for a brief outline of some positive proposals for how to design scientific communities that maximize the epistemic benefits of trust while minimizing its drawbacks.

4.1 OBJECTION 1: SHARING HAS NEGATIVE EPISTEMIC CONSEQUENCES

While I have argued that sharing increases the fertility of research in a community, one might object that sharing could have the opposite effect, on the grounds that sharing can lead to conformism. If everyone in a research community uses the same materials and technology, then the range of research questions pursued may be smaller than it would if a wider range of materials and technology were being used.[53] To use the example of stocks of model organisms, if everyone in the developmental genetics field is working with fruit flies, then the community suffers from lower fertility due to the absence of use of (e.g.) mice to study mammalian development. The critic may accept the arguments made in chapter three showing that sharing can have a positive effect on fertility, but nonetheless she may argue that that these benefits may be overshadowed by the negative consequences of conformist sharing. Thus, one might object that to the extent that trust fosters sharing, it can have a negative effect on the fertility of research. In addition, this kind of conformism could hurt decrease reliability if a community settles on sharing a flawed type of material or technology.

This objection is correct in pointing out that sharing *could* decrease fertility and reliability. However, the kind of incentives described by social epistemologies of self-interest will tend to work against trends towards strong conformism. As Kitcher (1993) argues, self-interested scientists are motivated by a desire to attain the prize of being the first to make a discovery. Kitcher begins with the simplifying assumption that the credit for a discovery is shared equally between the scientists who make it. If many of the community members are all using the same method for investigating a problem, then the credit that any individual could hope to receive would be so diluted that it is prudent for an individual to switch methods in the hopes of gaining more of the credit for herself. This incentive to switch methods can be present even if the probability of success of the new method is lower (Kitcher 1993, p.350). Thus, self-interest can prevent scientists from being too conformist. In this way, there can still be a variety of

[53] This argument can also be made in terms of sharing of technical gossip and shop talk.

materials and technology used within a community even if sharing occurs because scientists have an incentive to not use the same materials as everyone else. Thus, in an actual scientific community in which scientists are motivated by both self-interest and other-interest, sharing need not lead to conformism and a decrease in fertility and reliability. I return to this theme of a balance between self-interest and other-interest at the end of this chapter.

4.2 OBJECTION 2: TRUST MAKES SCIENCE INAPPROPRIATELY VALUE-LADEN

A second objection to the claim that trust has epistemic benefits for scientific communities rests on the observation that we tend to trust people who are like us. It is an unfortunate aspect of the human condition that we tend to be suspicious of people different from us and feel more trusting towards people who share our background. In some ways this makes sense. Recall from chapter two that in evaluating the trustworthiness of others, we often look for evidence that the trusted shares our values and recognizes the same norms as we do. This makes sense because one wants to know whether one can count on the trusted to live up to the relational obligations that one thinks are part of our relationship. However, we often take this sensible approach too far when we rely on stereotypes and prejudice as guides for whether someone shares our values. Thus, whether or not one takes a trusting attitude towards someone can be affected, consciously and/or unconsciously, by such factors as the person’s race, ethnicity, gender and socioeconomic background, among others. Because of this tendency, the critic, who sees no reason to think that scientists are any different from everyone else in this regard,[54] argues that to the extent that it is based on trust, science will be subjective and value-laden.

This objection is pressing only if it can be shown that the role of trust in science brings what Longino labels ‘contextual values’ into the context of justification. Those who defend the claim that science is, and ought to be, value-free are not concerned by the presence of what Longino calls ‘constitutive values’. These are values like accuracy, simplicity, truth, repeatability, unification, which “are the source of the rules determining what constitutes acceptable scientific practice or scientific method” (Longino 1990, p.4). The role of constitutive values in science is uncontroversial. What motivates the values-in-science debate is a concern that ‘contextual values’ play a role in determining which hypotheses, theories, explanations etc. are accepted as having met the appropriate epistemic standards. Contextual values are “[t]he personal, social, and cultural values, those group or individual preferences about what ought to be” which “belong to the social and cultural environment in which science is done” (Longino 1990, p.4). These are the values which have traditionally been thought antithetical to objective science. Even those who resist the claim that science is value-laden agree that contextual values can play a role in the context of discovery in which hypotheses are generated. The controversial claim is that contextual values play a role in the context of justification—that contextual values help determine which hypotheses are accepted as having met the standards of justification. Recall Merton’s principle of universalism, according to which “[t]he acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonists; their race, nationality, religion, class, and personal qualities are as such irrelevant.

[54] Allen argues that some of the distrust between T.H. Morgan and H.J. Muller had some basis in their very different backgrounds. Morgan came from a wealthy Southern family, while Muller came from a working class, immigrant background (Allen 1978, p.207).

Objectivity precludes particularism" (Merton 1942/1996, p.269). Therefore, if this second objection is to have teeth, it needs to show that trust opens the door for contextual values in the context of justification.

Longino's analysis of the need for transformative criticism provides the basis for such a pressing objection. Longino draws on underdetermination arguments to show that there is a gap between hypotheses and the data taken to support them (and also between our experience and the data that we extract from it). She describes the gap between data and hypotheses as follows:

> Data—even as represented in descriptions of observations and experimental results—do not on their own, however, indicate that for which they can serve as evidence. Hypotheses, on the other hand, are or consist of statements whose content always exceeds that of the statements describing the observational data. There is, thus, a logical gap between data and hypotheses…. [I]n the interesting cases of scientific reasoning, for example, that concerning the characterization of and relations among subatomic particles, hypotheses contain (as essential components) expressions not occurring in the description of the observations and experimental results serving as evidence for them. (Longino 1990, pp.58-9)[55]

Background assumptions are, for Longino, what fill this justificatory gap. Background assumptions "determine what states of affairs count as evidence for a hypothesis" (Longino 1990, pp.56-57). They can include assumptions about how the expressions occurring in the hypotheses are connected to the expressions occurring in the descriptions of the data, as well as general methodological principles like the correctness of enumerative induction. Longino argues that background assumptions are justified by having survived a process of critical scrutiny from a variety of perspectives. This variety of perspectives needs to include input from a variety of scientists from different racial, ethnic, gender, socioeconomic backgrounds because, as research into sexism and racism in science has shown, background assumptions can be based on contextual values (e.g. Fausto-Sterling 1985; Gould 1981; Lewontin, Rose, and Kamin 1984;

[55] For her discussion of the gap between experience and data see (Longino 2002, pp.99-103).

Lloyd 1993 and 2005; Longino 1990). Thus, in order to ensure that background assumptions, which provide the framework for justification in science, do not simply reflect the views and interests of one segment of society, it is necessary that a true diversity of scientists be included in the process of critical scrutiny.

The pressing version of this objection can now be presented. In chapter three I argued that trust has epistemic benefits because it encourages sharing of materials and technology, technical gossip and shop talk. I argued that such sharing increases the reliability of research because it enables more scientists to be part of the process of critically evaluating the materials and technology used and the ideas circulated. This includes critical scrutiny of background assumptions. Given Longino's arguments for the need to subject background assumptions to critical scrutiny, it would seem that she would appreciate the role of trust in fostering more critical scrutiny. However, sheer quantity of critical scrutiny is not enough. In order for critical scrutiny to uncover questionable background assumptions based on contextual values, those involved in the critical process need to come from a variety of backgrounds. But if scientists share, collaborate, and discuss their work only or primarily with members of their social circle, the kind of diverse criticism that Longino promotes may not take place. Therefore, to the extent that people trust others who are like them, the sharing encouraged by trust will be ineffective at uncovering assumptions based on contextual values. In this way, one might object, trust undermines the objectivity of science.

In responding to this objection, I will not deny that trust has the potential to undermine objectivity. Scientific communities that are 'old boys' clubs' that exclude women and minorities are likely to share in ways that do not subject contextually based background assumptions to scrutiny. In societies in which members of particular groups are treated with suspicion, they are more likely to be excluded from sharing networks, and thus some assumptions may go unquestioned. Therefore, this objection is right to point out that trust *may* undermine the questioning of problematic contextual values.

However, it would be wrong to conclude that trust *must* have this negative epistemic consequence. While we do have a tendency to trust people like us, it is by no means the case that we are incapable of trusting people who are very different from us. This suggests that the correct response to this objection is not to reject the arguments presented in chapter three for a positive epistemic role for trust in science. The correct response is to argue that an epistemically well designed scientific community needs more than just trust; it needs to encourage diversity and personal relationships between individuals of different backgrounds. In order for sharing to increase reliability by facilitating the kind of transformative criticism Longino describes, scientists need to create diverse sharing networks. This is one way that an epistemology of other-interest can be supplemented by Longino's 'contextual empiricist' social epistemology.

4.3 OBJECTION 3: TRUST ENCOURAGES EPISTEMIC LAZINESS

A third objection to the claim that trusting relationships have significant epistemic value rests on the worry that trust encourages epistemic laziness. The concern is that trust could be used as an excuse by negligent scientists who give unwarranted credence to what their colleagues tell them. This could be one means by which trust has a negative epistemic impact on the community. For instance, a concern one might have about the role of technical gossip and shop talk in science is that if scientists are too gullible in believing what their colleagues tell them, misinformation can spread throughout the community. This might happen through the unintentional dispersal of inaccurate information or through the malicious spreading of misinformation. Someone who objects to the arguments of chapter three might point out that information about results spread through the informal communication of technical gossip and shop talk does not go through the rigorous peer-review process, which makes it more likely to contain inaccuracies. In addition, one might worry that trust in one's colleagues prevents scientists from recognizing misconduct. Thus, epistemically lazy scientists who trust their colleagues allow fraud to go on under their noses.

The critic points to instances of trust being used as an excuse for ignorance and inattentiveness to one's responsibilities. For example, CEOs of large corporations have tried to evade charges of malfeasance by claiming that they should not be expected to have done more to prevent wrongdoing because they trusted their subordinates (Pasha 2006). The CEOs maintain that they were ignorant of any wrongdoing on the part of their subordinates and simply trusted them to follow the law. In a more commonplace context, the critic points to the mother who says she trusts her partner and is so absent from her family that she fails to see evidence that her children are being abused. In both of these cases, trust is used as an excuse for epistemically blameworthy behavior. Thus, one might be concerned that endorsing trust between scientists could lead to similar epistemically damaging behaviors.

To appreciate why this objection is pressing, one must recognize that justified trust in someone does involve letting one's guard down to some extent (Baier 1994, p.139). Part of the reason why trust makes us vulnerable is that when one trusts, one lets down one's guard a little and does not engage in constant monitoring of the trusted to make sure she is acting as expected. For example, suppose that I know that my colleague is particularly conscientious and considers it her duty to her colleagues to double-check her work, and I know of no reason why she would be prevented from doing so or motivated not to do so in this particular instance. In this case, it seems that I am justified in trusting her to provide accurate results, and I am not obliged to check and double-check her work myself. My justified trust in my colleague absolves me of the responsibility to engage in burdensome monitoring. Thus, objection three can be interpreted as asking the following questions: If trust absolves one of some of the responsibility to monitor, how is trust not just epistemically lazy? Why is justified trust not just an excuse to be lazy?

My response to this objection is that while justified trust does absolve one of *burdensome* monitoring, it still requires an adequate amount of monitoring, and so it is not epistemically lazy. In order for *A*'s trust in *B* to do *C* to be justified, A must have some reason to assume that *B* will do *C*. In cases of ongoing trust, when *A* trusts *B* to do *C* over a period of time (e.g. I trust my colleague to work conscientiously throughout our project), *A* has an epistemic responsibility to check periodically whether her reasons for trusting *B* to do *C* still hold.[56] Thus, the carelessness of the CEO and the mother are blameworthy even if they did trust the wrongdoers. They ought to be faulted for misplacing their trust and also for being neglectful in their responsibilities. The CEO has a responsibility to remain well-informed about the dealings of her corporation and to ensure that it pursues no illegal or unethical actions. The mother has a responsibility to determine whether her children are safe. Neither the CEO nor the mother can abdicate this responsibility to someone else, even if they trust that person.[57] Trust does not justify epistemic negligence.

[56] Recall from chapter two that *A*'s reasons for trusting *B* to do *C* can be evidential or pragmatic. In cases when *A* believes that *B* will do *C*, *A*'s belief is justified by evidence that *B* will do *C*. When *A* accepts that *B* will do *C* for pragmatic reasons, *A*'s acceptance is justified by evidence that trusting *B* to do *C* will, in fact, accomplish the pragmatic goals at issue. Thus, in both types of trust, the truster has an epistemic responsibility to discharge.

[57] This is not to say that there cannot be extraordinary circumstances in which we do not hold people to their responsibilities. For example, suppose the mother is drugged by her partner and cannot exercise due care to make sure her children are safe. However, in situations like these, even the critic of trust will agree that the mother was not lazy.

This is also the case in science where trust in a colleague does not absolve one of the responsibility to exercise due diligence to make sure that one's trust is justified. When one engages in technical gossip or shop talk, one has a responsibility to recognize that, as informal communication, it has not passed through the peer review and replication process. Thus, one has an epistemic responsibility to make sure that one has good reason to trust what a fellow scientist says. Similarly when one trusts one's colleague to provide solid research, one has a responsibility to remain reasonably well-informed about her activities. One's trust absolves one of the need to obtain independent verification of everything one is told and of the need to check up constantly on one's partners in research, but that does not mean that one's trust justifies laziness.

The threat of scientists giving undue credence to their colleagues' work and the need for scientists to exercise due diligence is illustrated in the controversy surrounding the South Korean stem cell researcher Hwang Woo Suk. Hwang gained international fame for his 2004 paper in *Science*, in which he claimed to have succeeded in cloning human embryos and extracting stem cells from cloned embryos. In 2005 he again published a breakthrough paper in *Science*; this time he reported that his team at Seoul National University had developed embryonic stem cell lines tailored to specific patients. With the publication of these results, Hwang staked his claim for priority in some of the most sought after breakthroughs in stem cell research.

Unfortunately, reports began to surface in 2005 that undermined Hwang's credibility. First, it was discovered that Hwang had used ethically suspect means to collect the eggs needed for his research. Eggs were donated by two of his junior researchers as well as some South Korean women who were paid $1,400 each ("Korea's cloning crisis" 2005). To solicit eggs from these sources is considered morally dubious, since it is questionable whether subordinates freely choose to donate their eggs, and the potential for exploitation of the poor exists when women are paid large sums of money for eggs. Concerns about Hwang's ethics increased when it was discovered that he had lied to cover up the source of the eggs.

It was hoped that Hwang's misdeeds were only related to the issue of egg sources. For example, on December 4th 2005 the New York Times published the following editorial comments:

> South Korea's high-flying stem cell researchers—reputedly the best in the world at cloning—have stumbled badly in handling the ethical issues of their controversial craft. Worse yet, the research team's leader, a national hero in his homeland, lied in an effort to hide his ethical lapses. We can only hope that he has not also lied about the astonishing scientific achievements of his research team….
>
> The key unresolved issue is whether lying about egg donations suggests that the Korean team may have lied about its scientific results. So far there is no evidence of that. Indeed, American collaborators and observers remain confident that the team's achievements were real. ("Korea's cloning crisis" 2005)

But only a few weeks later, these hopes were dashed as a university panel found that Hwang had forged much of the data in the 2004 and 2005 papers (Chong 2005). Even though the results of another one of Hwang's papers—this one reporting the first cloning of a dog—were authenticated, Hwang's reputation was ruined, and he resigned from his position at Seoul National University.

The scientific community's response to Hwang's deception illustrates Baier's claim that we recognize the pervasiveness of trusting relationships when the atmosphere of trust has been polluted (Baier 1994, p.98). When the fraud was uncovered, scientists were reported in the press as acknowledging their vulnerability to being deceived by their colleagues. For example, stem cell researcher Peter Andrews acknowledged the limits of the process of replication. "In the end, the progress of science depends on results being repeated in independent labs, but along the way we have to work by trusting our colleagues," he said. "It comes as a shock when occasionally we find that someone has betrayed that trust" (qtd. in Chong 2005). Another distinguished stem cell researcher, Irving Weissman, responded to speculation that Hwang's work was sabotaged by other members of his lab by noting how vulnerable senior scientists are to fraud on the part of people working for them:

> I've told people in my lab many times -- there is no doubt that you can fool me. It's a matter of trust between us that you are not doing that…You can be taken in by a very clever and overeager person whether they are a professor or an assistant professor or a trainee. (qtd. in Fox 2006)

Thus, the case of the South Korean stem cell controversy illustrates the extent to which scientists are vulnerable to having put their trust in the wrong colleagues.

This case also illustrates the ways that such ill-placed trust can cause personal harm for those involved and negative epistemic consequences for the community as a whole. Hwang's fraud caused considerable harm to many people and their scientific work. His laboratory was disrupted and doubt cast on the accuracy of the work of all the scientists involved. Junior scientists and other scientists who worked under him found their credibility undermined. Some of their loss of credibility may be deserved since it has been suggested that some of them aided Hwang in his deceit, or even, as some suggest, sabotaged his work. However, the authentication of their work on cloning the dog suggests that his lab was doing some legitimate scientific work, which has now been disrupted. A cloud of suspicion has also been cast over the South Korean scientific community who had held Hwang up as a national hero. Hwang's international collaborators have also suffered. The role that Gerald Schatten of the University of Pittsburgh played in getting Hwang's work published has received much scrutiny. While it does not appear that Schatten knowingly participated in Hwang's fabrication of results, he did help get the work published and appears as a senior author on one of the papers. Schatten's reputation has been tarnished by his association with Hwang, and also by his and Hwang's failure to disclose fully the conflict of interest involved in their application for patents on the technology reported in the publications ("Cloning Scam" 2006). The reputation of the field of stem cell research and, to a lesser degree, science as a whole has also suffered as a result of Hwang's untrustworthiness as public trust in the reliability of scientific results has been shaken.

Hwang's fraud also had negative epistemic consequences for the scientific community. Most obviously, Hwang's untrustworthiness negatively affects the *reliability* of the community's results since it introduces falsehoods into the scientific record. The harm to reliability can mushroom as other scientists make inaccurate predictions and ineffective interventions based on fabricated research. The acceptance of falsehoods into the scientific record has other significant indirect harms. As social epistemologists of self-interest note, there are incentives for scientists to try to be the first to make a breakthrough. Once it is reported that the breakthrough has been made by someone else, that particular incentive is gone for other scientists to keep trying. Thus, promising lines of research may be dropped, which negatively affects both the *speed* and *efficiency* of the community. Fabrication by untrustworthy scientists can slow the progress of science as the lines of research that would actually have led to the breakthrough are abandoned. The cost of actually making the breakthrough are also increased, harming efficiency, since it is costly both to change one's line of research once the fabricated discovery is published and also to change it again once the fabrication has been uncovered (Poling 1992, p.145). Fraud of this kind also negatively affects the efficiency of science because costly investigations must be undertaken to uncover the extent to which those associated with the dishonest scientists are themselves dishonest and thus deserving of diminished reputation. Similarly, the speed of scientific progress is slowed as these researchers must divert their time away from their work in order to defend their reputation. Thus fraud can make science less reliable, less efficient and slower.

Gerald Schatten's role in the South Korean stem cell controversy illustrates the role that trust can play in causing these negative epistemic consequences. Once the fraud was discovered, Schatten was criticized for helping the fraudulent papers get published and for putting his name on one of them as a senior author without being involved enough in the actual work to be confident that the results were genuine:

> Stem cell pioneer John Gearhart said Schatten's decision to accept the responsibility as senior author of the Korean paper and promote the research without overseeing the lab work was the Pitt scientist's biggest mistake. "If you are a senior author, you have to do more than just accept the integrity of the scientific process that you believe is going on," said Gearhart, a professor of medicine at The John Hopkins University's McKusick-Nathans Institute of Genetic Medicine in Baltimore. "You have to be part and parcel to it." ("Cloning scam" 2006)

In statements, Schatten has indicated that he trusted Hwang,[58] but he has not attempted to use this trust as an excuse. However, were he to attempt to do so, the quote from Gearhart cites the reason why such an excuse would carry little weight. While Schatten's trust in Hwang does absolve him of a responsibility to engage in constant, burdensome monitoring of Hwang, trust does not entitle a scientist to negligence. Thus, Schatten had an epistemic responsibility to make some effort to "do more than just accept the integrity of the scientific process."

The primary way trusters can discharge this epistemic responsibility and avoid epistemic laziness is by engaging in what I call 'participatory trust'. Participatory trust has three main components: 1) getting to know the trusted well, 2) participating with the trusted in the activities one is trusting her to perform, and 3) taking steps to address some of the structural features of one's relationship with the trusted that can undermine one's epistemic position. As I will show, judicious performance of these activities is epistemically praiseworthy.

[58] "In a written statement Friday, Schatten said he was quitting his 20-month relationship with the South Korea group after 'a breach of trust about possible egg-donor recruitment irregularities'" (Hall 2005).

Some of the problems with the negligent trusters like the mother and CEO can be solved by their getting to know the trusted better. The first component of participatory trust is getting to know the trusted better so that one can learn things like whether she recognizes relational obligations and finds them motivating, or whether there are any other aspects of her character or circumstances that might prevent her from acting trustworthily. Mothers who get to know their partners well before they expose their children to them are more likely to uncover character traits (e.g. violent temper) that would put the children at risk. Of course, mothers can be the victims of manipulative individuals who hide their true characters from them. So getting to know the trusted well is often not enough to justify continued trust.

The second part of participatory trust can help mothers and other trusters with this problem—by participating in the activities that the partner pursues with her children, she has a chance to see just how he interacts with them. By taking an active role in the family and the partner's role in it, the mother will gain information relevant to protecting her children.[59] By actively participating in family life, she gives herself the opportunity to discover evidence that falsifies her assumption that the partner can be trusted. For example, by dressing the children, she may notice bruises. Similarly, CEOs who get to know their immediate subordinates well and who are active participants in the affairs of their business are more likely both to know who can be trusted and also detect wrongdoings. Such CEOs can still be deceived, but they will have lived up to their epistemic responsibilities. Notice that participatory trust is less a matter of making spot checks and surprise visits to check up on the trusted than it is a matter of being involved on a regular, normal basis. Spot checks and surprise visits can be very useful in

[59] Certainly, there will still be unfortunate cases where the partner is so deceptive and manipulative that even an active mother is prevented from ever getting evidence of abuse, but in these cases we are not likely to see the mother's actions as blameworthy—she did all she could be reasonably expected to do to make sure her trust was well placed.

uncovering evidence of trustworthiness, but they have a tendency to make the visitor look suspicious. Thus, ideally, the responsible truster makes an effort to be regularly involved.

A third aspect of participatory trust involves taking steps to address some of the structural features of one's relationship with the trusted that can undermine one's epistemic position. One lesson of standpoint epistemology is that one's position in power structures and institutions can affect the evidence available to one.[60] Information does not flow equally among all levels in a hierarchy. Individuals in positions of power need to take steps, like having an 'open door' policy and encouraging organizational dissent, to make sure that subordinates do not hide problems from them (Harris et al. 2005, p.42, p.207). Being approachable is epistemically valuable to mothers and CEOs alike because they can learn useful information about people they trust if their children and employees feel comfortable coming to them. People in positions of relative powerlessness need to make an effort to find out what kind of information is typically hidden from them and, when necessary, to try to gain access to that information to which they are not ordinarily privy. This is not to say that trusters are required to escape the limitations of their epistemic location. That would be requiring too much. All that is required is that trusters take steps to recognize the limitations and do what is possible and reasonable to overcome them.

Participatory trust should also be expected of scientists. One natural interpretation of Gearhart's comment that "you have to do more than just accept the integrity of the scientific process that you believe is going on…. You have to be part and parcel to it" is that he is criticizing Schatten for not participating enough in the research. Scientists who are part of a collaboration can live up to their epistemic responsibilities by getting to know their partners well

[60] For a useful summary of feminist standpoint epistemology, see Harding (1991). Many standpoint epistemologists want to draw stronger conclusions about the privileged epistemic location of oppressed groups (e.g. Harding 1991, p.119). By recognizing one of the insights of standpoint epistemology, I do not endorse these stronger claims, which are not directly relevant to the point I make here.

and by participating in the work as much as is reasonable. Getting to know one's collaborator well involves knowing things like how committed she is to professional standards of ethical conduct, how conscientious she has been in the past, how desperate she is to find results that will improve her reputation, and how often she tends to cut corners in ways that threaten the integrity of the research. These are things that one can find out in part by talking to one's partner and the partner's past colleagues, but they are also things one can learn best by seeing one's collaborator in action, and this requires active participation in the work. This can include being in the lab when experiments are performed, being party to conversations where the results are analyzed and witnessing what changes are made to experimental design or analytical methodology.

Of course, there will be practical constraints that will limit scientists' ability to participate in the work of their colleagues. For example, international collaborations will face the challenges of distance and expense of travel. Such limitations do not make responsible trust impossible. In these circumstances, scientists can still actively participate in the work of their colleagues by using teleconferencing and other communication technologies to join the conversations in which experiments are designed, problems are addressed and data is analyzed. They can also, to the extent possible, make the effort to visit their partner's labs occasionally. All that participatory trust requires is that scientists make a reasonable effort to be part of the process of their partners' research. Finally, issues of power and epistemic location are important for scientists to take into account. Open door policies, encouraging dissent and understanding what information is being kept from one by more powerful members of the community are all policies that scientists would be well advised to follow.

Knorr Cetina's account of the role of the lead scientists in high energy physics provides a useful illustration of active participation on the part of senior scientists. As was discussed in chapter three, high energy physics is a field characterized by very large collaborations. Knorr Cetina describes the lead scientists' place within the collaboration as follows: "Graphically speaking, leaders were not the "top" of an experiment, not its "spearhead" pointing forward, but were centrally placed within it. Above all, they were *centrally located in the conversation conducted within the collaboration*" (Knorr Cetina 1999, p.180). The leader is the 'spokesperson' for, 'secretary general' of, and 'true participant' in the experiment (Knorr Cetina 1999, p.180). As spokesperson, the leader interacts with the community outside the experiment by attending conferences and interacting with institutional management. As secretary general, the lead scientist "is someone who gathers information and relays it—someone who functions as an intellectual administrator, who 'handles', knows, passes on matters, and can be contacted about them" (Knorr Cetina 1999, p.180). Finally as true participants, the leaders

> …took care, with co-workers, of parts of the equipment and their functioning; they were present not only at meetings that concerned them as leaders but also at relevant technical meetings; and they worked closely with some postdocs and younger collaborators. (Knorr Cetina 1999, p.180)

These lead scientists, who occupy this central position within the experiment, are active participants in the experiment. They may trust the postdocs and younger collaborators who are part of the experiment, but this trust is hardly passive or lazy. It is in this way that trusting scientists can live up to the epistemic obligations that are part of working in collaboration.[61]

A responsible truster will engage in some or all of these activities to a reasonable extent. Of course, too much or heavy handed practice of these activities will be taken by the trusted as a sign of distrust. Thus, epistemically responsible trust is a matter of balancing the need for

[61] One might think that participatory trust is simply common sense and must be already widely practiced in science, but this is not necessarily the case. Poling (1992) reports his experiences working with a collaborator who was later found to have fabricated his results. Poling reports that in the case of one paper, he visited the study site but did not observe experimental sessions, and in another paper he did not visit either of the two study sites while his colleague was supposedly collecting data (Poling 1992, p.143).

verifying that one's trust is well placed with the need to truly trust by refraining from excessive monitoring. There is no magical formula for determining how much of each kind of participatory activity one needs to perform in each instance of trust. Common sense, trial and error and learning from other participatory trusters are all useful guides in striking the right balance.

It is important to remember that the kind of trust discussed here as participatory trust is not the kind of trust described by the rational choice approach. The purpose of getting to know the trusted is not to learn what the trusted values so that one can provide incentives for trustworthy behavior. The reason for active participation is not to use one's regular presence as a deterrent that makes it in the trusted's self-interest to act as expected, although this may be a side-benefit. Nor is the purpose to be present to detect untrustworthy behavior so that it can be punished, although this may be another bonus. The purpose of these activities is to determine whether one's *trust* is justified. Were one to engage in the activities that support participatory trust simply in order to obtain these side-benefits, one would be *merely relying* on the individual. That one can acquire these side-benefits from these activities shows that they are also helpful to epistemically responsible mere reliers. Thus, there may also be participatory reliance, but it is important not to confuse it with participatory trust.

Interestingly, although participatory reliance is a possibility, it is not frequently encountered. Engaging in participatory activities tends to lead us to adopt an attitude of trust rather than reliance. This is because trust is morally inflected in the way outlined by Strawson. Recall from chapter two that reactive emotions are those emotions to which humans are naturally subject as a result of their inevitable involvement, or participation, in interpersonal relationships (Wallace 1994, pp 31-2). When we participate in relationships we tend to become susceptible to the reactive emotions. Thus, someone who actively participates in a relationship, when she is working on the assumption that the partner will do something, will tend to trust, rather than rely on, the other. This is why despite it being possible for mere reliers to discharge their epistemic responsibilities by engaging in participatory reliance, they tend instead to adopt another method of verifying that their reliance is justified. I call this other method the 'audit' method of verification.

Audit verification is importantly different from participatory verification. Whereas people who follow the participatory method meet their epistemic burdens by being part of the action and observing how the trusted/relied upon person acts, followers of the audit method stand outside the action and meet their epistemic burdens by requiring the trusted/relied upon to submit evidence that her actions meet specific standards. Take an example from the world of education. A principal of a school has an obligation to ensure that her students are being taught well. Even though her trust (or reliance) may absolve her of a duty to monitor every move of the teachers, she still has an epistemic burden to find out how well her teachers are performing. If she follows the audit method, then she may try to live up to her epistemic responsibility by instituting a series of student tests to measure how the students of various teachers are progressing. The teachers are informed that their performance will be measured by how well their students' test results meet the standards appropriate for their age. The students take the tests at several points in the year, and the principal looks to see if the results improve acceptably, thereby discharging her epistemic responsibility to know how the students are progressing. Through this method, the principal can meet her responsibility without ever setting foot inside the classroom.

Consider another option available to the principal: the participatory method. If she wants to know how the students are doing, she can get out of her office and into the classroom. She can co-teach a class with each teacher to learn how that teacher works and observe how the students respond. She can work with the teachers as they prepare their classes, grade some assignments and meet with parents. If the principal is diligent about participating in the teaching of her schools' students, she can hardly be said to be epistemically lazy. So in following the participatory method, she shirks no epistemic responsibilities.

Both methods have their advantages and disadvantages. One might argue that the audit method has the advantage of being less costly. It is true that there are significant costs to the principal's time for taking the participatory approach. Participating in all the various aspects of teaching could be very time-consuming. However, it is not obvious that the principal would save time by creating and administering the student tests herself. Of course, she could simply read the reports provided by a group that has been hired from outside to create and administer the student tests. However, outsourcing the testing process would be financially costly, so it is not clear that the participatory method would be more costly overall than auditing. There is more merit to the argument that the participatory method is, in some way, less objective[62] than the audit method. When the principal enters the classroom, she has only her own observational and analytical powers to rely on in assessing the teachers' performance. This means that she can bring her own biases and blinders into her assessment. In contrast, if she were to outsource the testing to a group of people, the assessment process would be subject to a process of criticism that could weed out individual biases. There is no way to outsource participation. Therefore, the participatory method does not have the benefit of using objective measurements and standards. However, the audit method has other problems, which the participatory method avoids. These

[62] 'Objective' is here being used in Longino's sense of being subjected to intersubjective critical scrutiny (Longino 1990).

problems are particularly pressing when the audit method is used, as it frequently is, in conjunction with mere reliance.

Audit reliance can undermine the goals of an enterprise by distracting people from their true goals and by providing perverse incentives. Onora O'Neill discuses these problems in her criticism of what she calls the 'culture of accountability' (O'Neill 2002). In the United Kingdom during the 1990s, reports of a crisis of public trust in public institutions like the educational and health systems led to calls for increased accountability. New legislation and regulations were put in place to make precise measurements of performance in these sectors and sanction schools and hospitals that fall short of the performance indicators. Essentially, legislators and regulators responded to the perceived crisis of trust by trying to improve the effectiveness of their reliance on the self-interest of teachers and health care professionals. They hoped that if the punishment for poor performance were stiff enough and auditing mechanisms for detecting poor performance were effective, performance would improve. This would cause public confidence in the system to return. O'Neill argues that this culture of accountability distorts the "proper aims of professional practice" (O'Neill 2002, p.50). This happens because the auditing process distracts professionals from the work they are really supposed to be doing:

> Much professional practice used to centre on interaction with those whom professionals serve: patients and pupils, students and families in need. Now there is less time to do this because everyone has to record the details of what they do and compile the evidence to protect themselves against the possibility not only of plausible, but of far-fetched complaints. (O'Neill 2002, p.50)

Under a culture of accountability, teachers spend more time administering standardized tests and less time actually teaching, and doctors spend more time filling out paperwork with less time remaining to address the needs of their patients. The concern is that this system of audit reliance may in fact decrease the quality of work because it decreases the amount of time that people have to do the work.

The problem of perverse incentives is a slightly different way that audit reliance can be self-defeating. Even the best performance indicators, the standards by which good performance is measured and verified, are chosen because they are “at the very best” surrogates for the real objectives of the activity measured (O’Neill 2002, p.55). Good test performance is not the real objective of education, but if the test is designed well, then it is a good substitute for the ultimate objectives. One of the reasons why it is a good substitute is that it is easily operationalized and measured (O’Neill 2002, p.54). Thus, by taking the audit reliance approach, we end up holding people accountable to surrogates for the real objectives of their work. Unfortunately, sometimes the easiest way to meet a surrogate goal turns out to undermine the real goals. This is problematic if, as good mere reliers, we are counting on people to act out of self-interested motivations. Since people will be rewarded or punished based on their ability to meet the performance indicators, a system of audit reliance creates incentives for people to act in ways that undermine the real goals of their work; in other words, audit reliance can create perverse incentives. O’Neill describes this phenomenon as follows:

> …the performance indicators have a deep effect on professional and institutional behaviour. If a certain [testing] board offers easier examinations in a subject, schools have reason to choose that syllabus *even if it is educationally inferior*. If waiting lists can be reduced faster by concentrating on certain medical procedures, hospitals have reason so to do, even if medical priorities differ. *Perverse incentives are real incentives.* (O’Neill 2002, p.55)

By holding people accountable to measurable standards, audit reliance may have the perverse effect of encouraging people to act in ways that undermine the goals one is trying to ensure are met. Even though verification by such standards has the epistemic benefit of being somewhat more objective than participatory trust, it can also be self-defeating by distracting individuals from the real purpose of their work and even providing perverse incentives.

Distraction and perverse incentives are not just problems for audit reliance in education and health care; they are problems facing an audit reliance approach to science. First, attempts to use a system of audit reliance to increase confidence in the validity of the predictions and interventions published in scientific journals threatens to distract scientists from their real work. The response to the crisis of trust in the public sector was a call for increased accountability. Similarly, revelations of scientific fraud are usually followed by calls for increased scrutiny of articles before they are published in journals. The South Korean stem cell controversy was no exception. That Hwang's fabricated results were published in such a reputable journal shocked many within and outside the scientific community. There were calls for independent testing of results before publication: "Journals are likely to ask for more supporting data on papers which purport to be major breakthroughs; and in some cases, scientists may even open up their work to independent experimental analysis before they submit a paper for publication" (Morelle 2006).

Such calls are not new: Daniel Freedman (1992) discusses proposals for editors to "[scrutinize] the bank accounts and laboratory notebooks of contributors" and for journals to "'randomly audit' laboratory practices (as does the FDA)" (Freedman 1992, p.185). Such independent experimental analysis and auditing might help journals uncover fraud before it is published, but it would also require scientists to take time away from their research to work with the independent testers. At the time of the revelation of Hwang's fraud, Donald Kennedy, Editor in Chief of *Science*, dismissed calls for independent replication of data by a third party, saying it was an unreasonable requirement for publication: "I think that to install a procedure by which replication by a third party was required for acceptance of a manuscript would impose simply an enormous load on our readership and the scientific community" (Rincon & Amos 2006). Auditing of results could thus potentially distract scientists. This would be self-defeating insofar as such distraction could negatively affect the speed, efficiency, power and fertility of the community. Distracted scientists would find their work on their research projects slower and more costly if they are required to submit to independent testing. The demands on their time and resources could also create practical barriers to their taking on new research projects, thus harming both power and fertility.

One might object that while extra auditing measures might harm the speed, efficiency, power and fertility of scientific research, audit reliance increases the reliability of research. If scientists are held to higher standards of evidence, then inaccurate predictions and ineffective interventions are more likely to be weeded out, which is a good thing. However, there is also a way that a culture of accountability in science can undermine reliability. Unsurprisingly, in the wake of scandals, such as the stem cell controversy, government regulators and scientific policy makers feel pressure to do something about the dramatic cases of falsification, fabrication and plagiarism (FFP) (De Vries et al. 2006). Thus, the audit reliance approach to science tends to focus on setting up procedures to detect FFP, standards to measure FFP and punishments for engaging in FFP. While these efforts may be effective in decreasing the prevalence of FFP, they might distract the scientific community from more significant causes of misconduct. In their 2006 empirical study of research misconduct, De Vries, Anderson and Martinson found a gap between the kinds of misconduct that receive a lot of attention from policy makers and the kinds of problematic behavior that scientists themselves view as most damaging to the integrity of science:

> It is particularly important to notice that when scientists talk about behaviors that compromise the integrity of their work, they do *not* focus on FFP; rather they

> mention more mundane (and more common) transgressions, and *they link these problems to the ambiguities and everyday demands of scientific research*. When policymakers limit their concern to the prevention of infrequently occurring cases of FFP, they overlook the many ways scientists compromise their work in an effort to accommodate to the way science is funded and scientists are trained. (De Vries et al. 2006, p.47)

In their survey and focus groups, De Vries et al. found that scientists are more concerned about the everyday decisions about how to deal with the gray areas of research. They worry about the tough decisions about when 'cleaning' one's data becomes 'cooking' the data (De Vries et al. 2006, p.47). They worry about sloppy record keeping and how to respond to results one cannot replicate. As one participant in their focus groups observes, spending time to make sure one does not engage in FFP can distract scientists from dealing with these other issues of scientific integrity:

> It's a question of over-commitment. These famous people are so busy, I think they are mostly ethical, they don't … violate FFP, but they don't sit for an hour and talk to their students about keeping a lab notebook. In fact they probably, you know, don't even look at the raw data, they just see the final figures and paper. (qtd. in De Vries et al. 2006, p.45)

If the scientists interviewed and surveyed by De Vries et al. are right, then it is the commonplace practices of cutting the corners and making questionable decisions in gray areas that affect the integrity of science more than the rare cases of FFP. This raises a real worry that an audit reliance approach could undermine the reliability of science. Holding scientists accountable to standards aimed at preventing FFP could distract them from dealing with the real problems that undermine reliability.

In addition, independent testing and auditing could provide perverse incentives to scientists. Social epistemologies of self-interest consider desire to publish as one of the primary motivations of self-interested scientists. If results must pass independent tests before publication, scientists have an incentive to pursue lines of research that can be easily tested independently. Thus, fertility could be harmed as research projects that are either difficult to test independently or slow and costly to test would be ignored in favor of projects whose results could be easily tested. Hence, audit reliance can have some negative epistemic consequences for science.

To summarize, one objection to the account I have provided of the positive epistemic benefits of trust in science is that trust may make scientists epistemically lazy. The proponent of this objection is concerned that trust may be used as an excuse by negligent scientists who give unwarranted credence to what their colleagues tell them. In response, I have argued that scientists have a responsibility not to uncritically accept everything their colleagues tell them, which includes information gained through technical gossip, shop talk and collaboration. By engaging in participatory trust, scientists can gain the kind of information they need to assess the validity of their colleague's work. They can do this by getting to know their colleagues well, participating in the research as it progresses, and being attentive to the weaknesses of their epistemic location. While participatory trust may leave some room for individual bias in evaluations of validity, it avoids the self-defeating problems of distraction and perverse incentives that plague systems of audit reliance. While this does not show that participatory trust is always a better approach than audit reliance, it does show that participatory trust is a viable alternative and that scientists who trust their colleagues need not be epistemically lazy. The analysis of the relative advantages and disadvantages of participatory trust and audit reliance suggests that the proper antidote to epistemic laziness is a combination of the two approaches.

4.4 EPISTEMIC COMMUNITIES: A BALANCING ACT

My responses to the previous objections refer to several epistemic benefits and costs of both trusting relationships between scientists and competitive relationships based on considerations of self-interest. This naturally again raises the question—should epistemologists adopt a social epistemology of self-interest or a social epistemology of trust? Should philosophers of science encourage scientists to rely on the self-interested motivations of their colleagues, or should they promote trust in the concern for others that motivates trustworthy collaborators?

This question presents a false dichotomy. We should not think of these two approaches to social epistemology as competitors. Rather they are complementary analyses. Both identify social practices (reliance and trust) that, when engaged in properly, confer epistemic benefits on the community. Each social practice also has the potential to weaken the epistemic position of the scientific community. These two approaches to social epistemology mirror each other. Some of the problems caused by too much reliance and too much self-interested behavior can be moderated by greater trust and concern for others. Similarly, some of the problems associated with too much trust can be mitigated by greater reliance on self-interest. Hence, the two approaches complement each other. A summary of the advantages and disadvantages of each approach will show how social epistemologies of trust and self-interest can balance one another.

As was argued in chapter three, trust confers several epistemic benefits on a scientific community. Trust facilitates sharing between scientists. Scientists who trust each other are more likely to share a number of key resources including materials and technology. Communities in which materials and technology are shared produce work that is more reliable and powerful through a faster, more fertile and more efficient process. Scientists who trust each other also share information, ideas and critical scrutiny though the practices of shop talk and technical gossip. These modes of communication also positively impact the reliability, power, speed, efficiency and fertility of the scientific community in which they operate.

Unfortunately, as the critics who present the objections discussed above argue, there are potential negative epistemic consequences to trusting relationships between scientists. First, the method of verification appropriate to trusters may be somewhat less objective than the method of verification favored by mere reliers. A scientist who fulfills her epistemic responsibility to check the validity of her partner’s work by engaging in participatory trust may overlook crucial pieces of information. When the scientist enters the lab, she has only her own observational and analytical powers to rely on in assessing the trustworthiness of her partner. This means that she can bring her own biases and blinders into her assessment. In contrast, if she were to outsource the evaluation to an institutional audit reliance process, the assessment process would be subject to a process of criticism that could weed out individual biases. Thus, a more suspicious scientist, who merely relied upon an objective system of audit reliance, might do better at picking up on the problems with her partner. Second, a community composed only of other-interested scientists who trust each other may have a tendency towards conformism. By sharing their materials and technology and ideas through technical gossip and shop talk, they may come to think alike and research the same questions.

Fortunately, a community of trusting scientists can overcome these drawbacks by also having a certain level of competitive incentives in place to appeal to the self-interested motivations of scientists. If scientists who engage in participatory trust also avail themselves of a reasonable[63] level of objective audit reliance, then they can both reap the epistemic benefits of working with people whose work they know well and also check their judgments of the validity of the work against the objective tests. Similarly, if, following Kitcher's advice, a community of trusters institutes incentives for being the first to make a discovery, then there is an incentive for scientists to branch out and use uncommon methods and pursue new research agendas. In these ways, some of the potential harms of trust can be overcome by sensible levels of mere reliance.

To turn to the other approach, social epistemologists of self-interest are right to say that self-interested behavior within a competitive environment has several epistemic benefits. First, once a practice of sharing has become part of a scientific community, self-interest can, as Rescher (1989) argues, provide incentives for individuals to reciprocate and keep the sharing going. Second, when scientists are merely relying on each other instead of trusting each other, they tend to use objective methods of verification to determine whether their colleagues are behaving properly. Third, as Kitcher and Railton argue, a community of self-interested individuals competing for priority will be a cognitively diverse one in which scientists are pursuing different research projects (Kitcher 1993, Railton 1994).

That said, self-interest and our reliance upon it also have negative epistemic consequences for a scientific community. First, self-interested competition tends to breed suspicion and undermine trust. Thus, the more that scientists view their colleagues as motivated by competitive instincts to garner reputation and funding, the less they trust each other. Thus, in trying to garner the epistemic benefits outlined by the social epistemologies of self-interest, we can miss out on the benefits of depending on the other-interest of scientists. Second, the role of

[63] Striking the right balance between reasonable levels of audit reliance and too much audit reliance is not an easy task since individuals have differing levels of sensitivity to audit reliance. Some individuals may feel distrusted by even a small amount of audit reliance, and one risks poisoning the relationship if one tries to implement an audit reliance process.

self-interest in motivating scientific fraud has numerous negative epistemic consequences. While the harms of fraud were discussed in section 4.3, it was not noted then that competition between self-interested individuals can be a significant cause of fraud (Judson 2004, p.146). The pressure to publish is often cited as one of the main causes of fraud (Judson 2004, p.146, Thelen & DiLorenzo 1992, pp.166-172). Thus, a system of competitive incentives for self-interested individuals can provide perverse incentives to fabricate results in order to increase one's reputation. One might think that the solution to the problems of fraud is increased auditing to detect fraud and increased punishments to deter fraud. However, as was shown in 4.3, such audit reliance can distract scientists from their aims and provide further perverse incentives—in this case incentives to pursue only the research projects that can be easily verified. In these ways the reliability of a competitive community can be harmed by incentives to engage in fraud and less egregious forms of scientific misconduct. A competitive community can also be slower, less efficient and less fertile if it tries to avoid fraud by merely relying on scientists to be motivated by self-interest to pass the tests put in place to verify the validity of their work.

Fortunately, trust and trustworthiness can counteract these negative effects of competition. In a competitive environment that provides perverse incentives to commit fraud, scientists are well-advised to work with trustworthy colleagues who are not motivated solely by self-interest. Trustworthiness thus decreases the incidence of fraud. Of course, it will not eliminate it entirely. Thus, scientists will still need to live up to their epistemic responsibilities to ensure that their partners are not engaging in fraud. The problems of verification can be addressed if scientists meet their epistemic burdens through a combination of participatory trust and moderate levels of audit reliance. When a scientist engages in participatory trust, she gains substantial amounts of information about the work of her partner. She, therefore, needs to rely less on methods of audit reliance. She may nonetheless still want to use some minimal level of audit reliance to check her judgments about the validity of her partner's work. But since the audit reliance will be minimal, it will not seriously impact the speed or efficiency of the work. Nor will it create overpowering perverse incentives that will decrease fertility. Therefore, the negative epistemic consequences of relying on self-interest can be mitigated by trust in concern for others. This means that social epistemologies of trust and self-interest complement each other.

Finally, as my response to objection two reveals, social epistemologists need to be attentive to issues of power and inclusion. A community which has stuck the right balance between competition and cooperation, reliance and trust, self-interest and other interest, will garner many epistemic gains in terms of reliability, power, speed, efficiency and fertility. However, if it is a homogenous community in which the critical abilities of oppressed and minority groups are not included and valued, background assumptions will go unquestioned and the science produced will reflect the values of the dominant group(s). To maintain the objectivity of science as a whole, scientists need to balance their natural tendency to trust people who share their values with openness to building trusting relationships across differences. Therefore, the social epistemology that emerges from my analysis describes a well-designed epistemic community as one in which inclusive, diverse practices and institutions balance fostering reliance on the self-interest of scientists with cultivating trust in their other-interest.

4.5 SOME PRACTICAL PROPOSALS

The social epistemologists' notion of a 'well-designed' scientific community is a useful fiction. It is fictional because there is no designer or group of designers who determine how scientific communities are organized and which practices will be observed by its members. But it is a useful fiction because it rightly identifies that one of the roles of social epistemology is to provide recommendations for how to improve the institutions and practices of science. While we cannot start from scratch and create the ideal epistemic community, we can make improvements so that our scientific communities better live up to the norms outlined by social epistemologies. In this final section, I outline two practical proposals that are a natural consequence of the social epistemology presented here.

First, epistemically beneficial trust can be fostered by encouraging socializing between scientists that is fully inclusive. Social interaction is not only important for initiating trusting relationships between scientists, it is also critical for ensuring that these relationships are based on justified, participatory trust. Therefore, scientific communities that provide opportunities for scientists to interact with each other both inside and outside the workplace foster trust and garner epistemic benefits. Scientific leaders (e.g. the leaders of labs, collaborative projects and research institutions) ought to create venues for the scientists they supervise to get to know each other. For example, T.H. Morgan is remembered for organizing group trips to Woods Hole that helped build personal relationships between the Drosophilists (Allen 1978, p.198). In addition, funding bodies can produce better science not only by providing resources for the work itself to be carried out, but also by funding conferences, workshops, seminars and other opportunities for scientists to get to know their colleagues in the field better. Funders who have a particular interest in reducing scientific fraud ought to not only require that their grantees have in place audit reliance mechanisms to deter misconduct, they ought also to provide travel funds for collaborators to visit each other. Of course, in order for interaction between scientists to foster transformative criticism, the avenues for socializing need to be fully inclusive of a diversity of scientists. This means that attention needs to be paid to the extent to which avenues of interaction exclude particular types of people. For example, exclusion of female scientists from socializing in male-dominated environments is often cited as one of the factors undermining the position of women in science. Therefore, recognizing that trust between scientists has epistemic consequences ought to prompt us to pay more attention to the context and dynamics of the environments in which scientists socialize.

Second, scientific communities ought to invest in the moral education of their members. Communities whose members are socialized to recognize moral obligations to their colleagues foster other-interested trustworthiness. Avenues for moral education include both formal ethics training and informal mentoring. Formal ethics training is frequently called for as an antidote to fraud (Freedman 1992, Vasgird 2007). While there is mixed data on the effectiveness of existing ethics training, there is some evidence to suggest that it can be effective when it is ongoing and integrated into the scientist's educational program (Anderson et al. 2007). Ethics training that includes discussion of relational obligations might help produce scientists who can be trusted. Similarly, informal training through mentoring can be helpful in reducing scientific misconduct.[64] Mentoring could also have epistemic benefits if mentors teach their students the importance of living up to one's relational obligations to colleagues. As has been shown, the

[64] Interestingly, Anderson et al. (2007) found that the kind of mentoring given makes a difference. They found that mentoring in ethics was correlated with a decrease in misconduct; however, mentoring in financial matters (e.g. how to obtain funding) and survival skills (e.g. how to build relationships) was correlated with an increase in certain kinds of misconduct (Anderson et al. 2007, p.856). This suggests that some research into the kind of mentoring that tends to increase other-interest in scientists would be useful.

community of Drosophilists garnered significant epistemic benefits from the role Morgan and other senior Drosophilists played in training their students in the rules of their community's moral economy. Finally, there may be some value in instructing scientists in the insights of social epistemology (including the social epistemology of trust), since it is scientists, rather than philosophers, who are primarily responsible for inculcating and maintaining the moral standards of the scientific community. Therefore, a social epistemology of trust brings attention to the epistemic value of moral education in science.

In conclusion, not only does paying attention to the role of trust in science reveal that trust and trustworthiness have epistemic benefits for scientific communities, but recognizing the negative consequences of trust in science reveals the value of a balanced approach to social epistemology. When social epistemologists recognize the role that both self-interest and other-interest play in science, a host of issues are raised which might otherwise be overlooked. Not only are epistemologists pressed to investigate the epistemic merits of practices like sharing, mentoring, and socializing; but we are also compelled to examine the epistemic significance of power relations and diversity.

In the introduction, I began by asking what ethics can contribute to the epistemology of science. I have shown that the ethical concept of trust provides the tools for a new avenue of research in social epistemology: the social epistemology of trust, in the moral sense. This new line of research differs significantly from past approaches to social epistemology and dominant accounts of trust, which both rely on a narrow and rather stark conception of human nature as primarily motivated by self-interest. Rational choice theorists and social epistemologists of self-interest and both frequently identify their work as merely providing a first step in the analyses of trust and the social character of science. Almost everyone recognizes that humans are motivated by more than mere self-interest. However, it appears that one reason why these investigations have not progressed beyond this first step is that we lacked an account of trust that could accommodate non-self-interested motivations. It is my hope that the account provided here will spur further research by ethicists into a thoroughly moral conception of trust, one which recognizes that our ability, and need, to trust is deeply connected to our sense of moral duty. In addition, I hope that this account of moral trust and my initial steps in a social epistemology of trust will prove to be useful tools for philosophers of science who want to expand our understanding of the epistemic consequences of the social character of science. The success of science is grounded on more than the self-interested drive of scientists; nobler motivations like concern for others and a sense of moral duty are psychologically real, ethically significant, and epistemically important.

BIBLIOGRAPHY

Adler, J. (1994), “Testimony, Trust, Knowing,” *Journal of Philosophy* 91: 264-275.

Allen, G.E. (1978), *Thomas Hunt Morgan: The Man and His Science*. Princeton, NJ: Princeton University Press.

Alliance for Human Research Protection (2006), “Cloning Scam: Dr. Gerald Schatten's Hand in Bogus Paper Detailed,” January 14. http://www.ahrp.org/cms/content/view/29/29/ (accessed January 23, 2007).

Anderson, M.S. et al. (2007), “What do Mentoring and Training in the Responsible Conduct of Research Have To Do with Scientists’ Misbehavior? Findings from a National Survey of NIH-Funded Scientists,” *Academic Medicine* 82: 853-860.

Axelrod, R. (1984), *The Evolution of Cooperation*. New York: Basic Books.

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

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

Barber, B. (1987), “Trust in Science,” *Minerva* 25: 123-134.

Barnes, B. and Bloor, D. (1982), “Relativism, Rationalism and the Sociology of Knowledge,” in Martin Hollis (ed.), *Rationality and Relativism.* Cambridge, MA: MIT Press.

Bergmann, J. (1993), *Discreet Indiscretions*. New York: Aldine de Gruyter.

Blais, M. (1987), “Epistemic Tit for Tat,” *Journal of Philosophy* 84: 363-375.

Bratman, M. (1992), "Practical Reasoning and Acceptance in a Context," *Mind* 101: 1-15.

Buss, S. (1999), "Appearing Respectful: The Moral Significance of Manners," *Ethics* 109: 795-826.

Chong, S. (2005), "South Korean Cloning Researcher Resigns," *ScienceNOW* Daily News, 23 December.

Crane, D. (1972), *Invisible Colleges*. Chicago: University of Chicago Press.

De Vries, R., Anderson, M., and Martinson B. (2006), "Normal Misbehavior: Scientists Talk About the Ethics of Research." *Journal of Empirical Research on Human Research Ethics* 1, 43-50.

Fausto-Sterling, A. (1985), *Myths of Gender*: New York: Basic Books, Inc.

Feyerabend, P. (1975), *Against Method*. London: Verso.

Freedman, D. (1992), "Editorial Processes, Safeguards, and Remedies," in David Miller and Michel Hersen (eds.), *Research Fraud in the Behavioral and Biomedical Sciences.* New York: John Wiley & Sons, Inc.

Fox, M. (2006), "*Science* formally retracts both Korean clone papers," *Reuters*, January 12. http://www.redorbit.com/news/science/356309/science_formally_retracts_both_korean_clone_papers/index.html (accessed June 4, 2008).

Giddens, A. (1990), *Consequences of Modernity*. Stanford, CA: Stanford University Press.

Goldman, A. (1992), *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: The MIT Press.

Gould, S.J. (1981), *The Mismeasure of Man*. New York: W.W. Norton and Co.

Hall, C. T. (2005), "'Stem cell hub' cloning network project folding U.S. organizer cites 'misrepresentations' by plan's collaborators in South Korea," *San Francisco Chronicle*, November 15.

Haraway, D.J. (1991), *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

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

Harding, S. (1991), *Whose Science? Whose Knowledge?* Ithaca, NY: Cornell University Press.

Hardwig, J. (1991), "The Role of Trust in Knowledge," *Journal of Philosophy* 88: 693-708.

Hardwig, J. (1985), "Epistemic Dependence," *Journal of Philosophy* 82: 335-349.

Harris, C.E., M.S. Pritchard, M.J. Rabins (2005), *Engineering Ethics: Concepts and Cases*, 3rd Edition. Belmont, CA: Thomson Wadsworth.

Hart, H.L.A. (1955), "Are There Any Natural Rights?" *Philosophical Review* 64: 175-191.

Hieronymi, P. (2008), "The Reasons of Trust," *Australasian Journal of Philosophy* 86: 1-24.

Holton, R. (1994), "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72: 63-76.

Hull, D. (1988), *Science as a Process*. Chicago: University of Chicago Press.

Kitcher, P. (1993), *The Advancement of Science: Science without Legend, Objectivity without Illusions*. New York: Oxford University Press.

Kitcher, P. (2002), *Science, Truth, and Democracy*. New York: Oxford University Press.

Judson, H.F. (2004), *The Great Betrayal: Fraud in Science*. New York: Harcourt, Inc.

Knorr Cetina, K. (1999), *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kohler, R. (1994), *Lords of the Fly*. Chicago: University of Chicago Press.

"Korea's cloning crisis," *New York Times*, December 4, 2005.

Kuhn, T. (1962/1996), *The Structure of Scientific Revolutions*, 3rd Edition. Chicago: University of Chicago Press.

Kuhn, T. (1977), "Objectivity, Value Judgment, and Theory Choice", *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Latour, B., and S. Woolgar (1979), *Laboratory Life: The Social Construction of Scientific Facts.* Beverly Hills: Sage.

Lewontin, R., S. Rose, and L. Kamin (1984), *Not in Our Genes: Biology, Ideology and Human Nature*. New York: Pantheon Books.

Lloyd, E. (1993), "Pre-theoretical Assumptions in Evolutionary Explanations of Female Sexuality", *Philosophical Studies* 69: 139-153.

Lloyd, E. (2005) *The Case of the Female Orgasm: Bias in the Science of Evolution*. Cambridge: Harvard University Press.

Longino, H. (1990), *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

Longino, H. (2002), *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

Luhmann, N. (1979), *Trust and Power*. New York: John Wiley & Sons.

Merton, R.K. (1942/1996), "Science and Technology in a Democratic Order," reprinted in *On Social Structure and Science*, P. Sztompka (ed.). Chicago: University of Chicago Press.

Morelle, R. (2006), "Science moves on from dog clone," *BBC News Online* March 9. http://news.bbc.co.uk/2/hi/science/nature/4786646.stm (accessed January 23, 2007).

O'Neill, O. (2002), *A Question of Trust*. New York: Cambridge University Press.

Pasha, S. (2006), "Lay's Tale: 'The Nail in the Coffin…' Enron's former CEO details his lost trust in CFO amid negative media barrage," *CNNMoney.com* April 26. http://money.cnn.com/2006/04/25/news/newsmakers/enron_trial/index.htm (accessed May 31, 2008).

Poling A. (1992), "The Consequences of Fraud," in David Miller and Michel Hersen (eds.), *Research Fraud in the Behavioral and Biomedical Sciences.* New York: John Wiley & Sons, Inc.

Rachels, J. (1975), "Why Privacy is Important", *Philosophy and Public Affairs* 4: 323-333.

Railton, P. (1994), "Truth, Reason, and the Regulation of Belief", *Philosophical Issues* 5: 71-93.

Rawls, J. (1971/1999), *A Theory of Justice*. Cambridge, MA: Belknap Press.

Rescher, N. (1989), *Cognitive Economy: an Inquiry into the Economic Dimension of the Theory of Knowledge.* Pittsburgh, PA: University of Pittsburgh Press.

Rincon, P. and J. Amos (2006), "Science will Stick with Peer Review", *BBC Online*. January 10. http://news.bbc.co.uk/2/hi/science/nature/4600402.stm (accessed 6/04/08).

Shapin, S. (1994), *A Social History of Truth*. Chicago: University of Chicago Press.

Solomon, M. (2001), *Social Empiricism*. Cambridge: MIT Press.

Sturtevant, A.H. (1965), *A History of Genetics*. New York: Harper & Row.

Strawson, P.F. (1974), *Freedom and Resentment and Other Essays.* London: Harper & Row.

Thagard, P. (1997), "Collaborative Knowledge," *Noûs* 31*:* 242-261.

Thelen, M.H. and T. M. DiLorenzo (1992) "Academic Pressures," in David Miller and Michel Hersen (eds.), *Research Fraud in the Behavioral and Biomedical Sciences.* New York: John Wiley & Sons, Inc.

van Fraassen, B. (1980), *The Scientific Image*. Oxford: Clarendon Press.

Vasgird, D. (2007), “Prevention over Cure: The Administrative Rationale for Education in the Responsible Conduct of Research”, *Academic Medicine* 82: 835-7.

Wallace, R. J. (1994), *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Webb, M. (1993), “Why I Know About as Much as You: A Reply to Hardwig,” *Journal of Philosophy* 90*:* 260-270.

Williams, B. (1973), “Deciding to Believe,” *Problems of the Self*. Cambridge: Cambridge University Press: 136-51.

Woods, J. (1989), “The Maladroitness of Epistemic Tit for Tat,” *Journal of Philosophy* 86: 324-331.

Wuchty, S., B.F. Jones, and B. Uzzi (2007), “The Increasing Dominance of Teams in Production of Knowledge,” *Science* 316: 1036-1039.