

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

A

TRUST AND THE LIMITS OF CONTRACT

by

CELESTE M. FRIEND

A dissertation submitted to the Graduate Faculty in Philosophy
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy, The City University of New York

1995

UMI Number: 9530869

**Copyright 1995 by
Friend, Celeste M.
All rights reserved.**

**UMI Microform 9530869
Copyright 1995, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI

**300 North Zeeb Road
Ann Arbor, MI 48103**

© 1995

CELESTE M. FRIEND

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

4.26.1995
Date

James E. Adelle
Chair of Examining Committee

April 26, 1995
Date

Robert M. Bensel
Executive Officer

Professor Virginia Held

Professor John Kleinig

Professor Sibyl Schwarzenbach

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

TRUST AND THE LIMITS OF CONTRACT

by

Celeste M. Friend

Advisor: Professor Virginia Held

Trust is morally basic. It makes cooperation between persons, to whatever degree, possible. In Chapter One, I define trust as being the relation between people bound by genuine goodwill, competency and vulnerability to each other.

In Chapter Two, I criticize Thomas Hobbes's understanding of society as founded upon a social contract which exclusively self-interested persons have reason to make in order to escape from the state of nature. I argue that on Hobbes's assumptions about the nature of persons, such a contract cannot be viable. Without the assumption that, prior to the contract, persons have some reason to trust one another, no contract can be expected to be made or survive.

David Gauthier, who presents a contemporary version of Hobbesian social contract theory in his *Morals by Agreement*, is vulnerable to a similar argument. In Chapter Three, I argue that given Gauthier's description of persons there does not exist the trust necessary to construct, even hypothetically, the social contract.

Chapter Four examines the very basic role which trust

plays in both personal and civic friendship. I argue that trust is what distinguishes "end friendship" from the instrumental friendships in which we are "used" or taken advantage of by others. I discuss *civic friendship* and argue that the trust which is basic to it, namely *social trust*, is a form of trust which can exist between people who are otherwise strangers to one another.

In Chapter Five I go on to *prescribe* trust. I argue first, that we should never be untrustworthy, as this implies treating people instrumentally, and hence violates Kant's Categorical Imperative. I argue, however, that we should not always trust, maintaining that trust, for normative purposes, is best understood as an Aristotelian *mean*, and can be described by the extremes of *naivete* and *cynicism* and that we should trust in a way which avoids both these extremes.

For my parents

Ann S. Friend and Richard L. Friend

Acknowledgements

I want to express my gratitude to my ever supportive family. To my parents I am especially grateful. They have always supported and encouraged me. My years in graduate school have not always been easy, but without their help they certainly would have been all the more difficult.

I am extremely grateful to my dissertation advisor, Virginia Held. She has been my guide and teacher throughout my graduate school career. I am especially grateful to her for originally suggesting to me the topic of trust. She has been tremendously helpful in the production of this work. I also wish to thank the other members of my dissertation committee. Sibyl Schwarzenbach has been a very good committee member and as good a friend, helping me become a better writer and better philosopher. John Kleinig read each chapter very closely and gave extremely helpful comments. All three members of my committee, Virginia, Sibyl and John, were generous with their time, reading chapters very quickly so that I could keep to my deadlines. Their influence is present throughout this work, sometimes explicitly in their own written work which informed many of my own thoughts on trust, but implicitly as well, in the revisions for the better which came about as the result of their thoughtful comments.

Adam Vinueza has been a valued friend from my first year in graduate school. During the time in which we were both writing, we met on a regular basis to discuss our work. His support has been extremely important. He helped me sort out some important distinctions which are formulated in Chapter One.

My mother was not only helpful and supportive *qua* mother, but *qua* research librarian as well. She spent many hours in the Boston Public Library searching databases for secondary sources for me. In addition, she made a special trip to New York in order to double-check all my citations because time constrictions made it too difficult to do myself. My aunt, Janet Cymbala offered her professional proofreading skills and proofread my final draft.

Finally, I am grateful to Assaf Tamir for his close and careful readings of the first and last chapters. This dissertation is, without any doubt, better for those comments. He has also been a constant source of support and encouragement to me, more so than these words can manage to convey. I have learned more about trust from him than writing this dissertation ever could have taught me.

Thank you, everyone.

CONTENTS

Chapter One
DESCRIBING TRUST 1

- §1. Trust is Basic 3
- §2. Trust versus Prediction 10
- §3. Defining Trust 20
- Notes 40

Chapter Two
HOBBS AND THE STATE OF DISTRUST 42

- §4. Hobbesian Man 45
- §5. Hobbes's Dilemma 59
- §6. Hobbesian Man Revisited 76
- Notes 89

Chapter Three
GAUTHIER AND THE LIMITS OF CONTRACT 92

- §7. Gauthier's Version of Hobbesian Morality 94
- §8. Hobbes's Dilemma Revisited 108
- Notes 129

Chapter Four
TRUST AND FRIENDSHIP 130

- §9. Friendship as an Exemplary Trust Relationship 134
- §10. Friendship is Essential 140
- §11. Trust Essential to Friendship 145
- §12. Civic Friendship and Social Trust 155
- Notes 172

Chapter Five
PRESCRIBING TRUST 174

- §13. When to be Trustworthy 174
- §14. When to Trust 187
- Notes 217

References 218

Chapter One
Describing Trust

Sitting in a New York City cafe, a young man puts his bag on a table next to mine and asks if I will watch it for him. He returns in a few minutes. He is not gone very long, but at least long enough for me to have left with his bag. When he returns, he opens the bag and pulls out a shiny new computer, which I would estimate to be worth at least a thousand dollars. I marvel aloud at how I--a complete stranger--was left in charge of such a thing. The young man smiles and tells me I have a trustworthy face. This encounter displays both trust and mistrust: mistrust of the general surroundings, in which bags left alone on cafe tables may be gone when we return, and trust in me, a stranger, to safeguard the man's possessions in such an environment.

Annette Baier tells a somewhat similar, although even more poignant story, of a man in an airport. He is loaded down with luggage and a small child when she offers him her help. Expecting to carry one of his bags, she finds herself instead entrusted with care of the child (Baier 1994e, 194).

These two stories are of persons trusting other persons with things that matter to them. Trust is indispensable; without it, the world is a hostile and inhospitable place in which to live. With trust, we affirm people's abilities to act toward us in positive ways.¹ It is a primary human good that makes interaction, on every level, between people possible. Trust at the very least enriches, and perhaps even makes possible, our abilities to work well with other persons.² Without trust between persons, social gestures as simple as the handshake, the social cooperation necessary for the existence of social institutions, and friendship would be impossible. Without trust, the simplest uses of human language would be in vain; communication would be impossible. Even knowledge, it can be argued, is only possible in an atmosphere of trust between persons.³

There is currently, however, a great deal of concern over the dissolution of trust between persons. When I was a child, there was no lock on the back door of my house. Now, we lock our front doors three times over, use clubs and alarms to discourage would-be car thieves, and remove tamper-proof seals from our bottles of aspirin. Without much doubt, mistrust and cynicism are on the rise.⁴ Nonetheless, trust between persons still exists. We still ask strangers to watch our bags while we get ourselves coffee. We still accept help from strangers in airports. And we still ask people on the subway for directions. In

general, we still believe that most people have at least enough goodwill toward us not to take advantage of our vulnerability to them.

Not only good things, however, need trust to survive. Dishonesty, deceit and exploitation, at least to a certain extent, also thrive in an atmosphere of trust. Trust can go awry and support institutions and relationships that are morally suspect. There is trust amongst thieves, and battered wives return to their husbands because they trust that this time they will keep their promises not to beat them. While I will argue that trust is generally desirable, there are instances in which it is inappropriate. So, we must attend to the morality of trust; of when it is appropriate and when it is not.

§1. Trust is Basic

Erik Erikson describes "basic trust" as the first stage of human development, necessary to psychological health. According to Erikson, infants must first learn to trust the world around them. In the extrauterine world, infants are faced with the facts that needs are not met instantaneously and caretakers are not constantly present. Given confrontation with these facts, infants must learn to negotiate between basic trust and basic mistrust, and learn to be more informed by the former. Erikson writes:

The firm establishment of enduring patterns for the solution of the nuclear conflict of basic trust versus basic mistrust in mere existence is the first task of the ego (Erikson 1950, 249).

To have basic trust, and not basic mistrust, is the first stage in the healthy development of the person. He writes:

For the first component of a healthy personality I nominate a sense of *basic trust*, which I think is an attitude toward oneself and the world derived from the experiences of the first year of life. By "trust" I mean what is commonly implied in reasonable trustfulness as far as others are concerned and a simple sense of trustworthiness as far as oneself is concerned. When I say "basic," I mean that neither this component nor any of those that follow are, either in childhood or in adulthood, especially conscious (Erikson 1959, 57).

This sense of basic trust is manifested both in a person's attitude toward herself and in her attitude toward others. Having a sense of basic trust that is stronger than a sense of basic mistrust means being comfortable with oneself and one's own abilities, and therefore not being discouraged by the temporary absence of caretakers or hostile to them upon their return. Erikson claims:

The general state of trust, furthermore, implies not only that one has learned to rely on the sameness and continuity of the outer providers but also that one may trust oneself and the capacity of one's own organs to cope with urges; that one is able to consider oneself trustworthy enough so that the providers will not need to be on guard or to leave (Erikson 1959, 63).

Trusting ourselves and our capacity to control our own bodies, we can then also trust others. Likewise, trusting others, we come to trust ourselves and our ability to practice self-control.

Also, for Erikson, basic trust is clearly something relational; its development depends on the appropriate care of an infant by her parents and the community at large. Persons are naturally relational, and so trust, which is an attitude toward oneself and toward others, is the first virtue learned by persons. Because we all start out as persons related to others, and because we stay related to others to at least some degree throughout our lives, we must learn to trust those others, and be trustworthy to them, in order that we may have healthy, morally sensitive relationships with them.⁵ "All individuals are borne by mothers; . . . everybody was once a child; . . . people and peoples begin in their nurseries" (Erikson 1959, 17). What children must learn from their parents in order to have trust, the "cornerstone of a healthy personality" (Erikson 1959, 58), is that their parents are committed to their welfare, and that they are safe, both with them and with the larger community.

By learning this basic trust from parents, children are able to grow into psychologically healthy adults who are informed by a sense of basic trust, rather than a sense of basic mistrust, toward others:

At any rate, the psychiatrists, obstetricians, pediatricians, and anthropologists, to whom I feel closest, today would agree that the *firm establishment of enduring patterns for the balance of basic trust over basic mistrust* is the first task of the budding personality and therefore first of all a task for maternal care (Erikson 1959, 65).⁶

Trust is psychologically basic. Persons without it, Erikson maintains, are likely to be psychologically damaged: persons for whom "withdrawal into schizoid and depressive states is habitual" (Erikson 1950, 248). Infants who lack appropriate care and therefore fail to develop a sense of basic trust which can win out over their sense of basic mistrust can display "acute infantile depression" (Erikson 1959, 62). Psychologist Julian Rotter, echoing Erikson, writes "The . . . truster is less likely to be unhappy, conflicted or maladjusted; he or she is liked more and is sought out as a friend more often" (Rotter 1980, 6). In order to develop into healthy adults, we need to be grounded in the sense of security that the world is a safe place and that people will care for us. This sense is the sense of basic trust and is the first step toward a psychologically well-balanced life.

John Kleinig has also argued, although in a way different from Erikson's psychological understanding, that trust is basic. Whereas for Erikson trust is psychologically basic, Kleinig argues that trust is normatively basic, informing other normative concepts, in particular, that of harm. He argues that the concept of social harm, which has ramifications for criminal law, can be cashed out in terms of trust. Kleinig writes:

The continued existence of society of any worthwhile kind will depend on the general maintenance of a fair measure of trust between its members. Trust is perhaps

the most important ingredient in the social glue (Kleinig 1978, 35).

Accordingly, "harm to society is to be understood as the erosion of these relationships of trust" (Kleinig 1978, 35). This claim is supported, Kleinig argues, by the facts of human beings' basic social nature and the necessity of trust to support and encourage those social relations that in many ways constitute the interests of beings like us. Because trust plays such a fundamental role in social cohesion, acts which erode it (for example, criminal attempts) constitute social harm even though there may be no assignable individual who has suffered.

Wittgenstein also considered trust basic. In *On Certainty* he discusses the grounds of knowledge and the possibility of skepticism. According to Wittgenstein, trust, or certainty, precedes uncertainty, or doubt. He refers often to trust and claims that we must have it before we can entertain the skeptical position. He writes:

How does someone judge which is his right and which his left hand? How do I know that my judgement will agree with someone else's? How do I know that this colour is blue? If I don't trust myself here, why should I trust anyone else's judgement? Is there a why? Must I not begin to trust somewhere? That is to say: somewhere I must begin with not-doubting; and that is not, so to speak, hasty but excusable: it is a part of judging (Wittgenstein 1969, ¶150).

As children, we learn facts; e.g., that every human being has a brain, and we take them on trust (Wittgenstein 1969, ¶159).

The child learns by believing the adult. Doubt comes after belief (Wittgenstein 1969, ¶160).

Doubting and non-doubting behaviour. There is the first only if there is the second (Wittgenstein 1969, ¶354).

Skepticism, according to Wittgenstein, is untenable without a foundation of trust or certainty. We cannot doubt unless there is something else that we do not doubt. Certainty precedes uncertainty; mistrust only exists in an environment of trust.

Wittgenstein goes on to describe this attitude as basic, as constituting our common form of life:

That is to say, the *questions* that we raise and our *doubts* depend on that fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.

That is to say, it belongs to the logic of our scientific investigations that certain things are *in deed* not doubted.

But it isn't that the situation is like this: We just *can't* investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put.

My *life* consists in my being content to accept many things (Wittgenstein 1969, ¶341-344).

According to Wittgenstein, in order to mistrust some things, we must not mistrust others. This constitutes our common way of life.

Erikson argues that basic trust is necessary to psychological health. Kleinig argues that social trust informs the normative concept of harm and holds society together. Wittgenstein argues that knowledge is possible because we have what we might call 'propositional trust.' There are, of course, a variety of 'trusts.' We can speak

of trust in God (and in fact, our currency does), although this may better be termed 'faith.'⁷ We can speak of trust in oneself.⁸ We may even be able to speak of trust in relation to animals⁹, such as trusting a family dog not to bite us or our children. And we do, of course, speak of trust between persons. While each of these may at root be quite similar, my intention here is to limit my discussion and exploration of trust and its moral dimensions to *trust between persons*.

Trust is relational; it exists in the context of "real and felt relationships between actual persons" (Held 1993, 38). Trust, "one of those relations . . . through which we express our humanity" (Fried 1970, 143), can and does hold between individual persons, between individual persons and groups of persons, between social groups and between individual persons or groups of persons and the social institutions with which they interact. Not only can I be said to trust or distrust a colleague or even an unknown person riding in the elevator, but I can also trust or distrust a group of persons, e.g., doctors, lawyers, men, and institutions, e.g., my government or a corporation. And social groups can trust or distrust each other, so that it makes sense to claim that Americans trust doctors, or that young black men distrust police officers. On the face of it, these may seem to be different kinds of trust. But I will argue that they are not, and that in order to

understand trust and encourage it to thrive on a generalized level, i.e., make it a general characteristic of a society, we must first look to its features on the particular level. Trust between individuals in personal relationships, bound to each other by obligations and expectations, can give rise to trust between persons at large, e.g., between persons and their government, or between strangers.

According to Erikson, infants first learn to trust in the context of particular relationships with their primary care-givers. Likewise, if trust is to be a quality in a society in general, it must find its first home in the actual relationships of actual persons. My argument will be that while personal trust may seem quite different from social trust, in fact, it is not. The question of how to create and sustain a more trusting, trustworthy society can therefore be explored by attending to trust between particular persons. (§12)

§2. Trust versus Prediction

'Trust' used in a predictive sense is not to be confused with 'trust' in its thicker, moral sense. The sentence, for example, "I trust my friend Lisa to be honest with me" is an example of the latter sense of 'trust.' But the sentence, "Since Kim has not missed any of the picnics this summer, I trust that she will not miss this one" is an

example of the former. In the second example, the terms 'assume', 'predict', or 'expect' could be substituted for 'trust' without changing the basic meaning of the claim made. But, as I will argue in this section, to trust a person or some group of persons is to do more than simply predict that they will behave in a certain way. Trust is a relation between persons, social groups, and institutions constituted by beliefs in genuine goodwill and competence and by vulnerability to persons in light of these beliefs. When one person trusts another, he believes: that person has genuine goodwill toward him, that person is competent to act out of respect for him and concern for his well-being, that person could act otherwise, and that person will, at the least, not act otherwise. To trust someone is to believe all this and to be willing to be vulnerable to him because of these beliefs. To predict that someone will choose to act in some way, however, or even to predict that she will bear goodwill toward you, is not necessarily to make yourself vulnerable to the choices she makes and so is not necessarily to trust her. In fact, such a prediction can be instrumental in removing oneself from a situation in which one is liable to be taken advantage of or exploited.

In "On the Meaning of Trust," Virginia Held has argued that the essential difference between trust and prediction lies in the difference between situations of certainty and situations of uncertainty. This distinction between

certainty and uncertainty is not to be understood as a psychological attitude one person takes toward another. Rather, it is an epistemological question of whether we can ascertain how a person actually will behave. She writes: "trust is **most** required exactly when we **least** know whether a person will or will not do an action" (Held 1968, 157). (emphasis added) She argues that to predict that a person will take a particular course of action and to act in a particular way based on that belief is to act in a situation of certainty; to trust someone, however, is to make yourself vulnerable to the possibility that she may act other than how you trust her to act and that this is to act in a situation of uncertainty.

For example, we need not trust a paraplegic not to run after us, because it is a situation of certainty; the paraplegic **cannot** run after us, and therefore we are certain that she will not. But when we are in a situation of uncertainty, when we cannot be sure of another's actions toward us, i.e., when we cannot make a certain prediction of his behavior, then we are in a situation of uncertainty, and trust becomes possible. We need not trust a paraplegic not to run after us, because we are **certain** that she cannot. We do need to trust a person not to lie to us, however, because it is **uncertain** whether he will tell the truth. To trust someone is to act on the belief that he will act in particular ways, despite your uncertainty.¹⁰

Laurence Thomas also argues that there is a difference between prediction and trust. He argues that to trust, as opposed to merely predict, is both to want and to expect that another person will behave favorably toward you out of goodwill for you where this expectation is based upon that person intentionally manifesting themselves as trustworthy; it is to make yourself vulnerable to this expectation, i.e., to give someone else the power (or simply recognize their power) to harm you precisely because you believe that they will not. Given this understanding of trust, Thomas argues that while trust entails a species of prediction which is intentionally manifested, prediction does not entail trust.

In light of this basic distinction between trust and prediction, Thomas goes on to argue that there are four more particular differences between them. First, he argues that trust presupposes that the object of trust knows he is trusted, whereas nothing of the kind is true of prediction. In trusting another person, we assume that, on at least some level, he is aware of the fact that we trust him. But no such relationship need hold between an object of prediction and the person making a prediction. Second, he argues that whether or not an object wants to be trusted is relevant to whether or not she **should** be trusted, but whether or not an object wants to have predictions made of her is not relevant to the predictions made about her. It would be unreasonable, he claims, to trust someone who makes it clear

that she does not, in some particular instance, want to be trusted. (In addition, it seems that it may not only be unreasonable from the perspective of the one doing the trusting, but unfair from the point of view of the one trusted as well, since it may place upon her obligations she does not want to have.) Third, Thomas argues that the desirability, on the part of the predictor, of whether or not a prediction is borne out is irrelevant to the prediction, whereas the same is not true in trust. So, we can predict events which we do not want to occur. But to trust someone to do something is, at least in part, to want them to do that thing. Fourth, whereas the necessity of an event occurring makes a prediction all the better and more certain, it makes trust superfluous. If something **must** happen, then it is unnecessary to **trust** that it will happen, although I can predict with complete assurance that it will.

I agree with Thomas that the wishes of an object of trust are relevant to whether or not one should trust a person. If I do not want to be trusted with a secret, perhaps because I do not *trust myself* to keep it, or because I just do not want to know something which someone else would not want me to know, it may be unreasonable, or perhaps even coercive, to trust me. But I am less certain that the knowledge and wishes of the object of a prediction must always be irrelevant to the prediction; that, as Thomas writes, "Equally irrelevant is whether the object of

prediction wishes to have predictions made about it" (Thomas 1990, 240). It is also unclear that the desirability of the outcome of a prediction must always be irrelevant to the prediction in question. While these claims of Thomas may be true of predictions about nonhuman objects, I do not think that they necessarily hold of human objects.

Imagine, for example, that I discover that I am the object of a sociological study trying to determine certain facts about female consumer behavior. Certain predictions are being made about me, which, if borne out, will support a certain theory of consumerism that I consider wrong and sexist. Now, not wanting to have such predictions made about me, coupled with the discovery that exactly such predictions are being made about me, I very plausibly may change my behavior in order to reduce the likelihood that the theory in question is confirmed. Or I may change my behavior simply out of resentment at being used in such a study without my consent. (Or I may change my behavior just for the fun of it!) In fact, it is for this very reason, that persons's behavior is likely to change when they know that they are the objects of prediction, that psychological studies and the like are more successful when conducted without the knowledge of the objects of prediction that they are such objects. (For the sake of the present discussion, I ignore the potential ethical problems with studies conducted in this way.)

I think that the reason Thomas makes this mistake, as becomes obvious from the pattern of his examples, is that he thinks of objects of prediction as inanimate objects which conform to laws of nature. But if we assume that human beings, of which predictions can be made, are free to choose, within certain limits, how to behave, i.e., that they are not inanimate objects, then there is no reason to believe that the knowledge that one is the object of prediction and the desire to be such an object cannot be relevant to the making of predictions. Thomas is fundamentally correct in arguing that prediction and trust are distinct, and that trust involves more than mere prediction. My point, however, has been to show that Thomas's following claims do not hold: whether or not an object knows that it is having predictions made about it and whether or not it wishes to be the object of such predictions are always completely irrelevant to the making of such predictions.

The difference between trust and prediction is also evident in the different reactions persons have when the expectations involved in both trust and prediction are not met. Baier writes that "trusting can be betrayed, or at least let down, and not just disappointed" (Baier 1994b, 99). When we trust someone to behave in a certain way, to be loyal to us, for instance, and this expectation is not met, i.e., she is not loyal to us, the usual reaction is a

feeling of betrayal. We may feel betrayed because we thought she had the right intentions toward us, and yet acted otherwise. On the other hand, when we merely **predict** that someone will do something, but he does not, we are not as likely to feel betrayed as we are to be disappointed, and perhaps curious as to why things did not go as we thought they would. Suppose, for example, that John predicts that a friend of his, Chris, will bring his guitar to his party because he always brings his guitar to John's parties. If, despite his prediction to the contrary, Chris does not bring his guitar, it would not be reasonable for John to feel betrayed by him, given that Chris made no promise, explicit or implicit, to bring it. He may, however, be disappointed that Chris did not bring his guitar. Prediction can be disappointed, but only trust can be betrayed.

The two basic differences between prediction and trust are first, trust implies and involves a particular **relationship** between persons: a relationship in which persons are bound to each other by obligations and expectations. To be trustworthy is to be worthy of this sort of relationship with other persons. To trust is to choose to enter into (as in friendship), or to simply be in (as in parents and children), this sort of relationship with another person. And second, trust in such a relationship involves having certain intentions toward the other person,

whereas in the case of predictions, intentions are not necessarily at stake.

First, no personal relationship between persons need be implied by one person making a prediction of another, and certainly no relationship is implied when predictions are made of inanimate objects. To predict, for example, that Newt Gingrich will run for President in the year 2000 is not to make any implicit claim that there exists any relationship between Newt Gingrich and the person making the prediction. There need be no relationship bound by obligations and expectations in order for such a prediction to be made or borne out. In fact, when predictions are made of persons, it is sometimes important that there *not* be any personal relationship (for example, predictions which play a role in a scientific study).

In trust, however, the assumption of a relationship of obligations and expectations between the truster and the trusted is at least implicit. This is, I believe, consistent with the distinction made by Held. To be in a relationship with someone, whether that relationship is one we choose to enter or one which we simply find ourselves in, we make ourselves vulnerable to that other person despite the fact that it is uncertain how she will behave. In relationships of obligations and expectations, we are necessarily vulnerable. This is why, as Thomas points out, the wishes of an object are relevant to whether or not one

should trust it. To trust someone with something despite his wish **not** to be trusted is to try to impose on him a relationship into which he would rather not enter; it is to give him obligations which he does not want. But to make a prediction of something or someone is in no way to impose such a relationship on her.

Second, trust involves persons entering into relationships in which they are bound to each other by obligation and expectations by having the right sorts of intentions toward one another, whereas prediction involves no such intentions whatsoever. (Which is not to say that we cannot make predictions **about** intentions. I may, for instance, predict that one person will develop good or bad intentions toward some other person. But predictions themselves do not depend on the predictor or the object of prediction having certain intentions toward one another.) To trust someone, we must believe that he has morally appropriate intentions toward us. And at the same time, to be trustworthy is, among other things, to have the right sorts of intentions toward others.

It should be clear by now that my concern here is with the moral dimensions of trust, and not the more ordinary prediction. Predictions do not necessarily involve persons in relationships other than that of subject and object, whereas trust, consisting in part of the belief in another's genuine goodwill, does. The real difference between trust

and prediction lies in the relational character of trust; trust is composed of a belief about you coupled with my choice to be vulnerable to those beliefs about you, whereas prediction involves no relationship of vulnerability whatsoever. Prediction is not trust. The question remains however, what is trust?

§3. Defining Trust

Both Baier and Thomas define trust as a three place relation. Thomas writes:

Person X trusts person Y to act in a certain way with respect to a set of things (situations or objects) F, where F is valued by X (Thomas 1989, 176).

If X trusts Y with respect to F, then X makes himself vulnerable to Y with respect to F in that if, contrary to X's expectations, Y should fail to act in the appropriate manner with respect to F, then X has good reason to believe that he will be worse off in some important respect, and Y knows this (Thomas 1989, 177).

Baier writes that "trust . . . (is) a three-place predicate (A trusts B with valued thing C)" (Baier 1994b, 101).

Thomas also argues that trust, as described above, is to both want and expect that another person will behave favorably toward you because he has intentionally manifested himself as trustworthy. Trudy Govier defines trust in a manner similar to both Baier and Thomas, writing:

When we trust, we believe that those trusted are unlikely to harm us and are likely to do those beneficial things we expect them to do; we believe that they will not knowingly act against our interests, that

they count our well-being as something that matters. Though we may know there is a risk they will not act as we expect, we accept this risk and our vulnerability to the actions of the other (Govier 1992, 52).

Govier's definition of trust is similar to both Baier's and Thomas's in its stress upon the belief that the one who is trusted will act in particular ways despite the uncertainty of this. It is different, however, in its stress upon the person as a whole and what it is that they believe when they trust.

One problem with these definitions is the supposed connection between trust and interest. (By interest here I mean the relatively narrow sense in which something is in my particular interest if it materially benefits me, and is against my particular interest if it materially harms me. I do not intend the wider sense of interest in which we may say that being treated in accordance with respect is in my interest.) Govier's definition of trust explicitly claims that in order for a person to act in a trustworthy way toward us, she must act in our material interests and not against them. Baier's and Thomas's definitions, by defining trust in terms of behavior toward some valued object, seem to assume more or less the same thing, although less explicitly. On the contrary, however, we can trust a person even though we believe that she may not do what is in our best interest. For example, we can trust our employer even though she does not give us a promotion that we very much

want, and instead gives it to someone else because she considers him more deserving or better qualified. We would not trust our employer, however, if she led us to believe that we would get the promotion simply in order to make us willing to put in extra hours. Or a student may continue to trust his teacher even after the teacher gives the student a grade worse than what the student had hoped for, because it was not what the student in fact deserved. A student would not continue to trust his teacher, however, if he discovered that the reason she did not give him the grade he wanted (and in fact deserved) was because he did not bribe her.

Being trustworthy need not imply that we will always act in the interests of those who trust us, and so in order to trust, we should not be required to believe that another will **necessarily** act in accordance with our best interests. As Barbara Herman writes, "it does not follow that I violate [someone's] trust if I refrain from doing some things that will benefit him (because they are wrong or unfair) or if I act for someone else first" (Herman 1993, 191). Rather than concentrating on the preservation of one another's interests as the heart of trust in relationships, it may be better to describe trust as essentially involving not taking advantage of other persons; trust involves not using persons as only a means to an end. This pushes the understanding of trust in a more Kantian direction than any of the preceding definitions. Govier recognizes this when she writes:

If the Kantian imperative that others must always be treated as ends in themselves and never only as a means were to be universally observed, and were known to be universally observed, virtually all warranted distrust would disappear--there would be no dishonesty, deceptiveness, promise-breaking, or manipulation of others (Govier 1992b, 19).

Kant's second formulation of the Categorical Imperative prescribes that no one treat anyone else in a way which only takes advantage of or "uses" that person for an end outside of herself. This better describes what is presupposed when one person trusts another: the belief that she will not take advantage of his vulnerability to her, that she will not treat him purely instrumentally or treat him only as a means to an end.

This suggests that trust is better understood in Kantian terms than in interest-relative terms. In contrast to this, however, Thomas uses trust to argue against Kantian morality by arguing that the Categorical Imperative gives rise to an understanding of a person who "always chooses to do what is morally right" (Thomas 1990, 243). Thomas argues that such a person need not be trusted since we can simply predict, with complete assurance, that they will do the right thing. In essence, trusting a Kantian moral agent, or what Thomas calls a "wholly rational moral self" (Thomas 1990, 242), is superfluous, since they are incapable of doing wrong. And since to trust someone it must be at least possible that he could behave in ways which we trust he will

not, we need not 'trust' the Kantian moral agent. He writes:

a wholly rational moral self just is a self that is metaphysically constituted in such a way that there is no need to trust it to act morally, since a wholly rational moral self is conceptually unable to act otherwise (Thomas 1990, 246-247).

Kant's understanding of morality, according to Thomas, makes trust, a necessary quality of basically moral relationships and a basically moral society, superfluous.

Part of the problem with Thomas's critique of Kant, however, is that it does not take into account the second formulation of the Categorical Imperative. The second formulation, which Kant argues is equivalent to the first and the third, obliges us to treat each person, including ourselves, as an end in herself, instead of merely as a means to an end outside of herself. Kant's ideal moral agent is not simply obliged to follow abstract universal laws but also to take into account the inherent moral worth of all persons--this is an essentially trustworthy attitude. Moreover, given that what Kant describes is a moral *ideal*, it offers actual persons something against which to measure their moral actions. It does not **describe** how actual persons will in fact--or even **could**, given free will--behave. Rather, it prescribes an ideal against which persons **should** evaluate their actions. There is in Kant's understanding of morality, therefore, still room enough left

for the uncertainty which Thomas and Held rightly argue is necessary to trust.

There is, besides the problem of defining trust as relative to interest, another problem which I see with the definitions offered by Baier, Thomas and Govier. Each of them characterize trust as univocal. (This is less true of Thomas than it is of Baier and Govier. Thomas does distinguish between "basic trust" and "intimate or privileged trust" (Thomas 1989, 105). Nonetheless, he still offers a single definition of trust.) In a review of Baier's *Moral Prejudices*, John Dunn notes this assumption in her discussions of trust. He writes that she "does not distinguish as sharply as she might have done between [different] senses of the term" (Dunn 1995, 45).

The inadequacy, I believe, of defining trust univocally as a three-place relation is the oftentimes indeterminate nature of the object entrusted. Baier notes this indeterminacy when she writes that her definition

will involve some distortion and regimentation of some cases, where we may have to strain to discern any definite candidate for C (Baier 1994b, 101).

While many cases of trust do involve particular objects to be cared for, many others do not. In the most basic, and perhaps most important, cases of trust, such as those which exist between spouses or close friends, there is no assignable object to fill the place of the third object of relation. Govier suggests this when she writes: "Trust

implies expectations that have an open-ended character; there is no determinate list that a trusted friend should do" (Govier 1993, 105). The three-place relation implies that the object which fills the third place is some particular object or person other than the person who is trusting, such as a plant, book, or child. But what we trust people with in many important cases of trust is nothing in particular, or in other words, everything. To trust one's spouse, for example, is more than to trust him to take good care of the children, or the plants, or even your feelings. It is to trust him despite the fact that he may sometimes act in ways which hurt your feelings, for example. It is, among other things, to trust that he is with you for the right reasons; ultimately it is to trust him with your well-being. These cases of trust cannot be forced into a three-place definition of trust (unless, of course, we insist on defining it as "A trusts B with A").

The basic trust described by Erikson seems also not to be attached to particular objects. The kind of trust which he describes, rather than being attached to particular objects, encompasses instead a general sense of well-being and safety.

In fact, rather than trust being univocal, there are at least four senses to the word 'trust'. On the one hand, we can speak of 'trust' in the sense of prediction, as discussed above. (Prediction, is not, for reasons already

discussed, morally problematic, and so what I have to say below will not be addressed to that sense.) On the other hand, 'trust' has three senses which are morally interesting. First, there is *evaluative trust*; we can speak of trust in the sense in which we simply evaluate someone as trustworthy, but do not act on that evaluation. In this sense, then, "I trust you" has the same meaning as "I believe you are trustworthy." Second, there is *specified volitional trust*; we can speak of trust in the sense which is relevant to a particular set of circumstances, or context. For example, we may trust a person with some things but not with others. I may trust my mechanic with my car but not with my children. In this case, then, I have *specified volitional trust* for my mechanic, with relation to my car. (Specified volitional trust is well described by the three-place relation as given by Thomas and Baier.) And third, there is *unspecified volitional trust*; we can trust someone *per se* or generally, i.e., across the board. In this final sense, we trust someone not simply with particular items which need to be cared for but to take care of whatever may come along. This is the trust which can hold between spouses, lovers and close friends.

Unspecified volitional trust is not inconsistent with the lack of some specified volitional trust. For example, someone could say of her husband, "I trust him, except with the plants." There may be, however, some 'critical mass' of

negative specified volitional trusts which would undercut unspecified volitional trust. For example, if a man were to say, "I trust my wife, but I don't trust her with the children, or the household finances, or the car, or my laundry, or . . . ," we would be puzzled by what he meant in the first place when he said that he **did** trust her.

Evaluative trust is purely a matter of judging someone to have a certain character or disposition. Volitional trust, on the other hand, whether specified or unspecified, is trust in which we act on our evaluation of others as trustworthy, whether in some particular context, or in general.

What the definitions of Thomas, Baier and Govier have in common is that they are each composed of two parts: a set of beliefs which the truster has about the trusted **and** the choice of the truster to make herself vulnerable to those beliefs. Trust is a quality of 'relationships of vulnerability' in which the person remains or makes herself vulnerable to the one she trusts because she believes that he is trustworthy. We are vulnerable to the persons we trust because by trusting them, we give them some power over our well-being. Our well-being, to a greater or lesser extent, can be affected by the persons we trust. By trusting someone then, we are vulnerable because he **could** act in ways which fail to express his respect for us and concern for our well-being.

First, in order to trust another, a person must perceive or evaluate that other as trustworthy. But evaluating someone as trustworthy does not, by itself, constitute trust. This evaluation, coupled with the actual choice to make oneself vulnerable to another's trustworthiness, constitutes volitional trust, whether it is unspecified or specified.

So, half of volitional trust is believing that another person is trustworthy. In other words, to trust someone in either of the volitional senses, specified or unspecified, we must first trust him in the evaluative sense. And evaluative trust is constituted by a constellation of beliefs.

First, before we can trust another person, we must believe that he has genuine goodwill toward us; he must value us for our own sakes, and not merely as a means to his own ends; he must not treat us well only because his interests may be furthered in this way. To have genuine goodwill for another person is to accord that person intrinsic value, not merely instrumental or extrinsic value. Genuine goodwill means to see and treat another as an end in and of herself, not only as a means; it is, in essence, to act toward her in accordance with Kant's second formulation of his Categorical Imperative, "Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only" (Kant 1785, 47).

And so the first belief necessary to believing someone is trustworthy is that they have *genuine* goodwill, and not just instrumental goodwill, toward you. But, goodwill, even when genuine, is simply not enough for a person to be evaluated as trustworthy; genuine goodwill alone is not enough to ensure that someone will act toward you in a trustworthy way. Genuine goodwill is a necessary but not sufficient condition for evaluative trust. Intentions are never enough. As the well-worn proverb tells us, "The road to hell is paved with good intentions."

Second, in order to evaluate someone as trustworthy, we must judge or evaluate her **competent**. Competency is not an aspect of trust carefully attended to by the other definitions. A trustworthy person must be competent to meet the expectations we would have of her if we are to trust her: she must be "competent to care for my affairs and to further my necessities" (Kant 1775, 204). Someone may, for instance, have our best interests at heart, have genuine goodwill for us, yet simply be unable to keep a secret, be weak, or not have the courage to stand up under pressure in difficult situations and be loyal to us. An actively alcoholic parent, for example, may have nothing but goodwill for his children, and yet be incompetent when it comes to caring for them in general, or be incompetent when it comes to more particular things, such as driving them home from a friend's house in the evening. To take another example, we

may believe in a teacher's general intentions to be a good teacher to his students, but not be able to trust him to teach a particular subject because he is not competent in it.

Part of judging a person trustworthy is judging that he has the requisite skills and abilities to competently care for us, e.g., strength of character, will power necessary to keep secrets. Competency is therefore a normative category. Govier writes that when one person trusts another to do something, the truster believes that the trusted "is sufficiently knowledgeable to do so" (Govier 1993, 104). Knowledge alone is not enough, however, since a person could **know** how to do something, but nonetheless not be **able** to do it. For example, we may trust our dentist to care for our teeth. But if one day she has an accident and loses her vision, we would no longer trust her to be our dentist, even though she would presumably still **know** everything that she knew before. Sibyl Schwarzenbach better describes competence as "the capacity to perceive, to understand, as well as to respond with good will (both abstractly and in the concrete) to other persons who are very different from ourselves" (Schwarzenbach 1995, 27). Competence involves both the perception or knowledge of what a person needs and the ability to act on that perception or knowledge. More generally, competence can be understood as having *the means to further a person's general well-being*, where these means

can include a variety of practical and material skills (such as the mechanic has) as well as the cognitive skills necessary to perceiving what is necessary to a person's well-being.

Third, in order to judge a person trustworthy, we must believe it possible that she *could* choose to act in a way other than how we trust her to act. To judge a person trustworthy is to do more than simply *predict* what she will do. We must therefore believe that it lies within her actual power to act contrary to how we want her to act. For example, I cannot properly be said to *trust* an infant not to physically overpower me, since I do not believe that she is physically capable of doing so. To borrow from Thomas's example, it makes no sense to 'trust' Spock, a person fully incapable of acting immorally, to do the right thing, since he is constituted as being incapable of doing the wrong thing (Thomas 1990, 244). Likewise, it would make no sense for me to trust a friend to do something which I believe she cannot do. For example, I cannot reasonably trust someone not to sneeze while I am engaged in some delicate painting project because persons do not choose to sneeze (although I may trust her to do what she can to minimize the bad consequences of her sneezing, such as turn away). "We trust another person who *could* betray us not to do so" (Held 1984, 66). Trust involves the belief that someone could act other than how we trust she will. This is reflected in our

ordinary language; expressing trust, we often say "She would not do that," as opposed to "She could not do that."

And fourth, to trust someone in the evaluative sense, we must believe that, despite the fact that he could harm us, he will not. To trust someone is to believe that he will act in ways which express his respect for us and concern for our well-being despite the fact that he could act otherwise. Given his genuine goodwill and competency, it follows that we come to believe that he will behave toward us in ways which reflect his respect for us and concern for our well-being. We believe that he will act in ways which reveal the fact that because something matters to us it therefore matters to him. This condition is essentially the predictive component or precondition necessary to trust (§2) This fourth condition is essentially implicit within and follows from the other conditions delineated above. It simply follows that if we believe a person is competent to care for and respect us, and that it lies within his power to act in ways which do not express his respect for us and concern for our well-being, but we also believe that he has genuine goodwill for us, then we believe that he will most likely not act in these ways. So, a positive prediction of a person's behavior is a necessary, albeit not sufficient, condition of trust.

To summarize then, to trust someone *volitionally* is in part to judge her trustworthy or to trust her *evaluatively*. To evaluate someone as trustworthy is to believe that she: 1) has genuine goodwill for us, 2) is competent to act in the way we trust her to act, 3) could act so as to fail to express her respect for us and concern for our well-being, and 4) will act so as to express her respect for us and concern for our well-being despite her power to do otherwise. (We need not, however, believe that respecting us and being concerned for our well-being will be all that matter to her. When we trust someone, rather, we trust that she "will count [our] well-being as among the basic facts that determine what [she does]" (Herman 1993, 191).

Such an evaluation of someone's trustworthiness need by no means always be an explicitly held set of beliefs. It may be explicit, but in many cases, perhaps in most, it is only implicit; such beliefs are simply implied or presupposed by acts of trust. As Erikson notes, the evaluation of a person's trustworthiness need not be "especially conscious" (Erikson 1959, 57). Wittgenstein, too, notes the often unconscious nature of trust when he writes "certain things are *in deed* not doubted" (Wittgenstein 1969, ¶342). Persons behave in ways which simply imply that they have the beliefs, whether they are aware of them or not, which constitute trust.

Having the belief that someone is trustworthy on these four grounds, volitional trust involves making ourselves vulnerable to the person we trust because we believe that she is trustworthy. To trust someone volitionally is to believe her trustworthy (i.e., have evaluative trust for her) **and** to give her the power to care for your well-being, be it specified or unspecified. By trusting someone in this sense, I put myself at the risk of being harmed, either materially or otherwise, by the one I deem trustworthy.

I can, however, judge someone trustworthy without in fact trusting her in the volitional senses, for judging her trustworthy is only part of volitional trust. In other words, I can trust someone in the first sense, i.e. of evaluating them as trustworthy, but not in either of the volitional senses, which include **acting** on trust. Thomas writes, "from the fact that one knows that a person is trustworthy, it does not follow that one actually trusts that person" (Thomas 1990, 241). I can, for example, consider a person trustworthy but choose not to be vulnerable to her and therefore not to trust her perhaps because I do not fully trust my own evaluation of her, or because any risk of betrayal is too great.

To trust another person in either of the volitional senses, I must be willing to be vulnerable to her because I believe she has genuine goodwill toward me. I must be willing to **act** on my evaluation of her as trustworthy. So

it is not inconsistent to say "I think you are trustworthy, but I still just do not want to place my trust in you." For example, we may evaluate someone as altogether trustworthy and yet choose not to trust him with a secret about someone else because we simply do not want to divulge the other person's secret to anyone. Trusting someone volitionally, then, is in part to give power to another person, at the very least the power to betray you. It is to put yourself, to some degree or other, in his hands because you believe him trustworthy.

The distinction being drawn here is between **evaluating** someone as trustworthy and being willing to **act** on such an evaluation. We can trust someone in the sense of evaluating him as trustworthy, yet not act on this trust, and therefore not make ourselves vulnerable to him. We can consider someone wholly trustworthy, and yet simply choose not to enter into a relationship of vulnerability with him, for reasons which have nothing to do with our evaluation of him as trustworthy. To judge someone as trustworthy, but nonetheless not trust him, is to choose not to enter into a particular relationship in which you and he are bound to each other by obligations and expectations, despite your belief that he is capable and/or worthy of such a relationship.

Do these characterizations of trust apply to infants and young children? Baier writes: "One constraint on an

account of trust which postulates infant trust as its essential seed is that it not make essential to trusting the use of concepts or abilities which a child cannot be reasonably believed to possess" (Baier 1994b, 110). Infants, of course, do not choose to make themselves vulnerable to the persons they trust; they simply **are** vulnerable to their caretakers. But being vulnerable by itself does not constitute trust. (For example, I need not trust the governor of my state although, because I am employed by the state, I may be vulnerable to him through his budget.) As well, in the case of infants, the concept of 'belief' may be too strong. In infants, it may be better to understand this as the sense that their caretakers will continue to provide the nurturing and care that they need, in other words, the sense of safety, as Erikson describes it. For this reason, it makes sense to speak of an infant trusting her mother, even though it would not make sense to ascribe an explicitly held set of beliefs to an infant. But whether we understand it as belief or some more general sense, I think that the general point is the same. Infants can have the 'sense' that their caretakers are trustworthy or not, and can then respond to them in light of this sense. Infants respond differently to different persons in light of their 'sense' that different persons are better or worse at keeping them safe, protecting their well-being, and so on. So infants can exercise volitional trust, albeit not as

fully as children and adults, by responding differently to different persons in light of their sense some persons are trustworthy and others not. Infantile trust then, "is the primitive basis of other, more judgement-mediated, trustings" (Baier 1994e, 195).

This then is the two-fold definition of volitional trust:

I trust someone when first, I believe that she: has genuine goodwill toward me, is competent to act in the ways which I trust her to act, could act in such a way as to fail to express her respect for me and concern for my well-being, will not act in such a way; and second, I make myself vulnerable to her because I have this set of beliefs about her. In other words, I trust someone to the degree that I make myself vulnerable to her because I believe that she is trustworthy. Or in yet other words, I trust someone volitionally when I am willing to act on my evaluative trust of her.

This definition of trust clarifies the relationship which exists between persons involved in relationships of trust. The relation of trust exists when persons, believing that each has genuine goodwill for the other, and believing that each is competent to care for the other, make themselves vulnerable to each other. In a **relation** of trust, trust is reciprocal; it exists **between** persons. While trust can be unmutual, ideally it is the reciprocal, mutual vulnerability of persons, one to another, who are vulnerable precisely because they each believe that the other has genuine goodwill and is competent to care for them. This is the *relation* of trust.

As I said at the beginning of this chapter, it is trust which allows persons to interact in mutually beneficial ways. Because persons can trust one another--even when they are complete strangers to one another--we can ask each other to watch our bags and hold our children. Given a general background of trust, we can ask strangers on the subway whether we are riding on an express or a local train. And even if we do not trust the answers we get from those persons, we can trust the map on the wall to guide us safely on our trip (which is, of course, to trust the map maker, map distributor, and so on). (§12) Trust, whether personal or social, is necessary to human interaction. And, as I will argue in the following two chapters, to understand persons in ways which rule out their being trustworthy or trusting is to understand persons as incapable of constituting a viable society.

Notes

1. For a lovely representation of trust's ability to affirm persons, see Hal Hartley's film *Trust*.
2. Corporate America is certainly aware of this. See "Team Spirit is New Message at Olds" in *New York Times*, June 23, 1994. Oldsmobile is holding workshops and retreats intended to foster more trust between its employees, on the assumption that doing so will ultimately increase the company's ability to sell more cars. Saturn, another automobile manufacturer, is consciously fostering trust in Saturn car owners. See "Saturn Invites the 'Family' to a Party" in *New York Times*, June 20, 1994.
3. Lorraine Code argues that the construction of knowledge, which is "an intersubjective process," requires that there be trust between persons. She writes: "A shift toward granting due significance to cognitive interdependence reveals that the art of discerning whom one can trust is one of the most essential arts of personhood. . . . The matter of determining what human sources of knowledge are trustworthy is vital to responsible knowing" (Code 1991, 132). Because knowledge is the result of cooperation, we cannot have knowledge without first trusting. Code writes: "Like all cooperative enterprises, the division of intellectual labor depends on the cooperators' ability to trust one another to play their parts responsibly" (Code 1991, 182-183).
4. For a discussion of the necessity of increased security in modern life, see "Big Brother's Here, and--alas--we embrace him" by Floyd Abrams, in the *New York Times Magazine*, 21 March 1993. For a discussion of the erosion of social trust and its implications for government and business, see (Lipset and Schneider 1983).
5. Carol Gilligan is in many ways quite critical of Erikson. She notes, however, that his notion of basic trust is the exception in a developmental model which is in many ways a model based on individuation and separation, which she argues is male biased. She writes that "the initial crisis in infancy of 'trust vs. mistrust' anchors development in the experience of relationship" (Gilligan 1982, 12). The four stages of development which follow this one are, according to Gilligan, quite different from it in that they are primarily modeled on separation and individuation. Trust however is fundamentally born of the infant's relationship with her primary caretaker. Trust is a relational virtue.
6. In the 1985 foreword to *Childhood and Society*, Erikson sees a problem with assigning child care exclusively to women. He writes "It is now possible that only an equal, well-polarized

maturation of all-human maternal and paternal CARE can save mankind" (Erikson 1950, 10).

7. Baier writes that when philosophers have granted philosophical discussion to the subject of trust, it is often in association with God (Baier 1994a, 97). Erikson argues that basic trust gets institutionalized in society in organized religion. Religion, according to Erikson, fulfills the psychological need to feel that one is safe, and that the universe is a trustworthy place in which to live (Erikson 1950, 250 and 1959, 66-67). A commentator on Erikson, J. Eugene Wright, Jr. writes, "It is out of the sense of basic trust that faith is born" (Wright 1982, 56). For recent discussion of trust in God, see (Adams 1993, 121-123) and (Adams 1987, 16-23). Adams and Baier also argue that a nontheistic faith is necessary to morality in general. See (Adams 1995) and (Baier 1985b). Baier writes: "The persistent human adult tendency to profess trust in a creator-God can also be seen as an infantile residue of this crucial innate readiness of infants to initially impute goodwill to the powerful persons on whom they depend. So, we should perhaps welcome, or at least tolerate, religious trust, if we value any form of trust" (Baier 1985b, 207).

8. Trudy Govier discusses self-trust in "Self-trust, Autonomy, and Self-esteem." She shows that self-trust is a necessary condition for autonomy, which is the ability to control one's life (Govier 1993, 103). Govier argues as well that "There is... a strong positive analogy between self-trust and interpersonal trust" (Govier 1993, 109). Allan Gibbard, in *Wise Choices, Apt Feelings*, argues that the only alternative to self-trust is "hyperskepticism" (Gibbard 1990, 178). He also maintains that trusting ourselves, we then can trust others, we can "accord some fundamental authority to others" (Gibbard 1990, 180). Aristotle points out that "distrusting ourselves as not being equal to deciding," we "call in others to aid us" (Aristotle 1112b10). Also see (Baier 1994c, 179).

9. See (Thomas 1989, 177) and (Baier 1994b, 193).

10. The "certainty" referred to here is, of course, only relative, and not absolute, certainty. We cannot be **absolutely** certain that a paraplegic will not run after us.

Chapter Two

Hobbes and the State of Distrust

It is fair to say that Hobbes, "the true parent of rational morality and politics" (Gauthier 1988, 72), invented egoistic social contract theory, at least in its modern form. A critique of this type of social contract theory therefore starts with him, and proceeds to the present, with Hobbes's direct intellectual descendant, David Gauthier.¹ While over three centuries separate these two contractarians, they nonetheless make the same assumptions and draw similar conclusions. Just as Gauthier appropriates for himself Hobbes's assumptions about the nature of persons, he inherits the critique I will make against Hobbes. To make an argument against Hobbes is, then, to argue against Gauthier; in important ways, they are vulnerable in the same ways. I begin, therefore, with Hobbes.

I will begin this chapter by describing the salient features of Hobbes's argument for the social contract which gives rise to civil society, beginning with a description of the state of nature. There are, according to Hobbes, two closely related features of the state of nature which make

it intolerable: the exclusively self-interested nature of its persons and the lack of a common moral standard between these persons. In virtue of these two features, the state of nature is unbearably brutal. This leads Hobbes to his political conclusion that it is necessary and preferable for men² to join together to establish a contract by which the state, or Leviathan, is formed. Hobbes's argument is that the rationally enlightened self-interested person will consent to live under the absolute authority of a Sovereign because such is the only means of escape from the state of nature, which is far worse.

Macpherson, in his introduction to the *Leviathan*, writes:

it may still be argued that individual self-interest, however rational and however long-run, is not a sufficient cement to hold a society together (Macpherson 1968, 62).

I would go further than Macpherson, and argue that self-interest is not only insufficient to hold a society together, but is also a force which pulls it apart. I will argue that trust, as opposed to self-interest, while perhaps not a sufficient cement for social cohesion, is a necessary one. Exclusive self-interest and distrust go hand in hand, for trust by its very nature implies that persons take some genuine interest in one another. Trust, as has been argued, involves acting on the beliefs that another person has genuine goodwill, is competent to do what is expected of

her, and will act in a way to express her respect for us and concern for our well-being, even though she could act otherwise. (§3) When distrust runs rampant, when people believe that others have no goodwill for them, and therefore perceive themselves as being at risk of being harmed or exploited by others, then the desire to protect themselves from others runs rampant as well. This is the world as Hobbes saw it. Trust is something which holds of and between persons; it is therefore Hobbes's description of persons which, more than anything else, leads to his vision of the state of nature as essentially a state of distrust.

I will argue that what is implied by Hobbes's assumed features of persons is that the state of nature is essentially a state of complete distrust. I will go on to show that the civil society which proceeds from the state of nature thus continues to be a state of distrust and that this distrust presents an insurmountable obstacle to Hobbes's argument that the contract makes possible escape from the state of nature. First, I will argue that the contract cannot possibly be made on Hobbes's assumptions about persons. Second, I will argue that even if the contract could be made, it could not survive based on self-interest alone. I will maintain that some antecedent social conditions, one of which is trust, are necessary in order for cooperation between persons to be possible. And since Hobbes does not imagine these conditions to exist, his

argument for civil society based on a contract between men is unsuccessful.

§4. Hobbesian Man

The person, according to Hobbes, is essentially constituted by two things: passions and reason. First, the passions are what give rise in persons to actions:

These small beginnings of Motion, within the body of Man, before they appear in walking, speaking, striking and other visible actions, are commonly called ENDEAVOUR. This Endeavour, when it is toward something which causes it, is called APPETITE, or DESIRE; the later, being the generall name; and the other, oftentimes restrayned to signifie the Desire of Food, namely *Hunger* and *Thirst*. And when the Endeavour is fromward something, it is generally called AVERSION (L119³).

So, it is desire which impels persons forward, desires which give rise in persons to actions. And from this premise Hobbes draws the conclusion of ethical subjectivism. All it means for something to be a good or an evil is to be desired or not by some person:

words of Good, Evill, and Contemptible, are ever used with relation to the person that useth them: There being nothing simply and absolutely so (L120).

For Hobbes, these passions, which impel persons forward, are never-ending; the person is infinitely appetitive:

Nor can a man any more live, whose Desires are at an end, than he, whose Senses and Imagination are at a stand. Felicity is a continuall progresse of the desire, from one object to another; the attaining of

the former, being still the way to the later. The cause whereof is, That the object of mans desire, is not to enjoy once onely, and for one instant of time; but to assure for ever, the way of his future desire. And therefore the voluntary actions, and inclinations of all men, tend, not only to the procuring, but also to the assuring of a contented life; ... So that in the first place, I put for a generall inclination of all mankind, a perpetuall and restlesse desire of Power after power, that ceaseth only in death (L160-161).

This first part of the picture of the person which Hobbes draws, then, looks like this: he is moved by his desires, these desires cause him to love some things and hate others, these desires give rise in him to actions, and these desires are without end as long as he lives. So the person is an infinitely appetitive utility-maximizer, where utility is defined in terms of what he takes to be good, which is, in turn, defined by his desires. Hobbes's view of the person thus far is mechanistic; the person is a being driven by his emotions and passions; by love and hate, attraction and aversion, in much the same way that a machine is driven by springs and gears, an engine by pistons. This then is one half of what it is to be a person.

Second, the person is governed by the instrument of reason, which performs the function of obtaining for that person the objects of his desires. Reason, according to Hobbes, performs a calculative function. It figures out and weighs the consequences of actions, and in this way helps the passions be realized as fully as possible:

When a man *Reasoneth*, hee does nothing else but conceive a summe totall, from *Addition* of parcels; or

conceive a Remainder, from *Subtraction* of one summe from another (L110)

For the Thoughts, are to the Desires, as Scouts, and Spies, to range abroad, and find the way to the things Desired (L139)

Reason for Hobbes is, just as it is for Hume⁴, only instrumental; it exists in order to serve the passions in the pursuit of their desires. Reason is not itself substantive or normative; it does not evaluate or critique the objects of desire, nor does it issue moral requirements. Its sole function is to discover the best route by which to pursue that which is wanted, that which a person's passions cause him to desire. Reason alone issues no moral evaluations or commands; it only heeds the call of desire.

Reason, which calculates the costs and benefits of actions and in that way guides the desires to their most effective satisfaction, is most particularly and explicitly described by Hobbes in his "laws of nature." The laws of nature are, according to Hobbes, general precepts or rules of reason which exist in men in order that they may best defend their lives, and ultimately escape the state of nature. He writes:

A LAW OF NATURE, (*Lex Naturalis*,) is a Precept or generall Rule, found out by Reason, by which a man is forbidden to do, that, which is destructive of his life, or taketh away the means of preserving the same; and to omit, that, by which he thinketh it may be best preserved (L189).

The laws of nature, which are no more than precepts of reason, are guidelines which function as a means by which a

man can maximize his utility. Through them, men come to realize what is best for them, and hence seek it out, and likewise come to realize what is worst for them, and hence avoid it.

The first and fundamental law of nature is that men seek to end the war of every one against every one and thereby seek to establish peace between men by instituting a Sovereign authority. He writes:

That every man, ought to endeavour Peace, as farre as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre (L190).

Seeking peace is in every man's best interest insofar as other men are willing to seek it as well. When they are not, however, reason tells each man that it is in his interest to pursue the war of every one against every one, and secure what he can for himself.

Hobbes describes a total of nineteen laws of nature. They prescribe that men be gracious, compleasant, not prideful, not arrogant, and so on. These ways of acting towards other men, according to Hobbes, constitute the way out of the state of nature. As 'thick' as some of these laws may seem to be, e.g., to be gracious, they are nonetheless ordained purely by instrumental reason and self-interest. They are produced and approved of by reason, which in its purely instrumental function, seeks to satisfy man's desires, the chief of which is to leave the state of

nature. The laws of nature create the path out of this state.

Bernard Baumrin, in "Hobbes's Egalitarianism: The Laws of Natural Equality," has argued that the laws of nature betray a deontological inclination in Hobbes's political theory. He argues that Hobbes's commitment to genuine equality is necessary in order for the laws of nature to be laws at all. Baumrin writes:

So I will argue that without equality, genuine equality, none of Hobbes's laws of nature would be laws whether in or out of a state of nature, in fact they would not even be lawlike (Baumrin 1989, 119).

Building upon the claim that Hobbes's laws of nature cannot be understood without taking him to be endorsing a normative ideal of equality between men and not simply a material one, Baumrin goes on to argue that Hobbes is in fact a deontologist and that he endorses something like a categorical imperative when he puts forth his "one easie sum," which tells men to not endorse rules which they themselves would not want to live under (L214). Baumrin writes:

The only substantial difference between Hobbes's negative golden rule, the easie sum, and Kant's categorical imperative is that Kant wants us to go beyond these agents here and now, to all agents everywhere. ...the easie sum is a rule about rules; in other words, an imperative about general maxims, while the categorical imperative is a method of dealing with subjective maxims in order to make them general (Baumrin 1989, 125).

The problem with Baumrin's argument is that it fails to take adequate account of what Hobbes himself says about the character of his "laws of nature." According to Hobbes, as imperatives, prescriptions for men's behavior, the laws of nature are what Kant described as hypothetical, not categorical.⁵ Hypothetical imperatives are only valid in reference to some purported end outside of themselves, whereas categorical imperatives are valid in their own right. As hypothetical imperatives, Hobbes's laws of nature have a hold on men only insofar as men seek to best secure their lives, best gratify their desires. So even the first law, that men must seek peace if others are willing to do so as well, is conditioned by its second half, that when men have no hope of securing peace, they may use all advantages of war. Men need not seek peace categorically, i.e., absolutely, but only when others are willing to do so as well, when doing so best secures their ends; at all other times they are free to abandon their pursuit of peace without being irrational, immoral or unjust. The laws of nature only have a hold on men as long as they accord them the greatest satisfaction of their desires, i.e., only when they help him escape the state of nature. This is why they are hypothetical--they can always be stated with the preceding clause of "If you wish to live in a civil society..." or "If you desire a more commodious, pleasant

life..." Concluding his second of two chapters on these "laws of nature," Hobbes writes:

These dictates of Reason, men use to call by the name of Lawes; but improperly: for they are but Conclusions, or Theorems concerning what conduceth to the conservation and defence of themselves (L216-217).

Baumrin says Hobbes must be advancing genuine equality as a political ideal in order for his laws of nature to be laws at all. But, importantly, Hobbes himself says his laws of nature are not lawlike. So Baumrin's argument that Hobbes is arguing for a normative ideal of equality is unconvincing. The equality of which Hobbes writes is not normative, but purely empirical. Even the weakest man "has strength enough to kill the strongest" (L183). And reason for Hobbes is not substantive and does not give rise to moral imperatives to treat men in any particular way; instead it is purely instrumental.

These then are Hobbes's persons: individuals constituted by nothing other than their passions and the reason which serves these passions. Given what has been said thus far concerning his view of the person, it comes as little surprise that Hobbes advances as well the view of psychological egoism. Gauthier, in an historical essay on Hobbes, writes: "That persons are conceived to take no interest in one another's interests is implicit in Hobbes's account of the value of a man" (Gauthier 1990, 13). For Hobbes, the person is essentially self-interested; every

action the person performs is performed in virtue of a supposed benefit to himself. Even when acting in a way which might seem on the surface to be for the good of another, the motivation which underlies it is a desire for self-gain: the instrument of reason can chose a path to self-gain which also may incidentally benefit another.

Hobbes writes:

Whensoever a man Transferreth his Right, or Renounceth it; it is either in consideration of some Right reciprocally transferred to himselfe; or for some other good he hopeth for thereby. For it is a voluntary act: and of the voluntary acts of every man, the object is some *Good to himselfe* (L192).

No one **ever** does anything from any motivation or belief other than the desire for self-gain. Persons, for Hobbes, are purely self-interested; they have no genuine interest in others' well-being; they take no interest in others for the others' own sakes. Hobbes's moral agent "has no natural care for the preservation of well-being of his fellows" (Gauthier 1988, 77). Persons within the state of nature, have no relationships to each other which are not determined by the power they can gain from one another, and this power is the power to better acquire the objects of their desires. He writes:

Againe, men have no pleasure, (but on the contrary a great deale of grieve) in keeping company, where there is no power able to over-awe them all (L185).

The moral subjects inhabiting Hobbes's state of nature are autonomous, self-interested utility-maximizers who are

governed by nothing other than their desire to increase as much as possible the satisfaction of their own desires and by the instrument of their reason which allows them to most effectively do this. They take no genuine interest in others; their interest in others is confined to their interest in what others can do for them. Again, Gauthier writes, in describing Hobbes's view of the person: "He may have to accommodate their interests to attain his own, but in themselves their interests are not his concern" (Gauthier 1990, 13).

Hobbes extends his view of persons beyond the political realm into the personal sphere of particular relationships. While for the most part, Hobbes seems to be describing only **political** persons, this is not always quite so clear; his view of persons, namely their exclusive self-interest, includes for Hobbes the relations between mothers and their children. For Hobbes, the choice by a mother of whether or not to nourish her infant is informed by the fact that if she does choose to do so, the child will then be under her command and obligated to her. In Hobbes's own words:

Again, seeing the Infant is first in the power of the Mother, so as she may either nourish, or expose it, if she nourish it, it oweth its life to the Mother; and is therefore obliged to obey her, rather than any other; and by consequence the Dominion over it is hers. But if she expose it, and another find, and nourish it, the Dominion is in him that nourisheth it. For it ought to obey him by whom it is preserved; because preservation of life being the end, for which one man becomes subject to another, every man is supposed to promise

obedience, to him, in whose power it is to save, or destroy him (L254).

In essence, a mother who nourishes her child guarantees for herself a servant when that child is old enough to be able to serve her. Every relationship, even that of a mother and her child, is understood by Hobbes in self-interested and contractual terms.

In addition to this picture of persons, namely that they are subject to the forces of their desires and the instrument of their reason, that they are self-interested, infinitely-appetitive, utility-maximizers, it is also true in the state of nature that there are no moral standards which give rise to obligations between persons. Persons find themselves, in this state of nature, free from moral or political obligations to other persons, and hence free from the constrictions, moral or otherwise, which such relationships bring with them. (Hobbes's description of the implicit agreement between mothers and the children whom they choose not to expose and not to starve seems to be contrary to this. Hobbes implies an ambivalence on this subject. On the one hand, he argues that in the state of nature, there is nothing right, nothing wrong. Yet in the passage quoted above, he argues that the child **should** obey his mother, as he owes his life to her. If the child, then, does not obey his mother, does he act unjustly? Hobbes is unclear. Since the implicit agreement between the child is

only a covenant, it would, by Hobbes's own account, be unenforceable in the state of nature.) There is no common moral standard which binds persons to one another--practical reason may be constrained by the laws of nature, but since reason is instrumental and not substantial (i.e., it does not evaluate ends, instead it only determines how best to achieve them), actions are not constrained.

Hobbes assumes no prior relationships between persons which give rise in those persons to constraints on their actions towards others. Hobbes himself writes that we can

consider men... as if but even now sprung out of the earth, and suddenly, like mushrooms, come to full maturity, without all kind of engagement to each other (Hobbes 1651b, 109).

In Hobbes's state of nature, before men come together and invent the state by contracting with one another, they have no political or moral obligations to one another; there is no justice, no right, no morality whatsoever. Obligations only exist in the Leviathan. He writes, "where there is no Common-wealth, there nothing is unjust" (L202). Between the inhabitants of the state of nature, there is no common moral or political standard to which they are obliged, nothing is right, nothing is wrong. In the state of nature

this also is consequent; that nothing can be Unjust. The notions of Right and Wrong, Justice and Injustice have there no place. (L188)

The state of nature, the world inhabited by self-interested utility-maximizing persons ungoverned and morally

and politically unaccountable to each other, is a state of unceasing conflict. Hobbes rightly calls the state of nature the state of war, as there is no common morality or polity and hence no one's actions are subject to moral evaluation or political enforcement. While each man in the state of nature has reason and can therefore recognize the laws of nature, there is no enforcement mechanism to force men to abide by those laws. Morality remains internal, embedded in each man's mind as a way to get out of the state of nature if others are willing to conform their behavior to the laws of nature as well. The first law of nature prescribes that if others are willing to seek peace, then each man **must** seek peace as well. But if others are not, then men are reasonably justified in not adhering to the laws of nature, and in doing so they do not commit any moral trespass or rational mistake. And so each man is constantly in danger from every other man. There is a scarcity of the things which men desire, and yet each man comes equipped in the state of nature with infinite appetite. Hobbes concludes that:

therefore if any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies; and in the way to their End, ... endeavour to destroy, or subdue one an other. And from hence it comes to passe, that where an Invader hath no more to feare, than an other mans single power; if one plant, sow, build, or possesse a convenient Seat, others may probably be expected to come prepared with forces united, to dispossesse, and deprive him, not only of the fruit of his labour, but also of his life, or

liberty. And the Invader again is in the like danger of another (L184).

The state of nature, given the more or less material equality of persons, the scarcity of goods, the lack of a common moral standard and the self-interestedness of utility-maximizing persons, is a state of war, of endless conflicts between men for the objects of their desires. And in this state, "the life of man, [is] solitary, poore, nasty, brutish, and short" (L186).

This is the state of nature, the natural state of self-interested persons living without a common moral standard or a common government to which they can appeal. It is, however, obviously undesirable. And so persons in the state of nature, according to Hobbes, can be expected to make agreements and contracts to give up some of the rights they have in the state of nature in order to get out of it. They do so because they are rational. Given their rationality, "they would accept an institutional structure that provides the coercive force needed to motivate compliance with the laws of nature" (Gauthier 1988, 72). For since the state of nature is a state of war in which they may at any time be overpowered by any other, it is to their individual advantage to submit to the authority of some common power who can protect them from others. This gives rise to the state, the common power which is held over all men in order to protect each man from the encroachments of one another.

While in the state of nature, persons find reason to agree and contract with one another to form a government and give the Sovereign of such a government, whether it be an individual or an assembly, the power to enforce the laws which protect men from one another. The reason to make agreements with one another is the desire for self-preservation, and the greatest satisfaction of their desires, which rationality tells them, through the laws of nature, is best accomplished in a civil society, rather than the state of nature.

The finall Cause, End, or Designe of men, ... in the introduction of that restraint upon themselves, ... is the foresight of their own preservation, and of a more contented life thereby; that is to say, of getting themselves out from that miserable condition of Warre which is necessarily consequent ... to the naturall Passions of men, when there is no visible Power to keep them in awe, and tye them by fear of punishment to the performance of their Covenants (L223).

Hobbes offers a rational argument to justify the state and the Sovereign necessary to rule it. And this Sovereign, erected by the common will and justified by the individual interests of those who make up the common will, has absolute authority.

So it appeareth plainly, to my understanding, both from Reason, and Scripture, that the Sovereign Power, whether placed in One Man, as in Monarchy, or in one Assembly of men, as in Popular and Aristocraticall Common-wealths, is as great, as possibly men can be imagined to make it. And though of so unlimited a Power, men may fancy many evill consequences, yet the consequences of the want of it, which is perpetuall warre of every man against his neighbour, are much worse (L260).

Living in a state, even under a Sovereign of unlimited power, best furthers men's individual interests.

Rationality, whose role it is to determine the best means by which to secure the objects of desires, directs persons, as self-interested utility-maximizers, therefore, to agree to give up the rights they have in the state of nature in exchange for the benefits of living under the shelter of civilization. According to Hobbes, the rational thing for a man to do is to submit to a power greater than himself, that is the Sovereign, who then protects him from the encroachments of other, like-minded, self-interested persons.

§5. Hobbes's Dilemma

The two definitive characteristics of the state of nature are: the psychological egoism assumed of its inhabitants and the lack of any moral or political system to which those inhabitants might appeal. But it is just these two characteristics of the state of nature which block exit from it. This is Hobbes's dilemma--without a convention of promise-keeping,⁶ or truth-telling, i.e., some prior moral constraint upon persons, the contract, the agreement to establish the power of the Sovereign in order to escape the hardships of the state of nature, is not possible. Covenants within the state of nature have no binding force,

for there any man may take any action which he sees as conducive to his own ends. (Hobbes's description of the relationship between a mother and her child notwithstanding.) Covenants are only enforceable within civil society, within the Leviathan. He writes:

Covenants being but words, and breath, have no force to oblige, contain, constrain, or protect any man, but what it has from the publique Sword; that is, from the untied hands of that Man, or Assembly of men that hath the Sovereignty, and whose actions are avouched by them all, and performed by the strength of them all, in him united (L231).

Hobbes assumes that there are no moral constraints on his persons in the state of nature, but without at least some minimal moral codes of behavior, or some antecedent social conditions, in particular, trust, there is no reason to believe that his self-interested utility-maximizers would not break any covenant they have made if so doing suits their purposes and the maximal satisfaction of their desires. In essence, persons in the state of nature must have at least some minimal reasons to **trust** one another to abide by their agreements if the contract they make, which will release them from the misery of the state of nature, is to be viable. Persons must believe that others have at least some genuine goodwill for them, that they are competent, and that they will act as they are expected to act, i.e., honor their agreements, even though they could act otherwise. (§3) If men believe that others are exclusively self-interested, and that therefore they will

break their agreements as long as they believe it in their interest to do so, then they will have no reason themselves, given their own exclusive self-interest, to honor the agreements they have made.

Hobbes argues that this is why the power of the Sovereign needs to be absolute. (For example, he claims that the people cannot have the right to depose a monarch once they have established him as Sovereign (L229).) It is the absolute authority of the Sovereign that guarantees that people will not return to the state of nature. While the power the Sovereign is able to exact against his subjects may be severe and fearsome, life under his authority is still better than life in the state of nature. This is Hobbes's claim as to why the contract can be expected to survive. And it is why he argues that the agreement to bind oneself to such authority is ordained by rationality: it assures the best satisfaction of one's desires. My argument is that the viability of the contract can only be guaranteed by some relatively 'thick' antecedent social conditions, such as genuine goodwill, which contribute to trust. Men will only have reason to keep the agreements they have made if they believe that other men will keep them as well. But as long as men assume other men to be exclusively self-interested, and therefore untrustworthy, they will have no reasons to keep their agreements. If, on the other hand,

there is genuine goodwill and trust between men, then the contract they make is possible.⁷

I will argue that there are two arguments which can be made against the viability of the contract. The first argument is directed against the plausibility of the contract being made in the first place, given what Hobbes tells us is true of men in the state of nature. This argument, in essence, is that if what Hobbes tells us about men is true, they will never succeed in making a viable contract--it will always break down before it gets going. If men have no reason to trust each other, and therefore no reason to believe that other men will uphold their agreements, then no one will ever have reason to make agreements in the first place. "What reason could there be, [on Hobbes's view] . . . for fulfilling a promise when to do so would be to one's own personal disadvantage" (Dunn 1974, 286)? The second argument starts with the assumption that the contract *can* be made and goes on to consider whether or not it is sustainable on Hobbes's assumptions. Here I argue that Hobbes's contract, even if made, will not survive on the minimal assumptions he has to offer. Without trust between persons, i.e., without at least some genuine goodwill between persons, Hobbes's Leviathan "could never be created, nor, if miraculously brought about, sustained" (Baier 1985a, 294).

First, it can be argued that the contract itself cannot possibly be made given Hobbes's assumptions about the state of nature, in particular how he describes its inhabitants and the lack of a common moral standard. For if persons are essentially untrustworthy from the start, since they are purely egoistic and rational but have no genuine goodwill for each other, how can the contract itself be made? Carole Pateman writes:

The only contract that can be made in a contractarian world is a simultaneous exchange. If there is a delay in the fulfillment of the contract then it is extremely unlikely ever to be completed, and if one individual performs first, it is always in the interest of the other to break the contract (Pateman 1988, 57).

In the Hobbesian model of civil society, the Sovereign's role is to bring men out of the state of nature by making their contracts and covenants enforceable--hence the Sovereign is invested with the absolute authority necessary to do just this. But the original contract, or covenant between men, to leave the state of nature and place themselves under the authority of a Sovereign power is not enforceable in this way, for until the contract is actually made, there is no Sovereign to enforce it. While it is rational to keep one's covenants if others keep theirs, it is irrational to keep one's covenants if others do not. It must be remembered that, according to Hobbes, in the state of nature, covenants are "but words, and breath, [and] have no force to oblige, contain, constrain, or protect any man"

(L231).⁸ There is no enforcement of covenants other than that which the Sovereign provides, but it is covenant which is supposed to give rise to this Sovereign. There is, therefore, a temporal gap during which men can be expected to abandon their covenants with one another in favor of taking advantage of one another. This temporal gap exists between the time when men make the covenant to establish the Leviathan, i.e., leave the state of nature, and when they have actually done so and the absolute authority of a Sovereign has been established. During this temporal gap, men are vulnerable to the continued encroachments of other men. Given what Hobbes tells us of human nature, i.e., its fundamental self-interestedness, there is no reason to believe that men will not take advantage of such vulnerability. The temporal gap, therefore, never gets bridged, and men never leave the state of nature.

It makes no difference to this argument that Hobbes offers only a hypothetical social contract, or rational reconstruction of civil society. Hobbes does not describe the state of nature or contract to escape from it as historical events. Rather, he claims that we can hypothetically imagine men constructing his contract on the assumptions he gives us, and that therefore men are rationally justified in remaining under the authority of government, i.e., not dissenting from it by way of civil war. This argument is that if what Hobbes tell us about

persons in the state of nature were true, we could not rationally reconstruct society in the way which he describes. Hobbes's argument is in no way saved by appealing to the fact that he argues only for a hypothetical contract.

David Braybrooke has suggested that there are in fact two agreements present in the *Leviathan*, which Hobbes unfortunately conflates. The first is the covenant to establish a civil society governed by laws, and the second is the covenant to invest the authority of enforcing those laws in a person (whether individual or assembly) (Braybrooke 1976, 3). The argument here is that given the separate covenants which Braybrooke describes, and given what Hobbes tells us is true of persons, there is a temporal gap between the first and the second during which men can be expected to abandon the covenant to establish society and return to a state of war.

Braybrooke makes a similar argument. He argues that without different motivations or circumstances than the ones which Hobbes describes, the contract cannot be made. He writes:

In fact, rational agents, mindful of their symmetrical positions and of each other's rationality, will never arrive at the point even of professing commitment to the contract, ... To be sure, each agent will benefit from the contract, should the contract come into being... However, each agent (as both knows) stands to gain even greater benefits if he acts contrary to the contract while the other agent abides by the contract. A contract will not emerge. The agents' caution will

deprive them of a solution that both would prefer; yet they cannot, on rational grounds, help but be cautious (Braybrooke 1976, 17).

And again he writes:

fear that the other agents will at most only pretend to enter into the contract leads agents with the traditionally assumed motivations to renounce the project of obtaining a contract at the beginning (Braybrooke 1976, 24).

Braybrooke's claim is that given Hobbes's description of both the motivations of persons and circumstances in which those persons find themselves, the contract cannot be made. Braybrooke suggests instead that what is needed in order for the contract to be made is the assumption that there in fact do exist some prior social conditions between persons on the order of benevolence and altruism.⁹ Braybrooke suggests that in fact the conditions needed to establish the contract, and hence to support civil society, are already present in society. Therefore, the contract can be made, given the abandonment of Hobbes's description of persons and the state of nature. He writes:

It is true that in various real populations strong attachments to the requisite rules may not be widespread. On the other hand, their prevalence, if it were a fact, would hardly be inexplicable. Do not people in our society internalize the rule of keeping promises and the No Suckers Rule in childhood, from infancy and nursery school on? It is even easy to get them to do so; Children love to help; and they do not start out mistrustful--the difficulties children would have about keeping up a social contract have more to do with impulse and shortness of attention-span than with absence of good will (Braybrooke 1976, 32).

Thus the contract, without genuine goodwill and trust between persons, is not viable.

The second argument against the viability of the contract is also predicated on the supposed lack of genuine goodwill between persons and therefore the state of distrust between them. Pateman writes:

Hobbes' social contract brings Leviathan into being, whose absolute power is symbolized by the sword. A contractarian starting-point implies an absolutist conclusion. In the absence of all natural relations and all trust between individuals, the only way in which long-term associations can be sustained is through the force of the sword and through absolute obedience (Pateman 1988, 68).

As I will argue, even this is not possible. In fact, there is no authority powerful enough to enforce a contract made between purely self-interested persons.

Imagine that my first argument is unsound and that men actually are able to create a contract on Hobbes's assumptions alone, which affords them a way out of the state of nature. (As the above argument indicates, I do not believe that such a contract can be established. The point of the current argument is to show that even if it could be, such a contract would nonetheless not survive.) The contract is still unviable: not in the sense that it cannot be made in the first place, but in the sense that once made it will nevertheless collapse, and men will return to the state of nature. How absolute must the authority of a

Sovereign be in order to keep in line men who are exclusively self-interested?

There is one important exception which Hobbes makes to the absolute authority of the Sovereign. The subject must obey the Sovereign in whatever he commands, **except** when the Sovereign would try to put the subject to death. He writes:

If the Soveraign command a man (though justly condemned,) to kill, wound, or mayme himselfe; or not to resist those that assault him; or to abstain from the use of food, ayre, medicine, or any other thing, without which he cannot live; yet hath that man the Liberty to disobey (L268-269).

A man can disobey the Sovereign when the Sovereign commands him to kill himself or tries to execute him. Hobbes says this is so because a man cannot make a covenant to not protect his own life (L199). This is consistent with Hobbes's justification for the Sovereign: men concede to the authority of the Sovereign in **defense** of their lives. But if this is true, and the Sovereign's authority is not perfectly absolute, then all the more reason to suspect the ultimate viability of the contract. For the Sovereign cannot, in the name of protecting the contract, and hence protecting civil society, put subjects to death without them being free to resist him.

Without genuine goodwill and trust between persons, the enforcement mechanism itself is unviable. Revolution cannot be thwarted by authority; what keeps people in general from committing crime, harming one another, in essence acting

without regard to the welfare of others, is not the presence and threat of the police and the punishment meted out by the criminal justice system, but the fact that most people simply do not **want** to behave in ways which harm other people. Riots and looting are evidence for the claim that when social conditions do sufficiently dissolve, the police, and ultimately even the national guard, do not provide a bulwark against the possible breakdown of social cohesion.

Virginia Held, in *Rights and Goods*, writes:

For if mistrust engulfs government, or whatever agency would carry out the enforcement, there is nowhere to turn to supply the pressure that will keep those who think only of themselves from pulling society apart. In such a society there will never be enough inspectors, investigators, police, and prisons (Held 1984, 63).

There must exist between people some genuine goodwill, and the trust which such genuine goodwill makes possible, to make possible the social cohesion which provides a basis for civil society. And such goodwill implies that people are in general relatively trustworthy and trusting of others, including those charged with keeping the peace, such as the police, the military, and the courts. Were this not true, then chaos and bedlam would reign supreme, and the enforcement mechanism would have to be so large and powerful as to be itself unmanageable and untrustworthy. Held writes: "Any enforcement mechanisms put in place to keep persons to their contracts will be as subject to disintegration as the contracts themselves" (Held 1993,

204). A society which, in order to prevent revolution, is forced to devote most of its resources to its enforcement mechanism is not in fact viable. Is a society based on military rule in fact *civil*? If not, then the absolute authority of the Sovereign proposed by Hobbes does not enable men to enter into civil society. As long as Hobbes does not imagine people to have *any* genuine goodwill for one another, which is incompatible with the exclusive self-interest which he describes, and as long as he imagines the state of nature, and the 'civil' state as well, to be states of distrust, it is not at all clear how he can imagine escape from such a state of nature to be possible.

Hobbes takes as cynical a view of civil society as he does of the state of nature. In defense of his views of human nature, he writes:

It may seem strange to some man, that has not well weighed these things; that nature should thus dissociate, and render men apt to invade, and destroy one another: and he may therefore, not trusting to this inference, made from the passions, desire perhaps to have the same confirmed by experience. Let him therefore consider with himselfe, when taking a journey, he arms himselfe, and seeks to go well accompanied; when going to sleep, he locks his doores; when even in his house he locks his chests; and this when he knows there be lawes, and publike officers, armed, to revenge all injuries shall bee done him; what opinion he has of his fellow subjects, when he rides armed; of his fellow citizens, when he locks his doors; and of his children and servants, when he locks his chests. Does he not there as much accuse mankind by his actions, as I do by my words (L186-187)?

Hobbes clearly saw the world through a particularly cynical filter--he saw a world populated solely by inherently

untrustworthy persons, even when those persons are your children or members of your household. Civil society does not make these persons trustworthy, it simply offers a means by which to punish those whose untrustworthiness leads them to transgress the rights of other men (at least those who are caught).

The Leviathan is inherently unstable if built only on Hobbes's cynical assumptions. As long as people cannot trust each other, war seems inevitable.¹⁰

Another argument against the viability of the contract has been advanced by Pateman. She comments on Hobbes's view that even mother-child relationships are governed by self-interest:

The problem with the story is that, logically, given Hobbes' assumption that all individuals are completely self-interested, there seems no reason why any woman (or man) would contract to become a lord over an infant. Infants would endanger the person who had right over them by giving openings to their enemies in the war of all against all. Thus, all stories of original social contracts and civil society are nonsense because the individuals in the state of nature would be the last generation (Pateman 1988, 49).

Hobbes explains mothers caring for their infants by appealing to the implicit covenant formed between them. That covenant gives the mother (or anyone who keeps the infant from dying) dominion over him when he grows. So, Hobbes preserves his premise of exclusive self-interest, and nonetheless explains why persons in the state of nature care for their children. Pateman's objection is that whatever

advantage a person gets by the servitude of the child they preserved as an infant, it is not enough to justify, in terms of self-interest, such preservation. Caring for an infant makes a person in the state of nature all the more vulnerable to the aggressive encroachments of others. And so exclusive self-interest even precludes caring for children, despite its advantages, because those advantages are outweighed by the disadvantages presented by such care in the state of nature. Given self-interest alone, rationality would never endorse having or nurturing children because its advantages, i.e., having a servant, would be outweighed by its disadvantages, i.e., increased vulnerability to others.

Hobbes does consider 'trust.' But his account of 'trust', and his subsequent use, is not enough to explain how persons, motivated only by self-interest, can indeed be trustworthy enough, in any morally deep or thick sense, to be expected to keep their promises and live up to the contracts they have made. The definitions he offers of trust and other terms clearly conflate two separate concepts: reliance (or prediction) and trust. This is initially obvious from the definition offered for trust itself, where belief, trust and reliance are taken as synonymous with each other, and then equated with honor. It is also implied by a constellation of definitions offered by Hobbes. He writes:

To believe, to trust, to rely on another, is to Honour him; signe of opinion of his vertue and power. To distrust, or not believe, is to Dishonour (L153).

On the page previous to the one defining trust, Hobbes defines "honour" in terms of "value", such that to honor a man is to value him at a high rate (L152). And in turn:

The Value, or Worth of a man, is as of all other things, his Price; that is to say, so much as would be given for the use of his Power (L151).

And honor is also defined as: "to pray to another, for ayde of any kind" (L152).

The progression of definitions is this: to 'trust' a man is to honor him, to honor a man is to value him highly and to ask for his help or aid, and the value of a man is his price. Hobbes defines 'trust' in terms which ultimately come back to a reference to one's own self-interest; to 'trust' a man is to value him at a high price. The more one person 'trusts' another, the higher a price they would pay for the use of his power. As counter-intuitive as it seems, a man with more power, who is hence more valuable, is more reliable, more 'trustworthy,' whereas a man with little power is less 'trustworthy,' i.e., less reliable in the matter of keeping the agreements he has made. The more power a man has, the more 'trustworthy' he is, insofar as we understand 'trustworthy' to mean simply 'reliable.' Men 'trust' other men more when they value them more, and when they are in more of a position to be helped by them. For a man with more power than us may in fact be more relied upon

precisely because he has no need, or at least less need to harm us. And so the more power he has, the more we can reliably predict his behavior, the less capricious would be his actions. And, the more power a man has, the more he can be relied on because the more resources he has at his disposal to do what ever he is being predicted to do. This seems implied by this constellation of definitions: 'trust' is described by Hobbes in terms purely referential to the interest of the one who trusts. Trust based on genuine goodwill is not possible for the inhabitants of the state of nature, since there is no genuine goodwill to be found there. So, Hobbes's use of 'trust' has to be understood as equivalent to 'predict' or 'rely.' And this understanding is consistent with his defining trust in terms of value and power. For Hobbes, 'trust' is a relation between persons interacting as if they were in a market. It depends not at all on any notion of mutual genuine goodwill between persons, since for Hobbes no such goodwill exists.

That Hobbes uses 'trust' synonymously with 'predict' or 'rely,' and not in any sense which involves persons having genuine goodwill for one another, is clearly demonstrated in the following passage:

But because Covenants of mutuall **trust**, where there is a feare of not performance on either part,... are invalid; though the Originall of Justice be the making of Covenants; yet Injustice actually there can be none, till the cause of such feare be taken away; which while men are in the natural condition of Warre, cannot be done. Therefore, before the names of Just, and Unjust

can have place, there must be some coercive Power, to compell men equally to the performance of their Covenants, by the terror of some punishment, greater than the benefit they expect by the breach of their Covenant (L202) (emphasis added).

For Hobbes, the difference between a contract and a covenant is temporal. In a contract, there is a present "mutuall transferring of Right" (L192), whereas in a covenant, that which is contracted for may not be delivered until a future time, either on the part of one or both of the parties to the agreement (L193). In regard to a covenant, an uncertainty arises as to whether or not one or both of the parties will fulfill their side of the agreement. But, in the state of nature, a breach of such an agreement would not constitute an injustice, since until there is a commonwealth, and in particular a Sovereign who can force men to keep their covenants, there is no justice or injustice. In the state of nature, men cannot be 'trusted,' in even the predictive sense, to keep to their agreements, since they have no reason to do so. What gives them reason, and so makes them 'trustworthy,' i.e., predictable, is the presence of a Sovereign power who can punish them in such a way as to outbalance any benefit they might expect to gain from breaking their agreements. Not only are persons not trustworthy in the state of nature because there is no genuine goodwill between them, neither are they 'trustworthy' in the predictive sense because absent the presence of a Sovereign who can mete out punishment, there

is no incentive to keep to the agreements one has made. Even predictability cannot save Hobbes's men from the state of nature, since without the Sovereign, they are not even that. And so the 'trustworthiness' which arises in the commonwealth is not genuine trust of the kind which involves goodwill between persons but simple predictability based on the fact that people, out of self-interest alone, will keep to their agreements in order to avoid the wrath of the Sovereign. What the Sovereign offers, then, is not a common-wealth populated by trustworthy persons but a commonwealth populated by predictable persons, predictable precisely because the Sovereign promises punishment if they fail to live up to the agreements they make.

§6. Hobbesian Man Revisited

Macpherson, in "Hobbes's Bourgeois Man," argues that an essential feature of Hobbes's political theory is what I have argued is exemplified in his account of 'trust': that Hobbes is drawing on assumptions about persons which proceed from bourgeois society. Macpherson writes:

Hobbes's analysis of human nature, from which his whole political theory is derived, is really an analysis of bourgeois man (Macpherson 1973, 239).

Hobbes in the *Leviathan* is concerned to describe bourgeois men, the persons who make up bourgeois or capitalist

society. It is this person whom Hobbes describes when he draws the picture of the essentially self-interested, utility-maximizing, rational person who acts in such a way and only in such a way that accords with the greatest satisfaction of his own desires. It is the person acting as if in the marketplace, there to negotiate the satisfaction of his desires. And so, according to Macpherson, in the *Leviathan*, "all the relations between men tend to be the relations of the market" (Macpherson 1973, 242). Hobbes's representation of persons and the forces which drive them are historically specific; they surface in the context of the rise of capitalism. For example, Hobbes's claim that reason performs a solely calculative function is placed in the context of the rise of capitalism and parallels the rise of quantitative analysis. Macpherson points out that:

The predominance of mathematical thinking in the seventeenth century is closely related to the rise of capitalism. Quantitative analysis of the material world, of which mathematics is the purest form, was demanded increasingly from the fifteenth century in the service of capitalist technology and of the nation-state. The bourgeois mind was apt to be a mathematical mind; the mathematical mind was generally a bourgeois mind. The limitations of the one were the limitations of the other. The mathematical method is congruous with the reduction of all men to the equality of the market (Macpherson 1973, 246).

Although capitalism, at the time he is writing *Leviathan*, is still quite new, nonetheless, Hobbes offers an account of reason consistent with this description. It is purely calculative; it merely adds, subtracts, weighs the

consequences of our various actions in order to determine which of them best affords us the satisfaction of our desires.

My argument is that Hobbes's account of 'trust' offers an example of Macpherson's claim. For Hobbes, any sense of trust which relies or builds upon persons having at least some genuine goodwill for others is impossible. For the political purpose of agreeing to establish civil society by investing a Sovereign with the authority to enforce men's contracts, persons are understood to have no genuine goodwill for anyone other than themselves. They are purely self-interested. So the notion of 'trust' which Hobbes delineates gets spelled out in purely economic terms; it is a relation of the market. Trust, in any morally interesting or thick sense, then cannot be the 'trust' which Hobbes asserts is a manifestation or symptom of honoring a man. Hobbes's 'trust' is merely prediction, simple reliance, one man on another, in the marketplace, absent any goodwill between persons.¹¹

Macpherson's argument is Marxist. He argues that our understanding of the individual and moral agent is created by the economic forces which control production and labor. Structure, at the level of economics, generates superstructure, e.g., ideology. This is one means by which to criticize Hobbes's description of the person, who lives both in the state of nature and in civil society. While

this criticism is basically correct, it nevertheless does not tell the entire story. An additional criticism is feminist, and takes the perspective of gender. According to this critique, Hobbes's description of the person is biased, not only in historically and economically specific ways but also in gender specific ways. A feminist argument is that Hobbes's persons (and of course Hobbes is by no means alone on this count) are specifically market roles: roles which were traditionally occupied by men. Hobbes's persons betray a particular way of understanding the world around them and the relationships in which they partake.

This critical approach as applied to Hobbes is represented by Christine Di Stefano in her essay "Masculinity as Ideology in Political Theory: Hobbesian Man Considered." There, she examines facets of Hobbes theory, in particular his description of the person found in both the state of nature and civil society, and discovers just how historically-specific and masculine an image of the person it really is. Di Stefano's analysis of Hobbes as giving voice to an essentially 'masculine perspective' on moral theory is informed by recent theoretical work in the psychology of the differences between women and men. These differences, it is claimed, are at least in part caused by the division of child care along gender lines. Nancy Chodorow, in *The Reproduction of Mothering*, for example, analyzed the ways in which gender-identity formation and

women's primary responsibility for child care results in different psychological development for girls and boys. According to Chodorow, in the process of gender-identity formation, boys, primarily cared for by their mothers, tend to develop their sense of maleness through separation and autonomy: as being "other" than their mother. Girls, on the other hand, develop their sense of femaleness through identification with their mother, which leads to primary self-identification in terms of affiliation and connection (Chodorow 1978). Carol Gilligan postulates the theory, similar to Chodorow, that girls and boys, given their tendencies to develop differently in the way Chodorow describes, tend to develop their moral sensibilities differently as well. She argues, in *In a Different Voice*, that while men tend to understand moral obligations as being connected to issues of rights and the protection of autonomy, women tend to be more concerned with issues of connectedness, care, and the preservation of relationships (Gilligan 1982, 100).

Di Stefano's analysis builds upon her understanding of these gender differences in application to Hobbes. In particular, she emphasizes Hobbes's denial of the essential relatedness of persons and parallel atomistic understanding of the individual as features of the Hobbesian political man. In her penultimate paragraph, she sums up her discovery of Hobbes's masculine bias:

Hobbes's abstract man is a creature who is self-possessed and radically solitary in a crowded and inhospitable world, whose relations with others are unavoidably contractual and whose freedom consists in the absence of impediments to the attainment of privately generated and understood desires. Abstract man thus bears the tell-tale signs of a masculinity in *extremis*: identity through opposition, denial of reciprocity, repudiation of the mother in oneself, a constitutional inability/refusal to recognize what might be termed dialectical connectedness. Hobbes's genius and courage was to face the momentous and uncomfortable truth of this masculine revelation. His failure was the inability to recognize it as half-truth (Di Stefano 1983, 643).

Hobbes's analogy of men as mushrooms (§4) particularly illustrates what Di Stefano calls "the Case of the Missing Mother" (Di Stefano 1983, 638). She writes:

Mushrooms do seem to spring up overnight; they grow rapidly in the wild and require no special tending... they grow in clusters... Finally, mushrooms reproduce quietly and apparently asexually; spores are scattered by the wind and land haphazardly, sprouting when moisture and temperature conditions are right (Di Stefano 1983, 638).

For Hobbes, the mushroom metaphor perfectly sums up his view of the political person, especially in the state of nature. He owes nothing to anyone, for no one contributed to who he is. He is a separate, autonomous man, in control of the environment in service of his desires and through the use of his reason, and threatened only by other men, just as autonomous, just as free from debt to other persons. Hobbes conceives not only of persons' essential characteristics, (self-interested and autonomous, utility-maximizing, free from moral obligation to others) but also of moral solutions

to their inevitable conflicts in terms which are specific to market roles, which have traditionally been occupied by men.

In *Feminist Morality: Transforming Culture, Society and Politics*, Virginia Held, discussing Di Stefano's analysis, writes "From the point of view of women's experience, such a subject or self is unbelievable and misleading, even as a theoretical construct" (Held 1993, 47). Held argues that the model of 'economic man' offers a limited picture of human capability and that we would do well to replace it, in at least many areas of moral concern. She writes:

To see contractual relations between self-interested or mutually disinterested individuals as constituting a paradigm of human relations is to take a certain historically specific conception of "economic man" as representative of humanity. And it is, many feminists are beginning to agree, to overlook or to discount in very fundamental ways the experience of women (Held 1993, 195).

And this bias, of understanding persons along the model of "economic man" and therefore as seeing contractual relations as appropriate for negotiating between their conflicting interests, is ignorant of many people's experience, in particular women's, of morality and relationships. It is blind to the fact that what contributes to social cohesion is, among other things, trust. Hobbes's contractual approach, based as it is on the assumptions of exclusive self-interest and rationality, does not assume the genuine goodwill and trust necessary to make cooperation possible, to make the agreements between persons viable, whether

viable in the sense of being possible in the first place or in the sense of being able to withstand the threat of dissolution brought about by continued self-interest.

Again, Held writes:

The difficulties of developing trust and cooperation and society itself on the sands of self-interested individuals pursuing their own gain are extreme. Pure contractual society would be society perpetually in danger of breaking down. Perhaps what are needed for even adequate levels of social cohesion are persons bound together by relations of concern and caring and empathy and trust rather than merely by contracts it may be in their interest to disregard... at some point contracts must be embedded in social relations that are noncontractual (Held 1993, 204).

My argument is a species of this argument: that Hobbes's contract itself is neither possible nor sustainable on the assumptions that he makes of persons. Political agreement is only possible on the assumption that persons have at least some genuine goodwill for one another, and therefore that some genuine trust exists between them.¹² If Hobbes's claims about us are true, then we can never truly trust any other person or be trustworthy ourselves. And if this is true, then we will never manage to leave the state of nature, or at least not to return to it. Without trust, Hobbes's argument for the rational reconstruction of civil society is unsuccessful.

Given the philosophers to whom I have thus far referred, it may be tempting to think that criticisms of Hobbes's account of morality and civil society are all merely contemporary dissatisfactions with the notion of

persons as exclusively self-interested. But criticism of the very same sort represented here surfaced more than two centuries ago, a century after *Leviathan* was published. Catharine McAulay, writing in the late eighteenth century, was critical of Hobbes in ways similar to the criticisms discussed here in her "Loose Remarks on Certain Propositions to be Found in Mr. Hobbes' Philosophical Rudiments of Government and Society". McAulay and current feminist ethicists¹³ reach similar conclusions concerning the inadequacy of egoistic contractarian political theory. The similarity of these views points to the depth and importance of the concerns raised. Her remarks on Hobbes's philosophy are addressed to fundamental aspects of his thought and contractarian philosophy. McAulay addresses herself to the inadequacies of Hobbes's arguments in light of her understanding of morality and humans' obligations to one another. McAulay writes:

In his infant state, society is the only means of preserving his being; this makes him love it. In his maturer age, what Mr. Hobbes calls the dictate of right reason makes him capable of it (McAulay 1767, 3).

Her point is that at least part of what makes persons fit for society, and in Hobbes's terms fit to be a contractor, is their interrelatedness to others; because infants are first cared for by others, they come to acquire feelings of affection and goodwill for others; and having acquired such feelings, they then become capable of political agreement.

Prior relationships between persons, therefore, make cooperation, and in particular, contracts, between persons possible. An infant, born unable to care for and protect herself, becomes attached to the human community in which she finds herself cared for and loved by others. Later, as the grown child attains the use of reason, she becomes capable of full membership in such a community. Hobbes, according to McAulay, ignores this fact about persons. He does not take into account the socialization and upbringing which contribute so greatly to a person's fitness for society and political agreement. His understanding of the person as an exclusively self-interested utility-maximizer is contrary to the reality of the interrelatedness of human existence and society.

Another of McAulay's remarks is addressed to a claim of Hobbes that parents can claim no right over their children simply in virtue of their status as parents. As has been discussed above, Hobbes claims that mothers have the right of dominion over their children if they nourish them instead of exposing them. And if a mother does care for her child, that child is obligated to serve her when he is grown. (As was already mentioned, this is not entirely consistent with Hobbes's competing claim that, in the state of nature, there are no obligations.) In response, McAulay claims that while there may be civil laws which give parents rights to treat

their children as they desire or will, it is by no means a natural right for them to do so. She writes:

The mother's care for the preservation of her young is an invariable dictate of nature through all her works. ...the tenderness of maternal feeling is common to all, and is a compelling force to obey this dictate (McAulay 1767, 9).

According to McAulay, the feelings which parents, and in particular mothers, have for their children is a natural and moral fact. Mothers care for and love their children not because they see some gain in it for themselves, but because it is contrary to their nature as mothers to do otherwise. Hobbes characterizes the relationship between mothers and their children as one which exists within the bounds of self-interest; the mother cares for her children in exchange for their future obligation to serve her. According to McAulay, however, mothers care for their children not because of ulterior motives such as personal gain, but out of natural tenderness for their children. Hobbes's description of human behavior as being motivated solely from self-interest draws an inadequate picture of humans and morality.

McAulay's basic claim is that Hobbes's philosophy is based on assumptions which are contrary to human nature, and in particular, mother's natural tenderness for their children. Her argument supports the arguments of current feminist ethicists that much traditional moral theory, such as Hobbes's, is based on a particular perspective, that of

"economic man." It reinforces the soundness of these arguments, not only by offering yet another voice to the chorus but also by showing a temporal continuity to the concerns raised by contemporary feminist philosophers which extends well into the past.

Gauthier, in the same historical essay on Hobbes referred to earlier, suggests a remarkably similar position for which I have argued. He writes that it is possible that Hobbes's assumptions of human nature, and in particular his assumption of psychological egoism,¹⁴ do not tell the whole truth about that nature. In his own words:

But we may insist that it does not constitute the whole truth about human beings, and that where it does not hold sway, a richer morality may be established on the basis of sympathetic interests--not, of course, a fictitious universal sympathy, but real particular sympathies. Most important, we may suppose that without these sympathies, and the richer morality and genuine sociability which they make possible, human society as we know it would disintegrate into something approaching Hobbes's nightmare vision of the natural condition of mankind (Gauthier 1979, 22).

The argument presented in this passage is quite similar to the one I have presented here: that without the presence of some antecedent social conditions, in particular genuine goodwill and the trust which such goodwill makes possible, Hobbes's solution to the problem of how to escape the state of nature is unviable. Gauthier seems to recognize that "The ideology of radical contractarianism . . . is, from a practical point of view, bankrupt, and indeed that it will destroy us if we remain its adherents" (Gauthier 1977, 354)

By the time he works out his version of a Hobbesian contractual morality, *Morals by Agreement*, however, he seems to have overcome any misgivings about the Hobbesian assumptions made of persons, and their ability alone to make morality between people possible. By the time *Morals by Agreement* is published, Gauthier has abandoned the worry that what Hobbes tells us is true of persons is both incomplete and inadequate for morality. Gauthier proposes to take Hobbes's political conception of society based on contract and apply it to morality. Morality, too, not just politics, can be understood as a contract between rational, self-interested agents.

And insofar as Gauthier wholeheartedly adopts Hobbes's description of the person as essentially self-interested in *Morals by Agreement*, he also inherits his assumption of the essential untrustworthiness of persons. Given this, it is not clear how Gauthier can be expected to improve on Hobbes; how, given the initial premise of the self-interestedness of persons, constructing a moral society is possible.

Notes

1. Gauthier is not Hobbes's only descendant. Jean Hampton and Gregory Kavka have also taken up the Hobbesian task of constructing moral theory on a self-interested foundation. See (Hampton 1986) and (Kavka 1986). The recent revival of Hobbesian moral theorizing can be taken as archetypal of what Virginia Held describes when she writes that, "Contemporary Western society is in the grip of contractual thinking" (Held 1993, 193). Contractual thinking, and more specifically, the pervasive understanding of human relationships in contractual terms is well-evidenced in contemporary society. A men's organization in Brooklyn developed a "Consensual Sex Contract" in an attempt to protect men from false accusations of date rape. The contract was intended to clearly delineate the terms of a sexual relationship and was made to be signed by both parties in a sexual relationship before the sexual relationship commenced (see *Harper's Magazine* February 1993, pp. 28-29). In another instance, the relationship between novelist and reader has been described as a "fictional contract" (Attridge 1993, 59).

2. I prefer, when using pronouns which refer to persons generally construed, to use gender-neutral ones or to alternate in my use of gender-specific ones (as I do in Chapter One). In the present context, however, this is inappropriate for more than one reason. First, here I am describing for the reader Hobbes's views on the nature of the person, the state of nature, the social contract, etc. and when Hobbes lays out his views, he himself refers to persons as men. So my use of the terms "man" etc. is consistent with the view I am describing. Second, it is not at all clear, as I hope will become evident in the course of my arguments, that Hobbes is, or even thinks he is, describing persons generally, as opposed to men in the gender-specific sense. For this reason as well, then, it is appropriate to refer to Hobbes's persons as male, since that seems to be in fact what they are. (See Di Stefano 1983, 635) What I am trying to avoid here is what Susan Moller Okin refers to as "false gender neutrality," whereby superficial nonsexist language is imposed on what in fact is a gender-specific description of persons. (See Okin 1989, 10-13)

3. All *L* references refer to (Hobbes 1651a). All spellings are retained from the original.

4. Hume writes: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (Hume 1739, 462).

5. Kant writes, "All imperative command either hypothetically or categorically. The former present the practical necessity of a possible action as a means to achieving something else which one

desires (or which one may possibly desire). The categorical imperative would be one which presented an action as of itself objectively necessary, without regard to any other end" (Kant 1785, 31). Clearly then, Hobbes's "laws of nature" are only hypotheticalal.

6. Hume, in "The Original Contract," argues that Hobbes's contract cannot claim to be the original contract, since promise-keeping as a convention is necessary to provide such a contract with a foundation. He writes: "We are bound to obey our sovereign, it is said; because we have given a tacit promise to that purpose. But why are we bound to observe our promise? It must here be asserted, that the commerce and intercourse of mankind, which are of such mighty advantage, can have no security where men pay no regard to their engagements" (Hume 1777, 481).

7. Not everyone agrees that exclusive self-interest is incompatible with trust. Nicholas Rescher argues that for reasons of self-interest, evolution has selected trust. He writes: "A society of communicators is accordingly destined to evolve under the pressure of rational self-interest into a kind of mutual-aid association whose members are engaged in a collaboration to create and maintain a fabric of trust" (Rescher 1990, 47-48). Evolution thus creates "situations in which self-interest simulates altruism" (Rescher 1990, 50). Russell Hardin also associates trust with self-interest, writing: "I can trust you to do what is in my interest in certain contexts **because** your doing so will fulfill some interest of your own" (Hardin 1990, 191) (emphasis added). As should be clear, Rescher's and Hardin's understanding of trust is incompatible with the definition of trust in Chapter One (§3). Their understanding of trust in fact comes down to no more than prediction based on an estimation of what a person takes to be in his own interest.

8. It is, therefore, hard to imagine how men in the state of nature even talk to one another, as trust is necessary to communication, which is a cooperative endeavor. Baier writes: "Speech is our cooperative and trust-facilitated activity *par excellence*" (Baier 1994c, 175). Sissela Bok argues that communicative trust is the most basic form of trust. She writes, "trust in some degree of veracity functions as a *foundation* of relations among human beings; when this trust shatters or wears away, institutions collapse" (Bok 1978, 31). Gauthier disagrees, and writes: "self-consciousness and language must be taken as conditions, not products, of society" (Gauthier 1977, 332).

9. Interestingly enough, Braybrooke also explores the possibility that trust is the missing ingredient in Hobbes's social soup. However, as much as this certainly peaked my excitement when I first read the essay, I was eventually disappointed on two counts. First, Braybrooke seems to have in mind a notion of

'trustworthiness' different than mine. He writes: "Let us try begging the question by assuming trustworthiness (without assuming benevolence)." (Braybrooke 1976, 29) Whatever definition of trust he has in mind (sadly he does not offer one) is such that it need not involve genuine goodwill (i.e., benevolence) between people. Second, he concludes his inquiry into trust by suggesting that widespread or general trust between people concerned to establish a civil society is an unrealistic assumption (Braybrooke 1976, 30).

10. In the *Star Trek: The Next Generation* episode "The Gambit", in which the captain of a pirate ship controls the crew through an implant in their necks by which he can, through a remote control device, inflict great pain. It is a clever illustration, albeit probably unwittingly, of a Hobbesian civil society. The crew members, while essentially out for their own gain, (i.e. untrustworthy) are nonetheless kept in line by their sovereign's absolute authority. In the end however, Picard (*Star Trek's* ever-present hero) is able to manipulate the remote control device so that the captain, in trying to force the crew members to obey him, instead kills himself. The absolute authority he wielded was not only not enough to control the crew, but in the end, it led to his own demise.

11. It is not even clear, however, that the market can function without trust. Baier writes, "Even the most competition encouraging systems do rely on some trust in contracts, the most controlled economies on some trust in those administering the controls" (Baier 1994e, 185). There, she quotes the work of Jon Elster and Karl Moene, who write, "some amount of trust must be present in any complex economic system" (Baier 1994e, 185).

12. Locke certainly recognized this, and so portrayed the state of nature, which he did not think was only a hypothetical construct, as richer in morally 'thick' relationships between people. For a discussion of the role of trust in Locke's account of contractual society, see (Dunn 1974).

13. See, for example, (Baier 1988), (Held 1993) and (Pateman 1988).

14. Gauthier uses the economist's term of "non-tuism" to describe Hobbes's assertion that persons take no genuine interest in each other (Gauthier 1986, 87).

Chapter Three
Gauthier and the Limits of Contract

In David Gauthier's contractarian account of morality, as laid out in his influential 1986 book, *Morals by Agreement*, he explicitly takes himself to be a Hobbesian moral philosopher. The assumptions that Gauthier makes of persons as moral agents remain essentially unchanged from Hobbes's description of persons in the state of nature. However much Gauthier is like Hobbes in terms of the assumptions which each makes of persons, they nonetheless differ in their solution to the problem posed by cooperation between essentially self-interested persons. While Hobbes sees the authority of an all-powerful sovereign as necessary to maintain peace between these persons and enforce their cooperative agreements, Gauthier believes that he can show why, given rationality alone, persons have reason to cooperate with their fellows. Gauthier claims that Hobbes "underestimates (or denies) the potential efficaciousness of internal or moral constraints, and so overstates the necessary extent of external or political constraints" (Gauthier 1988, 81). He endeavors to offer a moral solution to the problem of cooperation, which he believes is

preferable to the **political** solution offered by Hobbes.

As I will argue however, it follows from the definition of trust offered earlier that Gauthier's self-interested rationality is not enough to get trustworthy moral agents. I will argue that, at most, Gauthier has described means by which persons may reasonably **predict** one another's behavior, but has not provided adequate grounds for persons to **trust** one another. As for Hobbes, Gauthier's description of persons fails to yield moral agents who can truly be trusted. This follows from the exclusive self-interest assumed of persons, which is in itself incompatible with a thick concept of trust between persons.

Gauthier illustrates his contractarian account of morality in a story of a woman and a man, Joanna and Jonathan, stranded on a desert island. Gauthier describes why, given the rational morality he has described, Jonathan does not rape Joanna and why Joanna can trust Jonathan not to do so. This illustration in fact shows why Joanna does **not** have grounds to trust Jonathan and therefore would not choose to cooperate with him. Thus, Gauthier is vulnerable to the same argument already made against Hobbes, namely, that absent some antecedent social conditions between persons, such as genuine goodwill and trust, cooperation, and the contracts or agreements which negotiate such cooperation, are unlikely to be made and equally unlikely to survive. Joanna will not trust Jonathan unless she believes

that he has at least **some** genuine goodwill towards her, minimal goodwill being the intention not to harm her for **her own** sake. Contractarian morality that takes as its starting point exclusively self-interested persons offers a diminished, if not impossible, way of understanding persons and morality.

§7. Gauthier's Version of Hobbesian Morality

In *Morals by Agreement*, Gauthier offers a 20th century account of Hobbesian morality. "Morality," he writes, "faces a foundational crisis. Contractarianism offers the only plausible resolution of this crisis" (Gauthier 1991, 15). In the first few pages of the book, he makes it clear that he is a Hobbesian. "In Hobbes we find the true ancestor of the theory of morality that we shall present" (MBA10¹). Gauthier, like Hobbes before him, assumes certain characteristics to be true of persons. "Radical contractarianism together with the conception of human nature as appropriative entail . . . the individualistic utility-maximizing conception of rationality" (Gauthier 1977, 344). And, in *Morals by Agreement*, he writes:

Morals by agreement, it has been said (for I have said it), are the morals of economic man. And economic man is the natural man of our time. . . . Economic man, whether capitalist or worker, is an indefinite appropriator, seeking to subdue more and more of the

world to his power exercised in the service of his preferences (MBA315-316).

Gauthier introduces an economic notion into his description of persons--"non-tuism." People can be assumed to be non-tuistic, i.e., not interested in the interests of others (MBA87). And people are described as rational insofar as they seek to maximize their own individual utility: "the rational person still seeks the greatest satisfaction of her own interests" (MBA7). Gauthier admits that people are in fact sociable and that they therefore desire the well-being of others, but this desire is simply counted as one among their many and varied desires, and so the person remains essentially self-interested. While a person may prefer that her friend do well, it cannot be for the friend's own sake, but only because it is one of that person's own desires. Persons's interests in each other are counted merely as members of their sets of desires, and so people have no **genuine** goodwill for each other, e.g., they do not wish their friends well for the **friends'** sake, but only for their own, because it is one of **their own** desires, and not one of their friends'.

Given these assumptions about human nature, i.e., that people are infinitely appetitive, self-interested and endowed with practical rationality which enables them to satisfy their desires and hence maximize their utility, Gauthier, much like Hobbes had before him, constructs a

moral theory which will allow these individuals to co-exist in mutually beneficial ways. Gauthier sees his project, in *Morals by Agreement*, as continuing the legacy left to him by Hobbes, "the greatest of English moral philosophers" (Gauthier 1990a, 11). He writes: "The task left to the moral and social theorist today is to establish the proper bounds of the moral and rational conventionalism that was first conceived by Thomas Hobbes" (Gauthier 1990, 23).

One of the clearest illustrations of Gauthier's intention to understand persons on the model of "economic man" is in an illustration of his interpretation of the Lockean Proviso. Gauthier, borrowing the original example from Nozick (Nozick 1974, 185), imagines sixteen Robinson Crusoes, each stranded on separate islands, each endowed with different natural characteristics, different natural environments, and different abilities to transfer goods and persons between their islands. Gauthier uses this hypothetical example to demonstrate what he argues would be rational prior constraints to interactions between persons. The assumptions implicit in the example, however, are that people can be understood as individuals **prior** to social interaction, and in fact, that they **should** be so understood. "To conceive all social relationships as contractual is to suppose that men, with their particular human characteristics, are prior to society" (Gauthier 1977, 331) This is a crucial assumption of the "economic man," that he

is a rational, exclusively self-interested, utility-maximizer who exists independently of others and independently of moral relationships between himself and those others. While common sense tells us that no man is an island, Gauthier seems to believe that man is best understood as **alone on an island**. Hobbes counsels us to imagine men as mushrooms: Gauthier revises this metaphor and counsels us to imagine men as separate individuals on separate islands.

Gauthier, however, sees a problem with Hobbes's solution to cooperation between persons. He writes, "The weakness of traditional contractarian theory has been its inability to show the rationality of compliance" (MBA15). Hobbes concludes that in order for cooperation between rational self-interested men to be possible, the power of an absolute authority is necessary. The solution for him is political. Gauthier sees his project as instead offering a **moral** solution to the problem of cooperation among self-interested men. What Hobbes offers is an *external* motivation for men to be just and cooperative, i.e., the threat of punishment by the sovereign. Gauthier, describing Hobbes, writes:

Hobbes transformed the laws of nature... into precepts of reason that require each person, acting in his own interest, to give up some portion of the liberty with which he seeks his own survival and well-being, provided others do likewise. But this agreement gives rise to actual constraint only through the efficacy of the political sovereign; from the standpoint of moral

theory, the crucial step requires the intervention of a *deus ex machina* (MBA10).

Gauthier wants instead for the motivation to cooperate to be *internal*, i.e., to come from a person's own rational capacities. If Gauthier can succeed in offering purely rational reasons for persons to comply with their agreements, then they are motivated to do so absent the threat of force; this is the moral solution to the problem of cooperation. And this is how Gauthier desires to improve upon Hobbes. According to Gauthier, Hobbes's persons are "condemned to seek political solutions for those problems of natural interaction that could not be met by the market" (MBA174). His intention is to show that "moral solutions are rationally available" (MBA174). Hobbes insists that persons are endowed with practical rationality, calculative and predictive in essence, which enables them to satisfy their desires as efficiently as possible. Gauthier develops this strain of thought in Hobbes and turns to rational choice theory to provide the basis for his contractual morality. For Gauthier, rationality is normative, i.e., it provides dictates for behavior. He writes: "a person acts rationally if and only if she seeks her greatest interest or benefit" (MBA7). So, a person who seeks to maximize the interests of others for the others' own sakes, and not her own, fails to act rationally, on Gauthier's view. To act

for the benefit of others as one's end, i.e., for their sake and not one's own, is to act irrationally.

Gauthier, in the desire to fix Hobbes's problem of motivation, employs the filter of rational choice theory, in particular the model of the Prisoner's Dilemma. For Gauthier, there is a purely rational solution to why self-interested agents should cooperate. Using the model of the Prisoners' Dilemma, Gauthier argues that there is a rational and hence moral solution to the problem of cooperation.

In the example of the Prisoners' Dilemma, first introduced by Albert Tucker in 1950 (Poundstone 1992, 116-117), two prisoners, whose exclusive desires are to spend as little time in prison as possible, are each being held by a prosecutor for both a serious crime and a less serious crime. The evidence needed for conviction of the serious crime is less than reliable, and so the prosecutor is in need of confessions from each of them in order to secure their convictions on that charge. The prosecutor is certain, however, that he can convict them easily on the lesser charge. Each prisoner is told by the prosecutor that if she confesses on the matter of the more serious crime and the other prisoner remains silent, she will get off scott free on the serious crime, and so only serve a short time for the lesser crime, while her counterpart will be harshly punished. If however, they each confess, then they will both be moderately punished. But if they both remain

silent, i.e., do not confess, then they will both get off scott free on the more serious charge. The dilemma can be represented by the following grid:

	<u>Thelma</u>	
	Confesses	Does Not Confess
<u>Louise</u>		
Confesses	5 years each	1 year for Louise 10 years for Thelma
Does Not Confess	10 years for Louise 1 year for Thelma	2 years each

In order to maximize the satisfaction of their own preferences and avoid the harsher punishment, each prisoner reasons that no matter what the other prisoner does, i.e., whether the other prisoner remains silent or confesses, she is better off confessing. The conversation each holds with herself is this: "If she confesses, then I should confess, thereby being sentenced to five years instead of ten. And if she does not confess, then I should confess, thereby being sentenced to one year instead of two. So no matter what she does, I should confess." (This, of course, assumes that each prisoner is relatively risk-averse.) Each reasons this way, and each confesses, thereby bringing about the consequence that they are each sentenced to five years in prison. The dilemma arises because had they both remained silent, i.e., not confessed, they would each have been sentenced to only two years. The dilemma arises if each prisoner acting alone tries to maximize her own interests by confessing, because then they both are punished

moderately. This result, that each is sentenced to five years in prison, is not as desirable as the consequence of each prisoner remaining silent, and thereby each being sentenced to only two years in prison. Based on a desire to maximize the satisfaction of one's own preferences, in this instance minimizing years in prison, the rational thing to do then is to remain silent, that is, to cooperate indirectly with the other prisoner by refusing to cooperate with the prosecutor, assuming that the other prisoner recognizes that this is the best way to maximize the collective good. Cooperation with the other prisoner yields the best possible outcome for each of the prisoners, provided that the other prisoner acts in this way as well.

Gauthier writes:

But if we abstract from this particular story, we recognize the seriousness of the problem--in any situation with a structure parallel to that faced by Fred and Ed [the prisoners], supposedly rational utility-maximizers do much worse for themselves than could supposedly irrational optimizers (MBA80).

Gauthier takes the lesson of the Prisoner's Dilemma to be that when engaged in interaction with others in such a way that their decisions affect the outcome of your own utility-maximization, it is irrational to act alone without taking into account the interests of others and trying to maximize your own utility. The lesson of the Prisoner's Dilemma is that acting alone results in suboptimal results for those interacting, including oneself. Hence cooperation,

represented in the Prisoner's Dilemma as not confessing, which results in the optimal utility-maximization for each party to the interaction, is the rational course of action.

This initial version of the Prisoner's Dilemma presupposes that the self-interested desires of the parties to the dilemma are self-directed. That is, the dilemma as represented above assumes that the preference of each of the prisoners is to gain for *herself* the least amount of prison time as is possible. The dilemma can be represented in a way, however, which abstracts out from such a self-directed description of their preferences. It can instead be represented in such a way as to allow that the preferences of the parties be either self-directed or other-directed. This more abstract version of the Prisoner's Dilemma assigns ordinal values to the various outcomes instead of cardinal ones, thus allowing the actual content of the preferences to be more generalized, and thus open to a variety of interpretations. This version of the dilemma is represented as follows:

	<u>Thelma</u>	
	Confesses	Does Not Confess
<u>Louise</u>		
Confesses	3rd for each	1st for Louise 4th for Thelma
Does Not Confess	4th for Louise 1st for Thelma	2nd for each

According to this more abstract and generalized version of the Prisoner's Dilemma, the preferences of the two parties

can be anything whatsoever and yet still generate the paradoxical result that acting alone, i.e., taking into consideration only what you expect the other to do and not what is in their interest, leaves you in a worse-off position than you would be if you took into consideration both what you expect the other to do and what it is in their interest that you do. Both prisoners acting in the former way gives each the outcome in accordance with her third preference, while acting in the latter way gives each her second preference. Hence, the dilemma "depends simply on the structure of interaction, on the way in which each person's endeavor to fulfill her own preferences affects the fulfillment of everyone else" (Gauthier 1991, 23). And what is unsettling about the dilemma is that "the common good is subverted by individual rationality" (Poundstone 1992, 216).

Gauthier names the two possible strategies represented in the Prisoner's Dilemma as "straightforward maximization" and "constrained maximization." To be a straightforward maximizer is to act in all situations in a way as to maximize one's own utility, given others's strategies, but not given their utilities. A constrained maximizer, on the other hand, acts in a way which takes into consideration not simply others's strategies, but others's utilities as well. To take into account the other's strategies is to act in accordance with how you expect the other will act. To take into account their utilities is to act in such a way as to

keep in mind how they will fare as a result of your action and allow that to affect how you choose to act. Gauthier writes:

Let us say that a *straightforward* maximizer is a person who seeks to maximize his utility given the strategies of those with whom he interacts. A *constrained* maximizer, on the other hand, is a person who seeks in some situations to maximize her utility, given not the strategies but the utilities of those with whom she interacts (MBA167).

To be a constrained maximizer, on Gauthier's view, is to have a certain disposition, i.e., the disposition to act cooperatively with one's fellows. It is not simply a rational stance one takes because it avoids the suboptimal results of straightforward maximization (as demonstrated in the Prisoner's Dilemma), but an habitual disposition one adopts and incorporates into one's psychology which leads one to be a cooperative person, which in turn is rationally justifiable. He writes:

we shall show that under plausible conditions, the net advantage that constrained maximizers reap from cooperation exceeds the exploitative benefits that others may expect. From this we conclude that it is rational to be disposed to constrain maximizing behavior by internalizing moral principles to govern one's choices (MBA15).

Having described morality in terms of "constrained maximization," Gauthier must go on to show how these constrained maximizers, as moral agents, can have reason to trust one another. Absent trust, there would be no rational reason to act in a way which makes one vulnerable to the choices of those around one. Absent trust, in the

Prisoner's Dilemma, the rational choice is to maximize one's utility without regard to the other prisoner--to act alone. Trust, therefore, is necessary if people are to act cooperatively.

Gauthier clearly believes that the moral agents he describes are trustworthy. They can be so, he maintains, as long as we include a constraint on their description. That constraint is *translucency*. Translucency lies on a continuum somewhere between transparency and opacity. To be translucent is to have one's motives neither completely hidden from nor completely bare to the inspection of others. According to Gauthier, transparency is too strong a constraint on persons. It is too unrealistic an expectation for persons to have of one another, i.e., to be able, with complete certainty to read each other's motives. He proposes instead that we assume translucency of persons in order to be able to judge them trustworthy. He writes:

We may appeal instead to a more realistic *translucency*, supposing that persons are neither transparent nor opaque, so that their disposition to co-operate or not may be ascertained by others, not with certainty, but as more than mere guesswork. Opaque beings would be condemned to seek political solutions for those problems of natural interaction that could not be met by the market. But we shall show that for beings as translucent as we may reasonably consider ourselves to be, moral solutions are rationally available. (MBA174)

It is this assumption, translucency, which Gauthier claims can protect people from Hobbes's Foole, who presents himself

as a constrained maximizer, when in fact he is not, in order to take advantage of others. Gauthier describes the Foole:

Is not the Foole's ultimate argument that the truly prudent person, the fully rational utility-maximizer, must seek to appear trustworthy, an upholder of his agreements? For then he will not be excluded from the co-operative arrangements of his fellows, but will be welcomed as a partner, while he awaits opportunities at their expense--and, preferably, without their knowledge, so that he may retain the guise of constraint and trustworthiness (MBA173).

According to Gauthier, this argument put forward by the Foole is answered by the constraint of translucency. For translucency allows us to see the Fooles for who they are; it lays their motives bare, or at least bare enough for us to see through them and steer clear of their exploitative traps. So, if we can more or less see persons as they in fact are, and not just how they appear to be, then we can safely trust those whose motives we can see well enough to judge that they are genuinely disposed to cooperate with us, free from the duplicity of the Fooles. Peter Vallentyne describes the role of translucency in Gauthier's theory:

For if our characters are sufficiently translucent, and we have not adopted a policy of complying with rational agreements, we will be excluded from beneficial cooperative arrangements, because others will not trust us (they will see that we won't keep the agreements we make). Thus rationality requires us to adopt a policy of complying (Vallentyne 1991a, 10).

Gauthier argues that translucency can give his moral agents reasons to cooperate with one another, and therefore cooperation between them is possible.

Gauthier's assumptions about persons are essentially two-fold: they are self-interested, even if some of their interests concern the welfare of others, and they are rational in the pursuit of their utility. Gauthier's argument is that these persons will, insofar as they are rational, dispose themselves to be constrained maximizers, rather than straightforward maximizers because, as the Prisoner's Dilemma shows, this yields optimal results for all parties and therefore a greater net gain in utility for themselves. In addition, in order to protect persons from the Foole who disguises himself as a constrained maximizer, when in fact he is not, Gauthier assumes translucency of his constrained maximizers. This assumption gives us reason, so Gauthier argues, to trust other persons, and hence to enter into cooperative agreements with them.

If Gauthier has been successful, then he has done what he set out to do, that is, construct a Hobbesian moral system which operates without the presence of the sovereign's absolute authority. In what follows, I will argue that he is not successful. I will argue that he has not given us reason to trust each other, and therefore that, on his barest assumptions about persons, cooperation between persons remains out of reach.

§8. Hobbes's Dilemma Revisited

Given that trust involves in no small part that persons take **genuine** interest, and not just instrumental interest, in one another, persons as Gauthier describes them cannot be trustworthy to one another. The moral agents he describes are translucent, prudent utility-maximizers who have disposed themselves to cooperate with others because that is the best way to maximize their own utility. The end of each one of their actions is to best maximize the satisfaction of their own interests, which means, given that they interact with others, they are rationally bound to constrain their maximization in order to fare well in Prisoner's Dilemma-like situations. They take into account the utilities, and not only the strategies, of those with whom they interact, not because the utilities of others matter **per se**, but because that is the most prudent way to maximize their own utility. This is the essence of their untrustworthiness: their rationally enlightened but nonetheless exclusive self-interest. Gauthier's moral agents do not, and by his description cannot, on pain of irrationality, take into account the interests of others independent of the good it will do themselves.

Gauthier, of course, thinks that his agents are trustworthy. His argument concerning translucency is that it can allow people to trust one another and protect them

from the obviously untrustworthy Foole. But the difference between Hobbes's Foole and Gauthier's constrained maximizer is not their basic attitude to the others with whom they interact, but only the means by which they seek to maximize their utility: their end is the same, only their strategies differ. The Foole is a straightforward maximizer who maximizes his utility without regard to the utilities of others. Gauthier's moral person, or constrained maximizer, instead maximizes his utility by disposing himself to consider the utilities of others and therefore affording for himself the benefits of cooperation. If the Foole can succeed in hiding his true aim from us, which is to take advantage of us in order to maximize his utility, Gauthier can offer him no reason why he should not continue his straightforward maximization. Translucency is necessary if we are to be able to protect ourselves from the likes of him. But, in fact, it will only protect us from the Foole not smart enough to hide his true intentions and motivations from us.

In an earlier article, "Reason and Maximization," where Gauthier develops the view he will eventually put forth in *Morals by Agreement*, he had assumed **transparency** of persons in order to avoid the problem of the Foole (Gauthier 1975, 227-230). In the later work, he recognizes this assumption as too strong and too implausible. (He credits Derek Parfit with pointing out to him this problem in the earlier

argument (MBA174).) He therefore scales down the assumption of transparency to that of translucency. Gauthier, describing why translucency is preferable to transparency as an assumption about persons, writes:

But to assume transparency may seem to rob our argument of much of its interest. We want to relate our idealizing assumptions to the real world. If constrained maximization defeats straightforward maximization only if all persons are transparent, then we shall have failed to show that under actual, or realistically possible, conditions, moral constraints are rational. We shall have refuted the Foole but at the price of robbing our refutation of all practical import (MBA174).

Transparency scaled down to realistic proportions, according to Gauthier, is translucency. And translucency, he believes, is a plausible assumption to make of persons. Given that Gauthier is concerned that his assumptions of persons be realistic, why does he continue to assume that persons are exclusively self-interested? Vallentyne writes:

Given that people have at least a limited sympathy for others, it would be inappropriate to assess norms on the basis of what their outcome would be if people had no sympathy for others. Rationality requires that one use realistic assumptions (Vallentyne 1991b, 72).

The first question to be raised in regards to translucency, therefore, is whether or not it is in fact a realistic assumption to make of persons. As Geoffrey Sayre-McCord writes:

Yet deceptive people will be careful to provide the requisite (though misleading) evidence for those with whom they interact. They will develop winning smiles, travel with a glowing reputation, and cultivate an honest manner. Sadly this sort of magic is worked (without a ring of Gyges) all too frequently. Such

people seem both translucent and trustworthy (Sayre-McCord 1991, 192).

People are notoriously opaque to one another, especially persons unknown to each other. That translucency of motives exists between persons who are unknown to each other seems rather far-fetched. And so while Gauthier makes the claim that translucency is a realistic assumption whereas transparency is not, what only seems clear is that translucency is less unrealistic, but perhaps too unrealistic nonetheless. Human experience is replete with stories of dissimulation and deception. There is rarely a person so untouched by these experiences as not to know just how skilled persons can be at representing their motives as different from what they in fact are, and how unskilled we can be at consistently judging those motives. Baier notes, "Our actual psychology diverges importantly from that required of the hypothetical progenitors of Gauthier's liberal individual" (Baier 1988, 326). At least one of the ways in which our actual psychology is different from Gauthier's description is that people are often far from being translucent even to themselves, much less to other persons: we often hide our true motives even from ourselves. The best we may be to ourselves is translucent. But it is a far cry from that to being translucent to others.

We are to some extent less opaque, more translucent, to those whom we know and who know us, those with whom we have established histories of mutual goodwill. But part of what establishes this translucency, to the extent that it exists, is trust. Gauthier's argument depends in part on his claim that translucency can give persons reasons to trust one another, so that they may discriminate between untrustworthy Fools and trustworthy constrained maximizers. But the causal chain which he describes, that translucency gives reason to trust, is not as simple and one-directional as he would make it seem. It is not simply that, absent trust altogether, we are translucent to each other and only then do we come to trust one another. We trust one another, believing that we have genuine goodwill for one another, and in this way make ourselves vulnerable to each other in ways which establish the groundwork for laying our motives bare to one another (to the degree that we can). Translucency does not simply give rise to trust, trust also gives rise to translucency. We judge the motives of persons towards our interests by trusting them with things which matter to us, and seeing how well they live up to that trust. When a person acts in ways contrary to the trust we have of them, we judge that their motives towards us may not be what we had hoped. When a person acts consistently with the trust we place in them, we judge that their motives towards us are appropriate. Persons are simply not, as Gauthier claims

they **must be** in order for cooperation between them to be possible, translucent to each other absent a history of interaction and evaluation.

Gauthier's argument about the necessity of translucency for coming to trust seems to presuppose the belief on his part that we are in a state of agnosticism with regards to each other's trustworthiness, rather than a state of 'default' trust. Translucency, according to his argument, gives us reason to mistrust the persons whom we can see have not developed the cooperative disposition, i.e., the Fooles, and reason to trust the constrained maximizers. However, as was indicated in Chapter One, it may be better (and in fact more accurate) to understand most persons, most of the time, as already trusting one another, i.e., in a state of default trust. We do not so much use translucency, to whatever limited extent we can, to judge others as trustworthy, as we do to judge them untrustworthy. In other words, we begin by extending people benefit of the doubt by trusting them. I am not translucent to the young man in the coffee shop who asks me, a complete stranger, to watch his bag while he gets his coffee, yet he trusts me anyway. And upon coming to have more insight into their dispositions and motives, we may judge them as untrustworthy, or we may confirm and indeed strengthen our initial attitude.

Even if translucency were realistically possible between persons unknown to each other however, Gauthier's

moral agents would still be untrustworthy. Just as for Hobbes, Gauthier's moral agents cannot be trustworthy given his description of them as essentially self-interested. What Gauthier gives us is a description of persons whom we can **predict** will behave in some general ways. Constrained maximizers, whom Gauthier claims are moral agents, can be predicted to act in ways conducive to other persons's ends when it is in their own interests to do so. Translucency simply makes our predictions concerning these people somewhat more reliable. Gauthier's assumption of translucency gives us the ability to 'size people up' with some degree of accuracy: our predictions about how others will treat us is not merely guesswork **if** they are translucent to us. (And, of course, the intelligent Foole may not be translucent at all.) But simply 'sizing people up' is not trusting people, it is merely making predictions of them. The predictions we make of moral agents, however, are rather general, and may themselves, therefore, not be as reliable as to generate cooperation between persons. The most specific prediction we may be able to make of persons on Gauthier's account is that they will take our utilities into account **as long as** this helps to maximize their own overall, long-term utility. But as soon as cooperating with us fails to benefit them, we can predict that they will abandon the cooperative endeavor.

Part of the lesson which Gauthier argues we can derive from the Prisoner's Dilemma is that moral agents concerned to maximize their own long-term utility should habituate themselves to take into account the interests of others. In other words, they should develop *dispositions* to be cooperative. Why then, are these dispositions, which are, after all, dispositions to be cooperative, not enough to make persons trustworthy? Ultimately, it is because of what justifies such a disposition. Persons are rationally prescribed to develop and adopt such a disposition because it is in their *own* interest to do so. The disposition itself, therefore, is only justified insofar as it continues to be in the self-interest of the person who has it. In other words, "the rationality of (the act of) forming a disposition is one thing and the rationality of acting in conformity with the disposition quite another" (Yi 1992, 486). Either their cooperative dispositions will lead them to sometimes act against their own interest (which Gauthier would describe as irrational), or they will abandon or at least override their cooperative dispositions when rationality requires that they do so. In any instance in which abandoning such a disposition, or overriding it, is the best way to maximize one's long term utility (as it would be for the Foole clever enough to convince people she has a cooperative disposition when in fact she does not), rationality would dictate that one should in fact abandon or

override it. This is true because rationality dictates, on Gauthier's view, that one should always act in ways which maximize one's, and only one's, long-term, overall utility. "Given the aim of showing that self-interested people can adhere to the requirements of an altruistic morality, one does not want an argument in which the people end up not in fact being self-interested" (Thomas 1989, 69). As long as self-interest, and only self-interest, is the exclusive justification Gauthier can offer for a person's actions, then those persons cannot be trustworthy, even with their cooperative dispositions.²

Gauthier appeals to the model of the Prisoner's Dilemma in order to show why constrained maximization is the prudent, rational way for persons concerned to maximize their long-term utility to behave. But the Prisoner's Dilemma assumes the self-interest of those involved (even when the agent's preferences are in part constituted by the well-being of other persons). Whether self-interest is construed narrowly so as to only include the maximal satisfaction of the preferences of the agent, or more broadly so as to include the preferences of others as well, the Prisoner's Dilemma assumes that each agent wants only what he wants and accords intrinsic value only to his own ends. So trust, which is in part constituted by genuine goodwill between persons, is excluded at the level of basic moral choices. As long as Gauthier describes persons as

essentially and exclusively self-interested, they cannot be trustworthy, since "trust is needed to resist the temptation to betray" (Held 1984, 69). The most that Gauthier has done, if he has even done that, is to describe conditions under which persons can to some extent **predict** one another's behavior, but not trust one another; this is clearly illustrated by a story which Gauthier uses to show how his version of the Lockean Proviso. (According to Gauthier, the proviso would determine what bargainers are allowed to bring to the cooperative endeavor. It would prohibit potential bargainers from bringing to the bargaining table the fruits of prior interactions which were forced on others, that is, it prohibits people from exploiting the ones they wish to cooperate with because such exploitation would limit their opportunities for cooperation, which they judge to be in their own interests.) Here a long passage is necessary:

Consider an example of interaction constrained by the proviso. Joanna and Jonathan find themselves castaway on a small and otherwise uninhabited island. Concerned to establish a basis for cooperation, each resolves to comply with the proviso. Thus Jonathan does not force Joanna to submit to his sexual desire, even though it might better his situation (in relation to her absence) because it would worsen hers (in relation to his absence.) But when Joanna expresses her willingness, Jonathan maximizes his gratification without concern for her arousal. He betters himself without worsening her situation. Later of course, they may agree to enhance each other's sexual enjoyment, finding this mutually beneficial. But such mutual enhancement is cooperative. Each takes whatever food he or she pleases, without concern for the other. Since in the other's absence each would have access to all the food the island provides, each would worsen his or her situation by refraining from taking whatever food he or

she wants and can get. However, when Joanna begins to cultivate a garden and grows a greater supply of fruits and vegetables, Jonathan refrains from seizing what he pleases, since he would then be bettering himself (for in her absence there would be no garden) by worsening her situation (since in his absence she would enjoy the garden alone). And when Jonathan makes a fishing rod and catches a greater supply of seafood, Joanna, for like reason, refrains from seizing what she pleases. Of course, by this time we should expect Joanna and Jonathan to have reached an agreement to co-operate--and a good thing, too, given the pronounced swelling in Joanna's abdomen. But only such an agreement--explicit or implicit--can bring them to an optimal state of affairs. In natural interaction, even constrained by the proviso, each assesses every situation separately in deciding what to do, so that only immediate reciprocity can be taken into account. In co-operation, they may extend their horizon of concern, accepting costs now for expected benefits later (MBA206-207).³

Jonathan does not rape Joanna, not because he sees that this constitutes grievous harm to **her** and he does not want to cause such harm to **this person**, but because it would violate the proviso and therefore prohibit cooperation between them in the future, which would be to his advantage. Jonathan's resolution not to violate the proviso by refraining from rape and pillage is predicated on the fact that he is concerned to establish a cooperative relationship which will benefit him in his pursuit of utility-maximization. And in order to establish such cooperation, Jonathan resolves to treat Joanna in such a way that she will later willingly comply with whatever cooperative agreement they may strike. Jonathan does not manage, however, to treat Joanna in a way which expresses anything other than her instrumental value. And taking her to have nothing other than instrumental

value, he takes her to be "valuable in the way that a pen or a typewriter or a hammer is valuable" (Hampton 1993, 243).

The question is, does Joanna, given Jonathan's reasons for not raping her and not stealing what she has grown in her garden, have grounds for trusting him? Imagine that Jonathan is in fact transparent, and not merely translucent, to Joanna, i.e., she can clearly see all his motivations and internal principles which determine his behavior towards her. What she sees if Jonathan is transparent is that he will not hurt her precisely because he expects to gain benefits from refraining from hurting her. She sees that insofar as Jonathan can expect future benefits from cooperating with her, he will not treat her in such a way as would prohibit such cooperation between them. He will not hurt her, either by raping her or stealing the products of her garden, because to do so is inimical to his own interests. To Jonathan, Joanna is someone who can be useful to him, someone who can, through cooperative interaction, help him to best maximize his own utility. To Jonathan, Joanna is "like a rock or a river or a lump of food" (Dawkins 1989, 66). This is what Joanna sees when Jonathan's motivations and intentions are laid bare to her. What Joanna does **not** see when she inspects Jonathan's intentions, is the intention not to harm her for her own sake, the motivation not to harm her because harming her would be **harming her**. While Gauthier rejects transparency

because it is too unrealistic as an assumption to make of persons, the point is that even if Jonathan were transparent to Joanna, she nonetheless would not have grounds upon which to trust him. The most she would in fact have would be grounds upon which accurately to predict his actions. And what she could accurately predict is that the moment her usefulness to him comes to an end, he will no longer be apt to be helpful to her. She can imagine, for example, that if he decided that he would be better off alone on a nearby island, then he may very well abandon any cooperative project they have begun, taking the boat they have been building together, and leave her to fare on her own, whether or not she would be better off alone. Even transparency would not give her reason to trust him. In fact, given what she sees if he is transparent, she has reason to be wary of him, reason to distrust him. And if she has reason to actually distrust him, instead of reason to trust him, as Gauthier maintains, then she has reason to be wary of entering into cooperative arrangements with him. If she can see that his only motive is to secure for himself maximal utility, she will not be disposed to cooperate with him (at least no more than she already has to in order to protect her own interests--just as we may 'cooperate' with a person mugging us with a gun and asking us to give her all our money), but will instead be prepared to protect herself from his advances, whether they be on her body or on her garden.

Distrusting him, the necessary conditions for her to cooperate with him are not present. (And of course, while Gauthier describes Jonathan as the active participant and Joanna the passive participant in their cooperative endeavor, all that holds true of Jonathan equally holds true of Joanna. He has as little reason to trust her as she has to trust him.)

This is, in essence, the same argument already directed at Hobbes. If there are not even minimal moral constraints on persons' behavior towards other persons, then the viability of the agreements those persons would make is suspect. Absent some antecedent social conditions between persons, in particular genuine goodwill and the trust to which it can give rise, cooperation, and the contracts or agreements which negotiate such cooperation, are unlikely to be made, and equally unlikely to survive. Joanna will not trust Jonathan unless she believes that he has at least **some** genuine goodwill towards her, minimal goodwill being the intention not to harm her for **her own** sake; she will not trust him as long as she believes that he accords her no value in her own person. As long as she believes he is purely self-interested, she will believe that he "would on occasion have sufficient reason to violate the moral relation" (Fried 1970, 83). And as long as Joanna does not trust Jonathan, she will not be inclined to make herself more vulnerable than she may already be to him by agreeing

to cooperate with him. In fact, she would have every reason to build her own boat and leave the island to get away from him. Other than perhaps the thin 'cooperation' described above, cooperation between them seems excluded given the very description of their mutual unconcern

Similar to Pateman's concern about Hobbes's moral theory, Baier has expressed the concern that the generation of agents of Gauthier's cooperative morality will be the last. Gauthier's assumptions can give no one reason to care for their young. Discussing the story of Joanna and Jonathan recounted above, Baier notes that given Gauthier's bare bone assumptions about persons and their reasons for acting, Joanna cannot expect aid from Jonathan if she has nothing to offer him in exchange. Baier writes:

Suppose we continue this moral lesson a bit further. Suppose a storm destroys Joanna's garden just as she goes into labor. She has nothing now to eat or to trade, or even, for the time being, very good sexual services. But rational Jonathan must not let her be a parasite... What then? What bargain is it rational for her to make? What is it rational for her to eat, or allow to be eaten? She might let her child feed at her breast, recalling that she must herself have been a similar parasite; she might hope that Jonathan might be moved by pity or paternal concern, so supply her with food till her garden can be recreated. But this would be mere parasitism-encouraging sentimentality, on Gauthier's account. So presumably either both she and the child die of neglect, or she trades her milk for Jonathan's surplus fish while the child starves, or she eats or trades the child, all without anything unfair or immoral on her or Jonathan's part, by the standards of Gauthier's modest morality for would-be market-goers (Baier 1988, 325).

Gauthier offers no account of personal affections, mutual concern and benevolence which would give persons reason to care for one another in the absence of personal gain. If Jonathan cannot gain from it, then he has no reason to help either Joanna or their child (unless he of course believes that he can, a la Hobbes, later make the child his servant and that the benefit of this outweighs the cost of caring for the child (\$4)). Given this, it would seem that no one would ever have reason to raise their children. And so, once the rational contract is established, to the extent that persons abide by it, it damns humanity either to its demise or to servitude to parents. Again, Baier writes:

Gauthier's version of morality, it seems, does nothing to ensure that any young helpless human beings survive, since they all begin as parasites or free riders. ... So the libertarian morality Gauthier is advocating would last only one generation, if we all adopted it, that is if we successfully pretended we were not affectively bonded, in order to freely choose a few rational bonds for which we could develop a free affection, and successfully disposed ourselves not to renege on our cool rational choices (Baier 1988, 327).

This is the point: absent antecedent social conditions, such as mutual concern, benevolence, affection between persons, and perhaps most importantly, trust, cooperation between persons is not possible. Whereas Gauthier proposes that having established morals by agreement, 'free affectivity' between 'tender, playful' individuals is then possible, in fact it is the other way around. Persons naturally have and display affection for each other (the

example of mothers and children relied on by Baier is an obvious, albeit not unique, example). Persons are constituted, at least in part, by mutual concern, benevolence and nonmalevolence and, perhaps more importantly, by the trust to which these give rise. Given these antecedent social conditions, cooperation between them is then possible; but not vice versa.

Michael Sandel, in *Liberalism and the Limits of Justice*, has argued that the liberal conception of the individual is untenable. While Sandel's critique is directed at Rawls, his concerns are germane to my argument against Gauthier. Sandel argues that the "antecedently-individuated self" (Sandel 1982, 62) which liberal theory posits is incompatible with human experience. He writes that such a self "fails plausibly to account for certain indispensable aspects of our moral experiences" (Sandel 1982, 179). The moral self of liberal theory, he contends, is necessarily prior to his experiences and socialization. The self, therefore, is not constituted by his relationships with others. According to liberal theory, persons can be understood as "distinct individuals first, and then (circumstances permitting) we form relationships and engage in co-operative arrangements with others" (Sandel 1982, 53). This is the self which Gauthier describes as much as it is the one assumed by Rawls. My argument is that such a self, who is not in part constituted by his genuine goodwill for

others, cannot be understood as trustworthy. Sandel argues that to imagine the self as prior to society is to "imagine a person wholly without character, without moral depth" (Sandel 1982, 179). Such a person is, at least, also untrustworthy.

Given that both Gauthier and Hobbes assume that morality can be derived from assumptions that persons are rational and exclusively self-interested, they are vulnerable to the same criticism. If there are no antecedent social conditions between persons, their agreements seem unmakeable to begin with or unmaintainable. The leap to morality simply cannot be gotten from Gauthier's assumptions about persons.

My argument against Gauthier, and Hobbes as well, is compatible with recent feminist positions. It is an argument, at its root, against the description of personhood offered by Gauthier and Hobbes which is counter to the experience of many real persons, including most women. The "Robinson Crusoe image of 'economic man' is false for almost everyone, but it is totally false in the case of infants and children" (Held 1993, 207). It is an argument which challenges the notion, long dear to the hearts of many moral philosophers, that morality can be built upon the assumptions of mutual unconcern, self-interest, noncaring. But no one can manage without trust. This claim comes to light recently because of feminist criticisms (although

certainly not only feminist) of traditional moral theories and the ways in which some (not all) of them argue that we **can** manage without the 'thick' moral concepts such as benevolence, care, and trust.

Gauthier, however, sees and understands his arguments as essentially compatible with feminist concerns about justice and exploitation. Gauthier understands his moral theory, at least in part, as responding to questions raised by feminist thought.

First, as has already been mentioned, in the preface to *Morals by Agreement*, Gauthier tells the reader that he has used a randomizing device to evenly distribute female and male pronouns. But the actual use of genders in the story seems something other than random.

Second, Gauthier quite explicitly claims his moral theory to be the morals of "economic man." Given this, how does Gauthier see his theory as responding to feminist concerns? His claim is that by building a moral theory without relying on human affective bonds, he has prescribed a society in which persons' feelings for each other are not a source of exploitation. He makes this claim at the beginning and end of *Morals by Agreements*. In the first chapter he writes:

the contractarian sees sociability as enriching human life; for him, it becomes a source of exploitation if it induces persons to acquiesce in institutions and practices that but for their fellow-feelings would be costly to them. Feminist thought has surely made this,

perhaps the core form of human exploitation, clear to us. Thus the contractarian insists that a society could not command the willing allegiance of a rational person if, without appealing to her feelings for others, it afforded her no expectation of net benefit (MBA11).

And then in the last chapter he writes:

The assurance that interaction is subject to the constraints of justice meets the concern emphasized in feminist thought, that sociability not be a basis for exploitation (MBA351).

Gauthier's argument here is that given the possibility of the exploitation of persons' affective capacities, we would be better off establishing morality without appealing to them. The argument I have offered here is that this may not be possible, much less desirable. While there is little doubt that persons have had their affections for each other exploited in unjust ways, it betrays a paucity of imagination on Gauthier's part to therefore propose that we dispose with those affections altogether. Certainly people's natural affection for each other can be used against them in exploitative ways. For the purposes of constructing moral theory, Gauthier jettisons social bonds *per se*, rather than examining the particular social bonds and particular social circumstances in which those bonds are played out, and how they have led to exploitation. Gauthier then appeals to feminist thought to buttress his claim of the righteousness of contractarian morality. Unfortunately Gauthier does not offer references to feminist writers when making the claim that his arguments for contractarian

morality are consistent with feminist concerns about the exploitation of women.⁴

In fact, feminist ethics points to the enormously rich potential for moral theory of attempting to build ethics upon the foundation of the great deal of affection, trust, mutual concern and compassion which in fact does exist between persons in at least some instances.⁵ Gauthier instead, in doing away with affection between persons as relevant to moral theory building, throws out the baby with the bath water. While affection between persons has been a basis of exploitation, not least of all for women, this need not condemn it to the category of irrelevance for human moral concerns.

In fact, far from being only a source of exploitation, persons' affection for one another can be liberating. Personal relationships based on affection and fellow-feeling can enrich our common lives in ways which make them quite indispensable to the human condition. Friendship, as I shall argue in the next chapter, is a compelling example of this claim: that affection and sociability need not be taken only as opportunities for exploitation, but can also be understood as integral to our possible liberation.

Notes

1. All MBA references refer to (Gauthier 1986).
2. This argument parallels David Lyons's argument that the rules in rule-utilitarianism do not save utilitarianism from the charge that it cannot take into account claims of justice, rights, and the like. Lyons argues that as long as the utilitarian really intends, through her actions, to maximize overall utility, then she will have to be willing to break the rules which generally maximize utility when following such a rule would fail to maximize overall utility. He goes on to argue that the implication of this is that rules must be evaluated in each possible instance of their application in order to determine whether following it in that instance will maximize overall utility or not. Lyons concludes, therefore, that rule-utilitarianism is extensionally equivalent to act-utilitarianism. (Lyons 1965) J.J.C. Smart, a utilitarian who agrees with Lyons's argument, writes: "to refuse to break a generally beneficial rule in those cases in which it is not most beneficial to obey it seems irrational and to be a case of rule worship." (Smart 1973, 10)
3. Note that while Gauthier, in the Preface to MBA explains that he used a randomizing device to assign gender to pronouns (in order to avoid the clumsiness of 's/he' or 'him/her'), nonetheless, when it comes to assigning gender to characters in these stories, the division of labor falls neatly in line with traditional gender roles. Jonathan is the potential rapist, Joanna the potential rape victim. Jonathan is the hunter, Joanna the gardener, and it is only Joanna's belly which swells. Baier asks "Is the 'appropriately programmed randomizing device' that Gauthier on p. vi says is responsible for his use of 'he' and 'she' responsible also for the constrained rapist's gender?" (Baier 1988, 324)
4. Jean Hampton's essay (which was published after Gauthier wrote Morals by Agreement), "Feminist Contractarianism," argues that contractarianism can be feminist precisely because it can be used to attend to the problem of distributive justice for women. Hampton argues that distributive justice is the essential problem of all human relationships and that therefore a version of contractarianism, which is both Hobbesian and Kantian, is necessary to ending women's oppression. See (Hampton 1993).
5. See for example, (Held 1993), (Baier 1994), (Schwarzenbach 1994) and (Tong 1993).

Chapter Four
Trust and Friendship

According to Gauthier, morality can be established without assuming that genuine goodwill, trust, or benevolence exist between persons. Persons are assumed to have affective bonds to other persons: nonetheless, they are characterized, for purposes of agreeing to a social contract, by their mutual unconcern. Even if some of their preferences include that other persons do well, this is only accidental to their description. These preferences are "non-essential, a sentimental residue from the past, or an emotional patina which affords a more pleasing aura to an otherwise bare artifice" (Gauthier 1977, 330-331). It is not for the sake of other persons that they act: they act only to secure their own greatest preference-satisfaction--this is what it means for them to be rational. (§7) As has already been discussed, according to Gauthier, there are no prior social relationships upon which the feasibility of contract, and hence social cooperation, rely. In the final chapter of *Morals by Agreement*, describing the liberal individual, he writes:

She is not bound by fixed social roles, either in her activities or in her feelings. Although social

affective relationships are essential to the liberal individual, *there are no essential social relationships* (MBA347) (emphasis added).

According to Gauthier, while persons may be understood to be social, and bound by affective ties to one another, those ties are the 'icing on the cake'; they are not necessary to the construction or development of cooperation between persons. "Only the relation of hostility is excluded from the scope of contract, and only it is natural to man" (Gauthier 1977, 329). Just as for Hobbes, Gauthier does not presuppose any antecedent social conditions between persons which make cooperation between them possible. Given the definition of trust developed in the first chapter, it follows that persons cannot trust one another in the sense that they believe that others have at least some minimal goodwill for them, such as the desire not to harm them for their own sakes. Hobbes and Gauthier each presume that the value that persons have for each other is purely instrumental, given their shared assumption of exclusive self-interest. There is no trust, mutual concern, or benevolence existing between persons prior to the morals by agreement which Gauthier seeks to establish.

One striking feature of both Hobbes's and Gauthier's descriptions of the moral agents at the heart of their theories is the fact that most of us would not want to be friends with them. The descriptions offered of them hardly

make them viable candidates for friendship. Michael Stocker writes that egoists:

cannot act for the sake of the intended beloved, friend, and so on; thus, they cannot love, be or have a friend, and so on. To achieve these great personal goods, they have to abandon that egoistical motive (Stocker 1976, 457).

In the same vein, Lawrence Blum writes:

A truly selfish person could not have friends in the fullest sense. If he were genuinely able to care for another person for his own sake, if he were able to give much of himself to the other freely and for his own sake, based on a genuine understanding of him, then he would not be selfish (Blum 1993, 197).

Hobbes's and Gauthier's moral agents, after all, are self-interested, rational, utility-maximizers--hardly qualities we look for in our friends. While these moral agents could have established, long-term preferences that others do well, the well-being of others is not--and on pain of irrationality cannot be--for the others' own sakes. Joanna, for example, having no genuine reason to trust Jonathan, because he has no genuine goodwill for her, would be discouraged from becoming his friend. And any attempts on his part to 'befriend' her will understandably be met with a great deal of suspicion and wariness on Joanna's part, given Jonathan's translucence (§8). And likewise, Jonathan would be ill-advised to befriend Joanna. Since, as has already been argued, we can trust neither Hobbes's nor Gauthier's moral agents (§8 and §5 respectively), we cannot imagine being friends with them, or imagine them being friends with

each other, for certainly, as will be shown, friendship depends on trust.

Gauthier, nonetheless, envisions a place for friendship in society based on his account of morals by agreement. First, he establishes the essentially just society grounded in the mutual pursuit of individual utility-maximization by delineating the rational reasons for constraining the maximization of individual preference-satisfaction. This paves the way for cooperation between essentially self-interested persons. Having done this, he argues that there is then room for "a genuine civic friendship to blossom" (MBA348). Because justice has been established solely on the basis of the rational pursuit of enlightened self-interest, friendship, and all the other relationships grounded in affection and mutual concern, need not play any constitutive role in justice or morality. Rather, it merely functions to enrich and make more attractive the essentially just society.

Gauthier's conception of the moral agent as the self-sufficient, atomistic, liberal self who relies on no one but himself is a person who may have attachments to others, whether of friendship or some other sort, but need not in order to establish contracts, and hence morality, with others. The liberal self may have attachments, but she need not have them--they are not essential to her, only accidental, purely optional. Thomas writes that therefore,

a theory such as Gauthier's "has no space at all for the invaluable good of trust, and thus we could not be [contractors] without incurring a considerable loss" (Thomas 1990, 236). And just as it has no place for trust, it will be shown that neither does it have a place for friendship, and so we could not be the contractors which Gauthier enjoins us to be without giving up the opportunity of being friends.

Friendship is, however, essential. In this chapter, I will describe an alternative to Hobbes's and Gauthier's diminished understanding of human relationships based solely on motivations of pure self-interest. I will argue that friendship, whether understood in personal or political terms, is essential to human flourishing, both individually and socially. I will also argue that the trust defined in Chapter One is a necessary, albeit not sufficient condition of friendship. And if friendship is essential to human flourishing, and trust is essential to friendship, then trust is thereby also essential.

§9. Friendship as an Exemplary Trust Relationship

Friendship can take a number of forms. We can have friends and be friends for a variety of reasons. Aristotle distinguishes between three types of friendship: those of utility, those of pleasure, and what he refers to as

"perfect friendship" (Aristotle 1156b7). Friendships founded on either utility or pleasure are incomplete. Aristotle argues that they exist and continue to exist only with reference to the useful or pleasant goods gained to the friends through them. Perfect friendships, on the other hand, are more stable. These friendships are characterized instead by the mutual love of the friends' characters. Friendship, according to Aristotle, is constituted by reciprocal goodwill. "Goodwill, when it is reciprocal being friendship" (Aristotle 1155b33). This is the friendship of people who wish good for each other independent of gain for themselves. These friends view and treat each other as ends in themselves and not as means to their own ends. And persons in these friendships then, "are most truly friends" (Aristotle 1156b10).

Neera Kapur Badhwar, in "Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship," distinguishes between two forms of friendship in a way quite reminiscent of Aristotle. She writes that there are "instrumental or means friendships" and "end friendships" (Badhwar 1991, 483). Her definition of the former is not significantly different from Aristotle's. Of the latter, she writes:

In an end friendship, one loves the friend as an essential part of one's system of ends and not solely, or even primarily, as a means to an independent end--career advancement, amusement, philosophical illumination, or greater happiness in the universe. In

such love, one loves the friend for the person she is, that is, for her essential rather than incidental features (Badhwar 1991, 483).

In the discussion of friendship here, I will adopt Badhwar's term of "end friendship" to refer to what Aristotle calls "perfect" friendship (or in other translations is sometimes referred to as "character" friendship). For the sake of simplicity, however, I will refer to "friendship" unqualified, using it to stand in for end friendship, unless otherwise noted. As I will argue, it is end friendship which necessarily involves trust, and, in fact, it is trust which can distinguish end friendship from mere instrumental friendship.

Friendship is constituted by a number of interdependent relations between people. Coincidence of interest, shared projects, pleasure in one another's company, intimate knowledge of one another, equality in terms of more or less reciprocal feelings towards one another (although it may be possible, but certainly not ideal, for friendship to be 'unrequited'), a generally shared world view or value system, i.e. comradeship, and, of course, mutual interest in the interests of the other, all contribute to the creation and maintenance of genuine end friendship. It probably does not have to be the case that all of these conditions must be satisfied in order for a relationship to constitute end friendship. But a sufficient combination of them must exist between two people for them to be friends.

One of the features which differentiates friendship from other types of human relationships is its voluntariness. Marilyn Friedman writes:

No consanguineal or legal connections establish or maintain ties of friendship. As this relationship is widely understood in our culture, its basis lies in voluntary choice (Friedman 1989, 298).

We may not choose our parents, our siblings, our children and to some extent, our colleagues, neighbors and the bibliophiles with whom we share the library. But we can choose our friends. Of course, at some point, we can choose whether or not to be friends with those people with whom we are involuntarily related, for example, our sisters and brothers. But the voluntary nature of the relationship in these types of cases is complicated by the history of involuntariness and emotional complexity inherent to such relationships. At their onset at least, friendships between people who are not otherwise related are free of such emotional complexity and histories. They are freely chosen and freely embarked upon. This quality of choice may be conscious or it may be more or less unconscious. It is conscious when we intentionally seek to form a friendship with someone and therefore seek out their company, and it is unconscious when we simply find ourselves more and more closely tied with someone although we never intentionally set out to become so. We can of course "fall into a deep friendship" (Thomas 1989, 102) in much the same way that we

can fall in love. We may be besieged by feelings of friendship for a person we would not have expected to respond to in such a way. But since we can do with those feelings whatever we choose, friendships are therefore voluntary. We may, for instance, choose not to follow up on them by establishing a friendship because we are presently too overwhelmed with other things to create and sustain a new friendship. Or we may not trust ourselves (§1 n7) to be competent towards them, and so choose not to take on the responsibility of friendship. We can choose to be friends or not to be friends with someone in response to feelings for them which are not necessarily voluntary. The friendly feelings we may experience for someone may not be a matter of choice; nonetheless, friendship is a matter of voluntary choice. As Aristotle writes of friendship, "mutual love involves choice" (Aristotle 1157b30).

This characteristic of friendship, its voluntariness, is what makes friendship perhaps one of the best and most interesting examples of a trust relationship. The trust which exists in an involuntary relationship (for example, between a parent and child) is to some extent itself involuntary. And such trust exists until there is reason to no longer accept it. And so the volition involved in such instances is to *trust no longer*. But in friendship the reverse is the case. The development of trust is itself voluntary, as it is implied by choosing friendship. And so

the volition involved in these instances is to *trust*. To choose to be friends with someone is to choose both to be trusted and to trust. It is to choose to be vulnerable to the decisions and motives of another and to be responsible for someone else being vulnerable to you. It is to believe that she has your best interests at heart and that she loves you for your own sake, and it is to take his interests to heart yourself and to love him for his own sake.

It need not always be the case, however, that the sequence is from the development of friendship to the volition to trust. There may, on the other hand, be instances in particular circumstances in which we trust persons who are not our friends, and only then develop friendships with them. In dangerous situations, for example, we may place a great deal of trust in persons with whom we have no previous personal relationship, and after trust has been established, become friends with them. People in the Nazi concentration camps may well have developed friendships with one another, which could last after their release, **after** having come to trust one another because their survival depended, at least in part on such trust. On the one hand, friendship involves the volition to trust. And, on the other hand, trust can give rise to friendship. Still, whether the trust precedes the friendship or develops in concert with the friendship,

friendship is naturally understood as being a matter of choice, as being voluntary.

Given this quality of friendship, its voluntariness, friendship is an exemplary case of a trusting relationship. More than perhaps any other relationship, it needs trust to exist. This is not of course to claim that other relationships do not require trust. My central argument throughout this dissertation is that relationships in which persons are bound to each other by obligations and expectations do need trust to get started and to thrive. My more particular claim here is simply that trust is especially central to the relationship of friendship.

§10. Friendship is Essential

Friendship is an intrinsic human good. It is the stuff which makes life tolerable at its worst moments and joyous at its best. Human life without end friendship would be empty and hollow. Friendship can be understood on either a personal or a political level; we can have personal, intimate friends, and likewise we can be civic friends to people who are otherwise strangers to us. Whether we look at friendship personally or politically, I will argue that it is inseparable from the good life.

Marilyn Friedman, in "Feminism and Modern Friendship: Dislocating the Community" argues that personal friendship

is a positive human value because of its potential for social change and disruption of unchosen communities which often oppress the people in them. She writes:

friendship, more so than many other relationships, can provide social support for people who are idiosyncratic, whose unconventional values and deviant life-styles make them victims of intolerance from family members and others who are involuntarily related to them. In this regard, friendship has socially disruptive possibilities, for out of the unconventional living which it helps to sustain there often arise influential forces for social change (Friedman 1989, 298).

Friendship occupies a morally important place in the human scheme of things because it can offer refuge to persons who may have nowhere else to turn and because it can, therefore, disrupt oppressive forces, giving rise to social change. C. S. Lewis describes friendship's capacity for disruption when he writes:

The pack or herd - the community - may even dislike and distrust [friendship]. Its leaders very often do. Headmasters and Headmistresses and Heads of religious communities, colonels and ships' captains, can feel uneasy when close and strong friendships arise between little knots of their subjects (Lewis 1960, 40).

Friendship makes for safety in otherwise hostile environments. For instance, friendships in the gay, lesbian and bisexual communities are of tantamount importance, for their members have historically been excluded from their original families, traditional communities, etc.¹

Friendship plays this morally crucial role for non-traditional people precisely because it is not constrained by preordained expectations and social rules. It is more

accommodating and can better fit the needs and desires of persons. Thomas writes:

Deep friendship and romantic love are the only two forms of interpersonal relationships in which the involved parties interact intensely and frequently, but yet, aside from the rules of morality, the nature of that interaction is not defined by this or that set of social rules (Thomas 1989, 100).

In friendship, we need not conform our behavior to particular rules or socially constructed patterns as much as we may need to in other relationships. Friendships are tailored by the persons in them through an exchange of thoughts, feelings, expectations, etc., and so they are more tolerant of persons who may not be tolerated in other kinds of more socially structured relationships.

Thomas has argued that end friendships (he uses the term "companion friendships" (Thomas 1989, 97)) are perhaps the most requisite for moral flourishing. Because end friendships are what he calls "minimally structured" (Thomas 1989, 100), friends come to have a "commanding perspective" over one another's lives. He writes:

deep friends come to have a commanding perspective of each other's lives, since the interaction between them is minimally structured. This is important because it points to why such friends are able to contribute significantly to each other's self-understanding (Thomas 1989, 119).

Free from the constraints present in so many of our other personal relationships, friends are more able to affect one another. Friends can challenge each other to become better people by aiding in their self-understanding, thereby

pointing out to each other both their strengths and weaknesses. "Friendship becomes a way of knowing as well as liking" (Sandel 1982, 181).

We cannot, it seems safe to assume, grow and flourish as moral agents without self-understanding, that is, without some insight into our dispositions and motivations. Self-understanding requires in the first place self-examination. And self-examination, if not in fact required, then is at least greatly aided by dialogue.² Persons, therefore, through conversation, can greatly contribute to their friends' self-understanding. And so friends, by contributing to one another's self-understanding, thereby contribute to one another's moral flourishing in reciprocal and more or less equal ways. Without friends, persons would not be presented with the possibility of having their motivations probed by someone who loves them for their own sakes, and therefore being morally challenged in the ways that they are. Because our end friends love us for our own sakes, we are therefore more likely to be able to accept their criticisms of us, since "loss of pride is thus held at a minimum" (Thomas 1989, 144). When our friends point out our failings, we are more able to respond constructively to them because, knowing that our friends love us nonetheless, we do not suffer humiliation. Implicit in the ideal of friendship, albeit not always realized, is the expectation that our friends can make us better. Kant writes:

We all have a strong impulse to disclose ourselves, and enter wholly into fellowship; and such self-revelation is further a human necessity for the correction of our judgements. To have a friend whom we know to be frank and loving, neither false nor spiteful, is to have one who will help us to correct our judgement when it is mistaken (Kant 1775, 206).

Lawrence Blum describes friendship as a basically moral phenomenon. Friendship, he argues, gives persons opportunities to be altruistic, to act for the sake of someone else. In fact, more than simply offering the possibility for altruism, friendship necessarily involves altruism, as friendship is in part constituted by the fact that friends do things for each other for their friend's own sakes. Blum writes that "friendship is an expression of moral activity on our part--of a type of regard for another person, a giving of oneself, and a caring for another for his own sake" (Blum 1993, 199). The altruism which is in part constitutive of friendship is itself a moral activity; "all friendships are morally good to the extent that they involve a genuine concern with the good of another for his own sake" (Blum 1993, 208). End friendships, being defined as reciprocated genuine goodwill, are therefore instances of moral activity. End friendship involves, by definition, altruism, the wishing of well to another for the other's own sake.

Similarly, Thomas claims that, besides being morally good in that it contributes to the friends' moral flourishing, friendship is also morally good in that it

prepares us to treat others, with whom we may not be tied by affection, morally as well. "The moral sensibilities that a flourishing companion friendship realizes enhance the quality of our moral interaction with others" (Thomas 1989, 155). This is so precisely because friendship requires us to live "in accordance with the precepts of an altruistic morality" (Thomas 1989, 155). Friendship can, when it is good, make us morally fit. It can make us capable of participating well in the sorts of relationships which define the morally good life.

Given the centrality of friendship in a system of human ends and goods and the necessity of trust to friendship, trust must be accorded recognition as a topic for moral theorizing. Friendship is bound by trust. Trust is a primary and necessary constituent of the friendships which we most value in our lives and in human interaction. In fact, as I will argue, it is the defining feature which distinguishes primarily instrumental friendships from the truly moral end friendships in which we are loved for our own sakes.

§11. Trust Essential to Friendship

In light of the descriptions just offered of friendship, it should be clear that trust³ is an essential

ingredient in any genuine end friendship. Aristotle, making it clear that trust is necessary to friendship, writes:

But it is natural that [perfect] friendships should be infrequent; for such men are rare. Further, such friendship requires time and familiarity; as the proverb says, men cannot know each other till they have 'eaten salt together'; nor can they admit each other to friendship or be friends till each has been found lovable and been trusted by each (Aristotle 1156b25).

The friendship of the good... is proof against slander; for it is not easy to trust any one's talk about a man who has long been tested by oneself; and it is among good men that trust and the feeling that he would never wrong me and all the other things that are demanded in true friendship are found. In the other kinds of friendship, however, there is nothing to prevent these evils arising (Aristotle 1157a20).

Without trust, there can be no end friendship.

We entrust to our friends different things. We trust them to do certain things for us, to act in particular ways towards us. We trust our friends to take care of us when we need to be cared for, to lend us a helping hand when we are in need of it. We trust our friends not only to be present in our lives when we are in need of help due to misfortune but also to share in our good fortune. And given the reciprocal quality of friendship, we trust that they will let us care for them when they need us to, and allow us to share in their good fortune when it comes their way. We trust our friends to help us be the kinds of persons we want to be, and to help us maintain the level of self-reflectiveness necessary to our growth and development as human beings. Perhaps most importantly, we trust that our

friends are our friends for the right reasons; we trust that they have the right intentions towards us.

One of the most common trusting expectations we can have of our friends is that they will help us when we are in need. When we need someone to help us move from one apartment to another, or to take care of our cat while we are on vacation, or to study for an exam, we expect that our friends, if they can, will help us. We also expect them to help us in more drastic circumstances. For example, if we suddenly take ill in the middle of the night, we expect that we can call our friends, wake them up, and ask them for the aid we need.

We also trust our friends to share in our fortunes, not just our miseries. We expect our friends to help us to celebrate our accomplishments and share in our joys. We want our friends to be the people to whom we can turn to share some good news, not only to be the people to whom we turn when we have bad news.

Friends also trust one another to advise each other on important matters. They trust each other to give such advice, however, from a particular point of view: the point of view of what is good for her friend, not what is good for herself. For example, if Sagel trusts Augusta to give her advice, she trusts that Augusta will give the advice from the point of view of what is good for Sagel, whether or not it is good for Augusta as well. Imagine that Sagel has been

offered a job in a distant state and she asks Augusta for advice on whether or not to accept it. If Sagel believes that their friendship is an end friendship, then she will trust Augusta to have her best interests at heart and to give advice for her own sake. If Augusta advises Sagel to not take the job, then Sagel trusts that she gives such advice because Augusta believes that taking the job is not in Sagel's own interests. Sagel trusts that Augusta would not advise her not to take the job simply because Augusta does not want her to move away, that is, because it is in Augusta's interests that Sagel not take the job. And if Sagel discovered that August advised her not to take the job **only** because Augusta did not want her to move away, Sagel might very well feel betrayed. Sagel might very well feel that, whereas she had taken Augusta to have genuine goodwill for her, to count **her good** as something that matters, in fact Augusta had only instrumental goodwill for her, only valued her as a means to her own ends. This is the heart of end friendship: we trust in such a relationship that our friends have our own interests at heart and do not value us only as means to their own ends.

An important quality of friendship is that it is reciprocal. We trust not only that our friends will let us support them but also that they will support us. Likewise, we trust that our friends will not only help us but also allow us to help them. We trust our friends to call us when

they need us. And if we discover that a friend has allowed himself to flounder alone, we may feel betrayed that he did not trust us enough to call upon us. We trust our friends to solicit our help because by so doing they express their trust in our abilities to come to their aid. This reciprocal component of friendship means that we want to be able to help our friends as much as we want them to be able to help us.

Imagine, for example, that Eli and Owen are friends. They have been friends for more than a year, during which time they have grown close and come to trust one another. One day, Eli calls Owen to tell him that he has just received some distressing news: Eli's sister is ill and Eli needs to go and be with her in the hospital and help take care of her children. Eli asks Owen if he will take care of Eli's apartment while he is away, watering his plants, picking up his mail, and so forth. Owen is glad to be able to help, and while Eli is away, he looks after Eli's affairs and calls Eli regularly to offer emotional support. Eli is grateful to have such a friend. Some months later, after Eli's sister has recovered from her illness and Eli has come home, Owen is in need of help. He has been laid off from his job. He needs to find new work, and in the meantime is in financial straits. Eli offers his assistance, both by offering to loan Owen some money (which Eli can well afford) and by helping him to find a new job. But Owen refuses

Eli's help, insisting that he does not want to impose. Eli is sensitive to the possibility that Owen is uncomfortable accepting money, and therefore tries to be of help in other ways. But still, Owen politely refuses help of any kind from Eli. It becomes clear to Eli that Owen is not simply being polite, he really does not want Eli to help. In a situation like this, Eli could easily feel that Owen had betrayed a fundamental trusting expectation of their friendship: the expectation that they would help **each other** when either one was in need of help.

Reciprocation of assistance is important to friendship. It keeps friends from feeling as if their friend is really in fact their benefactor. It is important that a friend be willing to accept help, and not only to offer it, for otherwise they make their friend into their mere benefactor, and "the role of benefactor always remains accidental, even a little alien, to that of Friend" (Lewis 1960, 44). Kant writes:

A friend who bears my losses becomes my benefactor and puts me in his debt. I feel shy in his presence and cannot look him boldly in the face. The true relationship is canceled and friendship ceases (Kant 1775, 204-205).

Ideally, friends do not want to be in debt to one another. Just as they do not want their friends to use them, to treat them only instrumentally, neither do they want to use their friends.

More importantly perhaps than trusting our friends to be with us in the ups and downs of our lives, we trust them to help us safeguard our moral characters. We trust them with our well-being, moral and otherwise. According to Thomas, friends are central to our moral flourishing because it is in the context of friendship that we are able to examine our own moral motivations and characters. But this is not possible without trust. Without trust, we do not open ourselves up to other persons in ways which give rise to their "commanding perspectives" of our lives.

Thomas writes, "Companion friends have a commanding perspective of each other's lives. Hence, deep trust and spontaneity of interaction are characteristic of such friendships" (Thomas 1989, 149-150, emphasis added). Here, Thomas implies that trust follows from or grows out of the deep knowledge which our friends have of us. But in fact, trust and such deep knowledge are mutually dependent and reinforcing characteristics of friendship. If we did not trust some people, we would not reveal ourselves to them in ways which allow them to gain such privileged knowledge of our lives. Likewise, the more a person is able to perceive our motivations, intentions, etc. in ways which strike us as correct, the more we are prone to trust them, assuming of course that we judge that they have genuine goodwill for us and that therefore, they will not use their insight into our characters to take advantage of us or others. So the deep

knowledge which our friends have of us and we have of our friends depends on our trust of them and their trust of us, and vice versa. In contrast to Gauthier's understanding (§7), our translucency to other persons and our trust of them are mutually reinforcing characteristics of our friendships. Were it not for the fact that we trust some people, our friends would not be able to come to have the knowledge of us which is so crucial to their contributing to our moral flourishing. We trust our friends, and so allow them to care for us, to express their concern for our well-being, moral and otherwise. The contribution to our moral flourishing which Thomas rightly attributes to friendship would not be possible without trust.

The most important thing which we entrust to our friends is the friendly relationship itself. Trusting involves the beliefs that a certain state of affairs holds. (§3) In friendship, it is the particular belief that our friend is our friend for particular reasons. Essentially, it is the belief that our friendship is an end friendship, and not merely an instrumental or means friendship. Trusting our friend as an end friend implies our belief that she has our best interest at heart, that she wants the best for us for our own sake. For Sagel then, this involves trusting that Augusta will advise her from the perspective of what is good for Sagel, and not just from the perspective of what is good for Augusta. And it involves the belief

that the continuation of our friendship is not dependent on the continuation of the good consequences which accrue to either of us as a result of our friendship. We trust that our friends do not merely value us as instrumental friends. We trust that we are loved for who we **are**, not for what we do. In essence, we trust that we will not be abandoned as soon as our usefulness or pleasantness comes to an end.

This, of course, is not to deny that the pleasures and good consequences inherent to friendship are important. No doubt, friends are useful and pleasant. Aristotle writes: "good people too are pleasant to each other... [and] the good are also useful to each other" (Aristotle 1157a1). But it is not this that defines end friendship.

Baier writes: "The most important things we entrust to others are things which take more than noninterference in order to thrive" (Baier 1994b, 103). I would add, perhaps tautologously, that the most important things we entrust to others are the things which are most important to us. So, one of the most important things which I entrust to my friends is the preservation and maintenance of my relationships with them. While I trust them to water my plants and feed my cat while I am away, more importantly I trust them to be genuine and sincere in their dealings with me and so to be the kind of friends I take them to be. In essence, I trust them to be trustworthy. In other words, we

trust them to continue to be the persons we take them to be when we evaluate them as trustworthy.

Thomas writes:

When deep trust is involved, the description under which we trust a person is of the utmost importance, as it is precisely because a person has revealed himself to us in a particular way that we are moved or disposed to trust him (Thomas 1989, 145).

Possibly the deepest betrayal we can experience in our friendships is to discover that our friends, whom we thought to love us for our own sakes, were in fact only friendly with us because it served their own ends. To discover that someone whom we took to be our complete friend is in fact only our 'friend' because she hopes that through her association with us she can get a promotion at work, or do well on an exam by studying with us, or fulfill some other end of her own, and so forth, is to have our trust betrayed in one of the deepest possible ways. What we ultimately trust our complete friends with, then, is the very intentions which we take them to have in being our friends.

Trust is essential to friendship. End friendship only survives when the parties to it can truly trust one another--trust that the other has genuine goodwill for them. Trust is a necessary, albeit not sufficient, condition of end friendship.

§12. Civic Friendship and Social Trust

The accounts of friendship discussed so far center on friendship as a personal relationship. But friendship need not only be understood in personal terms. Friendship can also be understood as a political relationship which serves as the basis for citizens' relationships to one another.

John M. Cooper, in "Political Animals and Civic Friendship," fleshes out Aristotle's notion of "civic friendship." Civic friendship, Cooper tells us, is the relationship which ties together members of a community. He writes:

Like all relationships deserving the name 'friendship,' civic friendship involves mutual good will, trust and well-wishing (Cooper 1990, 319).

In civic friendship, according to Cooper, we not only value the characters of our fellow-citizens because our self-interest is furthered by their being good; we value their good characters as ends in themselves. We wish our civic friends well, albeit to a significantly lesser degree than we do for our personal friends, for their own sakes, not merely as instruments for our well-being; their well-being is something which matters to us *per se*.

Each expects his fellow-citizens in their dealings with him (political, economic, and social) to be motivated not merely by self-interest (or other private particular interests) but also by concern for his good for his own sake (Cooper 1990, 319).

Civic friendship then, is analogous to personal friendship. Both involve feelings of goodwill, well-wishing and non-instrumental intentions. Personal friendship is, of course, personal. It involves people bound by ties of affection, intimacy and knowledge of one another. Civic friendship is not personal, as it can and does exist between people who are otherwise strangers to one another. Personal friends engage in *personal* activities; they share some common set of interests, they do things together, etc. Civic friends, on the other hand, engage in *public*, political activities together; they contribute together to the well-being of their polity.⁴ Civic friends, for example, participate together in parent teacher associations, serve in volunteer fire departments, donate blood, coach their local little league teams, and serve on juries. Cooper writes:

Active participation in a city's life is that single function which all the human beings belonging to that city perform together, and in the performance of which their character as political animals consists (Cooper 1990, 310).

Much like personal friends, civic friends *do* things together. Personal friends and civic friends both act in ways to create and maintain their common lives. Personal friends care for and value the characters of their friends, and, although perhaps to a lesser degree, so do civic friends. Civic friends, in caring for the well-being of the community in which they live, cannot help but care as well

for the characters of the persons with whom they share that community.

Sibyl Schwarzenbach, in "On Civic Friendship," has also argued that there is a political sense of friendship distinct from, although analogous to, a personal sense of friendship. She argues that friendship, or *philia*, is a form of "reproductive activity." Whereas productive activity aims at the creation of products, whether material or otherwise, reproductive activity aims instead at the reproduction of the human relationships which constitute society. Reproduction is here not merely biological, but political or ethical. Schwarzenbach, describing such reproductive activity, writes that it includes

all those rational activities (thinking about others and their needs, cooking for them, buying them gifts, etc.) which go towards reproducing a particular set of relationships between persons over time--in the best case relations of *philia* (Schwarzenbach 1994, 5).

This ethical or political sense of reproduction, or "political reproductive activity" (Schwarzenbach 1994, 5), then gives rise to a form of friendship, which while different from personal friendship, is nonetheless in some important respects quite similar to it. It is the basic relationship of civic friendship at which, in its political or ethical sense, "reproductive activity" aims. She writes:

The central difference between personal and civic friends turns out to be that among civic friends intimate knowledge and a close emotional bond are absent. In political friendship that is, the traits of "mutual awareness", of "wishing the other well for

their own sake", and of "doing things" for the civic friend are still retained (to whatever degree) only now they are evidenced in a *general concern* (Schwarzenbach 1994, 8).

Civic friends express their genuine goodwill for one another through the medium of public norms. Civic friends accept and maintain a general concern for each other, and act on such concern in concrete ways with each other. The basic understanding of friendship as persons wishing well for each other for the other's own sake is the same whether the friendship is understood in personal or in political or civic terms. Personal friends know each other intimately; civic friends do not. But both personal and civic friends value their friends' well-being for their friends' own sake, although perhaps to significantly different degrees.

An important claim which both Cooper and Schwarzenbach make is that the concept of civic friendship, as advanced by Aristotle, need not be taken to be merely a relic of the ancient world. The modern state is informed by the ideal of friendship between citizens, albeit in a weaker and less recognized way, than was Aristotle's Athens. Cooper points out that the moral characters of United States citizens matter to other citizens of the United States. Citizens of the United States take a certain kind of pride in the characters of their fellow citizens, which they do not take in the characters of persons of other nations. Cooper writes:

There seems no denying that ordinary Americans, for example, are characteristically quite a bit concerned about the moral standards of people prominent in government, business and industry, and concerned in quite a different way from the concern they presumably also feel about the morality of people in similar positions in foreign countries, even ones with which the United States has extensive trading and relationships (Cooper 1990, 315).

While persons may have a certain concern for the characters of persons in other countries, this is tied to concerns they have for themselves; for example, they wish not to be taken advantage of by people with whom they engage in trading relations, nor do they wish to be the victims of aggression. When it comes to people of their own country, however, the concern runs deeper. They want their fellow citizens to be good for their own sakes. Cooper maintains that "they also care what kind of people their fellow-citizens are. They want them to be decent, fair-minded, respectable, moral people" (Cooper 1990, 314).

Schwarzenbach argues that the modern concept of universal individual rights can actually be understood as giving expression to civic friendship. Such rights can be seen as based on a general regard for the well-being of the members of a society. "Universal individual rights and the general regard for them emerge as a necessary aspect of any adequate modern conception of civic friendship" (Schwarzenbach 1994, 16). Recognizing and giving weight to individual rights through the medium of a constitution may then be understood as an expression of civic friendship. It

embodies the ideal that members of a society should be concerned for the well-being of each other as *civic friends*. So the modern state, with its emphasis on rights and the equality of its citizens, while not explicitly recognizing the notion of civic friendship, nonetheless embodies it, at least to some extent, in its foundations. Schwarzenbach writes:

It is... neither mandated by the size, nor the new complexity, of the modern nation-state that political friendship between citizens becomes impossible. Rather, it seems that the notion of *philia* has faded into the background of modern political analysis where it may still be found to operate, but in a subtle and near invisible form (Schwarzenbach 1994, 16).

Civic friendship then, is not just an historical curiosity about which we can express nostalgia. Rather, it is still present, if not easily seen, in the relationships between fellow citizens in modern nation-states. It informs the relationships between persons who, while otherwise strangers to one another, nonetheless are invested in the well-being of their fellow citizens. When civic friendship, the basic, genuine goodwill that the citizens of a state have for each other, is not present, the viability of that state is threatened. As has been argued in the cases of both Hobbes and Gauthier, when persons are initially conceived as having nothing but instrumental goodwill for one another, it is not clear that they can formulate the contracts or agreements which make possible their cooperation. Similarly, the ongoing viability of the contracts people make exists, when

it does, because people have enough genuine goodwill and therefore enough **trust** for each other to continue to abide by their agreements, believing that others will as well.

The trust which is necessary to civic friendships is a form of *social trust*, which we could call 'civic trust.' **Because** we take our fellow citizens to be basically trustworthy, we can enter into the friendly relations that are necessary to create and maintain society. "The friendships on which the best political alliances are modeled can neither be created nor sustained without trust" (Code 1991, 183). Just as trust plays a constitutive role in personal friendship, so does it have a necessary place in civic friendship. Held writes:

At the level of society, there must be the possibility of social trust. At the level of persons in their immediate interactions with those close to them, there must be the possibility of personal trust (Held 1984, 62).

Trust is necessary in both personal relationships and political relationships: in personal friendship and civic friendship. The trust present in civic friendship is a thinner version of the trust we find in our personal friendships. We do not expect our civic friends, our fellow-citizens, to care for us in the particular ways we expect our personal friends to do, for example by listening to our woes or bailing us out of jail in the middle of the night. But the social trust which exists between civic friends is nonetheless analogous to the personal trust

necessary to personal friendships. We trust that we are bound to our civic friends by feelings of genuine, not purely instrumental, goodwill. For example, if we are being tried for a crime, we trust that the jurors, whose responsibility it is to determine our legal guilt or innocence, will carry out their responsibility out of a sense of fairness to us and in recognition of the norms expressed in the Constitution's prescriptions for criminal trials. We trust that they will not accept bribes, for example, or in some other way take advantage of their position. Cooper writes, "where civic friendship characterizes a population there exists, as a recognized and accepted norm, a certain measure of mutual goodwill, and also mutual trust, among the people making up the population" (Cooper 1990, 319). Civic friends, although they can be strangers to one another, nonetheless believe that their fellow-citizens do not bear them ill-will, and that in fact they bear them genuine goodwill. Civic friends trust that they will act towards each other in ways which express this goodwill and their respect and concern for each other's well-being. They do this by recognizing and respecting the public norms of the society which they share with their civic friends.

Social trust then, can also be expressed more generally than simply between individuals to one another. Social trust, and its counterpart, social mistrust, also gets

expressed through a society's principles and fair distribution of resources. Held writes:

A first requirement for general trust to be justified would have to be, then, a recognition, as a fundamental principle underlying social and political arrangements, that none of the members of the society should be denied what is needed for a decent life and for adequate self-development while the society makes it possible for others to have far more than they need (Held 1984, 76).

Likewise, social trust can be expressed through the inclusion of society's members in the activities of the political life. Cooper writes:

If those excluded from active participation in the political life of the city recognized their exclusion as an injustice, they would see themselves as being exploited by the rulers for the rulers' own benefit. Their trust would thus have been violated, and a natural consequence of an uncorrected violation of trust is its destruction. Since civic friendship consists in part of the mutual trust of the citizens that they are all effectively concerned for one another's good, the destruction of this trust destroys the friendship too (Cooper 1990, 322).

Genuine goodwill and mutual respect and concern for each other's well-being, i.e., social trust, can be generalized through a society's institutions and laws, as can social mistrust. If institutions regularly fail to protect persons' welfare and regularly violate persons' rights, those persons will have little reason to trust those institutions or the members of society who create and maintain them. Respecting one another's rights, both by individuals and by institutions, is an important means by

which civic friends can demonstrate their genuine goodwill for one another.

If society fails to provide fairly for the needs of its members, or if it does not accord an equality of opportunity to the benefits that society has to offer, then it seems reasonable for some of its members (namely, those who suffer from such injustice) to feel that they are not respected, that society is not concerned to further their well-being, and that their rights are not being protected. By excluding members of society from either the material resources or political process, social mistrust is fostered, civic friendship is eroded and the grounds for cooperation deteriorate.

Diego Gambetta has demonstrated the devastating effects of a lack of trust for a society. He argues that in an atmosphere of distrust, organized crime, i.e., the Mafia, flourishes. The Mafia, he argues, is in the business of "supplying private protection as a substitute for trust" (Gambetta 1993, 220). So where "public trust, the basis for a well-ordered society" (Gambetta 1993, 77), does not exist, the Mafia is able to exploit such mistrust to their own advantage. Because mistrust exists in a particular society, or culture (for example, Southern Italy), people feel threatened and unsafe. They are then likely to turn, if it is available, to private sources of protection, which the Mafia then happily provides. The continued success of the

Mafia then depends on the continuation of the conditions in which people seek private protection, namely, distrust.

Gambetta writes:

The mafioso himself has an interest in regulated injections of distrust into the market to increase the demand for the product he sells—that is, protection. If agents could trust each other independently of his intervention, he would, on this score at least, be idle. The income he enjoys are the benefits to him of distrust. Thus coping with, and at the same time recreating, distrust would seem to be the means by which the power of the mafioso has endured so long" (Gambetta 1988a, 173).

The price of distrust can be high indeed. The results of the Mafia's dominance in Southern Italy are high murder rates, the migration of the most talented people, who want to avoid the exploitation of the Mafia, low technological innovation, high rates of cheating, and so on (Gambetta 1988a, 173). Without public or social trust, society cannot flourish.

The trust which members of a society can and do place in their government and its leaders is an instance of the civic form of social trust. When we trust a particular person to represent us in Congress, for example, we take them to have genuine goodwill towards us, *qua* citizens. Political scientists report that "trust in a legislator is predicated on the belief that the representative is serving, and will continue to serve, the best interests of his or her supporters rather than acting out of self-interest" (Parker and Parker 1993, 443) We take them to be, ideally, our

professional civic friends. They need not know us particularly or individually to know that we, their constituents, take them to have the right intentions towards us--genuine goodwill--and their responsibilities to us. Abuse of such power, then, is felt as a betrayal of trust on the part of their constituency. Sissela Bok, in *Lying: Moral Choice in Public and Private Life*, argues that deception, for example, can be particularly problematic for those in government, "where those who deceive occupy positions of trust" (Bok 1978, 95). And more often than not, such betrayals involve elected representatives disregarding their political duties, the obligations of civic friendship, in favor of personally benefiting from their roles. In other words, what we trust, when we trust politicians (and likewise, the military, unelected government officials, etc.), is that they will not take advantage of the power which we have entrusted to them not to treat us instrumentally but to maintain genuine goodwill for us by safeguarding over our interests.⁵

It is not enough, however, simply to evaluate the characters of politicians and others in positions of power in order to trust them, since it is not just the character but "the position of ruler that is seen to create the danger" (Baier 1994d, 164). Therefore, we should also be concerned to establish and maintain roles which encourage trustworthiness and discourage untrustworthiness in the

persons who occupy them. Baier, concluding a discussion of the myriad ways in which we limit the power available to our leaders (for example, the Twenty-second Amendment, and the system of checks and balances prescribed in the Constitution), writes:

we are not entirely without guidelines on how to design roles for individuals that will help them avoid the worst forms of untrustworthiness, or of oppressively burdensome trust, or of overly vulnerable trusting (Baier 1994d, 165).

Even with such safeguards in place, which limit the amount of abuse of trust which is possible, we still must, in the end, simply trust those whom we elect to positions of power, and in turn, trust them to **appoint** other persons to positions of power, sometimes even, as in the cases of the Supreme Court justices, to lifetime positions of power. Ultimately, we must trust "their sense of our sense of the central importance of that particular bit of our total network of social and public trust" (Baier 1994d, 163). Democracy, it would seem, itself rests on trust. "For political power to serve the purposes for which men need it, it has to be subject to drastic and potentially destructive abuse" (Dunn 1974, 296).

Social trust explains why we can trust and why we are willing to make ourselves vulnerable to persons who are otherwise strangers to us. As civic friends, we often trust that strangers on the street will not try to thwart our travel plans by giving us the wrong directions or make us

late for our appointments by telling us the wrong time. Civic friends, to the extent that they trust one another, therefore expect that those persons will respect them and act out of concern for their well-being. The young man who asks me to watch his bag while he gets himself coffee takes me to be his civic friend, whether he would explain it that way or not. He believes that I have enough respect for him and concern for his well-being not to make off with his bag, which holds his valuable computer. Otherwise, he has little reason to entrust me with it. No doubt, his trust in me is not as deep a trust as we would expect him to have for his personal friends, family, or lovers--but it is trust just the same. It is social trust, exhibited by civic friends to one another, which provides the foundation for the countless ways in which people, who are otherwise strangers to one another, are willing to be vulnerable to each other, from watching each other's bags in coffee shops to carrying each other's children in airports.

Trust, however thinner a version it may be, is as necessary to civic friendship as it is to personal friendship. **Because** we take our fellow citizens to be basically trustworthy, we can enter into the friendly relations with them that are necessary to create and maintain society. Just as personal friends engage in personal activities together, civic friends engage in political activities together: they share in the civic or

public life--they vote together (although for many, not as often as they perhaps should); they call in to radio call-in shows and discuss public issues together; and they serve on juries together. And just as trust is a necessary, albeit not sufficient, condition for personal friends to be able to engage in those personal activities (such as contributing to each other's moral flourishing), so too is it a necessary condition for civic friends to engage in the activities which constitute their friendship.

'Civic trust' is not, of course, the only form of social trust. We can also trust those who are not our fellow-citizens. For all I know, the young man in the coffee shop who trusted me with his bag was only visiting from some other country, and so not my fellow-citizen. His trust in me need not be understood as an instance of 'civic trust,' but it is trust just the same. When we travel in foreign countries, we trust the persons we encounter, perhaps to different degrees than we do here (sometimes more, sometimes less.) Different countries have different "climates of trust" (Baier 1994b, 96), and so the amount of trust we have in our hosts (and guests) will depend on the extent to which we evaluate the people of other societies as trustworthy. But at the most basic level, we simply, I believe, trust other human beings *qua* human beings. We trust, given that they too live embodied existences more or less like ours, they will not display casual disregard for

our well-being. This is not to say that we cannot distrust people in other societies. But this is the exception, not the rule. We begin from a position of trust, 'default' trust if you will, and then, if we have reason to do so, move into a position of distrust.

Social trust at its most basic level is the simple trust we have for other human beings. At a somewhat thicker layer, we trust that our fellow-citizens, our civic friends, will express their respect for us and concern for our well-being by supporting and regarding the public, social norms which prescribe our common lives. The trust which constitutes personal friendship entails an even greater degree of concern for our well-being and willingness to extend ourselves towards the well-being of our friends. Without trust, absent these beliefs about why our friends are our friends, friendship, whether personal or civic, does not seem possible. Of persons we do not trust, we cannot believe that they have genuine goodwill for us, respect us, or are concerned with our well-being. And so while we may engage in an instrumental friendship with them, somewhat cautiously, or perhaps even foolishly, genuine friendship seems barred from development. Absent trust, friendship, whether personal or political, does not flourish. Being capable of friendship means having the right sorts of dispositions and intentions. It is necessary, therefore, to

pose, and perhaps more importantly to answer, the questions of when we *should* be trustworthy and when we *should* trust.

Notes

1. Friedman's discussion of the role of friendship is entwined with a discussion about the similar role that urban communities can play. She argues that modern urban communities also offer alternative means to find moral support for people otherwise alone because of their idiosyncratic natures or deviant life-styles. (Friedman 1989, 297-307) Her points about friendship and urban communities merge in the example at hand. Gays, lesbians and bisexuals survive through friendship and communities which are, by and large, found in cities. For a brief discussion of the role which 'community' plays in the lives of gays and lesbians, see Richard Mohr's *Gays/Justice: A Study of Ethics, Society, and Law* (Mohr 1988, 12-13). Janice Raymond, in her *A Passion for Friends: Toward a Philosophy of Female Affection*, discusses the ways in which women's friendships can be liberating because of their power to disrupt patriarchy. (Raymond 1986)

2. To reflect on the connection between self-examination and dialogue, one need only think of Socrates's claim that the "unexamined life is not worth living" and the fact that he went about the process of examination through dialogue with friends. Dialogue, he thought, would make them, and all Athenians, morally better.

3. In this section, unless otherwise noted, by "trust" I mean *volitional trust*, either specified or unspecified. *Evaluative trust* is simply to judge someone to have certain intentions, dispositions, and the like, whereas *volitional trust* is this judgment and the decision to act on this judgment. In friendship, we necessarily act on our evaluation of their trustworthiness, thus the trust which is essential to friendship is *volitional*. (For the definitions of *volitional trust* and *evaluative trust*, see §3).

4. That personal friends engage in **personal** or **private** activities together, and civic friends in **public** activities, implies a distinction between public and private per se. The implication here, however, is that the private and public spheres, at least with regard to friendship, are somewhat analogous. Civic friends, who engage in the activities of creating and maintaining the public sphere together, can be understood in ways quite similar to personal friends, who engage in more personal, and hence private activities. For discussions of the public/private distinction, see (Held 1984, 24-28), (Hampshire 1978) and (Benn and Gaus 1983).

5. Steven Kelman argues that many of the activities of Congresspersons are, contrary to much public opinion, in fact in keeping with what we would trust them to do. He argues that most representatives in Congress are in fact concerned to work for the

public good, and not personally benefit from their positions. See (Kelman 1990).

Chapter Five

Prescribing Trust

In the first chapter, I offered a conceptual analysis of trustworthiness and trust and discussed the psychological and normative ways in which trust is basic and defined trustworthiness and trust. Such a description sets the stage for the prescriptive questions which are to come. Trust is generally or for the most part a good thing, making cooperation between persons possible. Nonetheless, trust can also allow deceit and abuse to flourish. Thieves trust one another, and battered women often continue to trust their batterers. Care must be paid to the normative issues surrounding trust and trustworthiness. In the present chapter, I turn from the descriptive account of trust offered in Chapter One, and ask prescriptive questions: when should we be trustworthy? and when should we trust?

§13. When to be Trustworthy

I will begin with the first of these two questions: when should we be trustworthy? To review from Chapter One,

volitional trust was defined as evaluative trust and vulnerability. Evaluative trust, the belief that someone is trustworthy, is constituted by four separate beliefs. I can be said to have evaluative trust for someone when: *I believe that she: has genuine goodwill toward me, is competent to act in the ways which I trust her to act, could act in such a way as to fail to express her respect for me and concern for my well-being, and will not act in such a way.* In light of this general description of trustworthiness, it would seem that we should generally be as trustworthy as possible.

To be trustworthy is, at least in part, to subscribe to the Kantian edict "Act always so as to treat humanity whether in yourself or in that of another person, always as an end, never as a means only" (Kant 1785, 47). This is what it means to have **genuine** goodwill for another, as opposed to instrumental goodwill for another or goodwill for him in order that he may further your own interests. To have genuine goodwill for someone is to value her for her own sake, and not purely for what she can do for you or for someone else. It is to view her as intrinsically valuable and possessing inherent dignity, not just instrumentally or extrinsically valuable. If we should not value persons only in terms which relate to our own interests and only in terms of what they can do for us, then the first criterion of

trustworthiness, that of having genuine goodwill for others, is something to which we should always aspire.

The second criterion of trustworthiness, to be competent to care for others, is also something toward which we should strive. To be competent in this sense is to be sensitive to what another person needs from us in order to thrive and be well and to have the capacity to fill those needs. "Persons cannot increase the trust between them by misinterpreting each others' problems and pains" (Held 1993, 39). Understanding persons' individual and unique life-stories and being able to respond to the particularities of different persons is what is "at work in the competent moral agent" (Walker 1989, 18). Insofar as we should contribute to one another's well-being, we should be competent, and hence trustworthy, to care for them. To be competent is to be sensitive not only to the needs of the persons who trust us, but also to be sensitive to the moral complexities of the human situation. Baier argues that to trust someone is to give him discretionary or discriminatory power. She writes:

To be someone to be trusted with a promise, as well as to be trusted as a promisor, one must be able to use discretion not as to when the promise has been kept but, rather, as to when to insist that the promise be kept, or to instigate penalty for breach of promise, when to keep and when not to keep one's promise. (Baier 1994b, 118).

Being trustworthy in the sense of being competent, then, is to be able to discriminate, perhaps a great deal, between cases, and to know what a situation calls for.

Also, in some situations a general sense of competency can mean that we know what we are **not** competent to do. We should know our limitations, and knowing them goes some distance toward determining our more general competency. For example, if we were to take our car to a mechanic who was not competent to do the particular work required on it, we would expect her to be generally competent by admitting her lack of a more specific competence. And this display of general competence would make her all the more trustworthy.

The third criterion of trustworthiness is that it must be possible for the trusted to act contrary to the expectations of the truster. It must be in her capacity to act contrary to the way she is trusted to act. So, in order to be trustworthy, we must not be **incapable** of harming others. And of course, persons do have the capacity to harm each other. It is just this feature of the human condition which makes morality both necessary and possible. For if we are **incapable** of being untrustworthy, then it is not necessary to trust us. If we were different sorts of beings than we are, and were actually incapable of being untrustworthy, there would be no issue of trust between persons. It is the fact that people do express their respect for us and concern for our well-being despite the

fact that they often could act otherwise which leads Thomas to conclude that it is through trust that persons affirm one another's moral worth. He writes:

a person does not show us any respect in telling us the truth if he does so only because he believes that we already know the truth, whereas a person who tells the truth although she could have prevaricated without detection very much does (Thomas 1990, 250-251).

Because people can fail to respect us and fail to display concern for our well-being and do not, even when they may be able to gain from harming us and do so undetected, we can experience and appreciate their respect for us.

The last criterion of trustworthiness is the belief that those trusted will live up to the expectations of those who trust them. So in order to be trustworthy, or to be considered trustworthy, we should generate this expectation in others. We should convince others of our genuine goodwill for them and our competency to care for them. This is Thomas's understanding of trust as being intentionally manifested. In order to be evaluated by others as trustworthy, we need to demonstrate and make evident our intentions to value them for their own sakes and act so as to express those intentions. This is what we are called upon to be when we are called upon to be trustworthy. We can do this either verbally or through our actions. One of the most obvious ways in which we can demonstrate our trustworthiness to others is simply by being trustworthy. By establishing a history of goodwill and trustworthy

behavior, we convince others of our general trustworthiness. Of course, we need not have a history of goodwill with a person in order for them to believe us trustworthy. Such is the case with trust between strangers.

To summarize, the moral standard which trustworthiness calls upon us to live up to is to have genuine goodwill for others, be competent to care for them and generate the expectation in others that we will act in ways which express our respect for them. Given this, it seems hard to imagine why we should not always be trustworthy.

While trustworthiness is morally praiseworthy and something toward which we should always endeavor, the moral evaluation of that with which we are entrusted may nonetheless sometimes make it difficult to see just what trustworthiness calls upon us to do. Kleinig, making a parallel claim with regard to loyalty, writes: "What loyalty demands will depend not only on the object of loyalty, but also on the particular interests that are at stake" (Kleinig 1995, 150). In cases of apparent conflict, the trustworthy action, like the loyal, has to be discerned from the options at hand. While loyalty, a close cousin of trustworthiness, is in general good and praiseworthy, there are nonetheless instances in which loyalty can be trumped by other moral considerations. He writes:

We do not owe loyalty to those whose conduct has transgressed and undermined publicly acknowledged or professional standards any more than the sexually

abused child, physically abused wife or oppressed minority group owe loyalty to their abusers. The loyalty that might normally be expected has been forfeited by the abuser (Kleinig 1995, 150).

That we should never be disloyal, according to Kleinig, does not imply that we should always be loyal in each and every circumstance. Loyalty is a praiseworthy goal to which to aspire. Nonetheless, there are instances in which we can be 'non-loyal' to a particular person, without being disloyal, and therefore remain generally loyal persons. Trust is very similar to the way in which Kleinig describes loyalty.

While trustworthiness is laudable, we should not always be trustworthy to **everyone**, regardless of the worth of the objects which they entrust to us, or the reasons they want them protected by our trust. Being trustworthy in general may require that we be 'non-trustworthy' in some particular circumstances to some particular persons. In fact, while in some circumstances we may not be able to be trustworthy to some persons, we should never be **untrustworthy**. Being trustworthy and being untrustworthy do not exhaust all the possibilities. In other words, we can sometimes fail, and sometimes **should** fail, to be trustworthy, in some particular circumstance, without thereby being untrustworthy.

In some instances, it would seem that being trustworthy on the one hand requires being untrustworthy on the other. Such cases of conflict would seem to draw one to the conclusion that one must sometimes be untrustworthy. I

intend to argue that this is not the case. I will argue that in some circumstances, while we may be required to be 'non-trustworthy', we should never be untrustworthy.

To be never untrustworthy we have to develop and maintain the moral competency required for the discretion which trustworthiness involves. Given the contextual sensitivity of trustworthiness, it may sometimes be difficult to discern the trustworthy way to proceed. Being trustworthy involves **discretionary** power, which means having the competence to discriminate between what is worth safeguarding and what is not. Being trustworthy does not mean having to safeguard something which a person deems is in their interest when that something is itself morally suspect.

For example, imagine that a fifteen year old, Paul, is a good friend with Addie, a woman who is also a friend of Paul's parents. Paul respects Addie and often talks to her about things which are on his mind. One day, obviously upset, Paul approaches Addie, needing to talk to her about something important. He tells her he really needs to talk to someone, but that he can't talk to her unless she promises that she will not tell anyone, not even his parents, what he is about to tell her. Seeing that Paul is upset and needs to talk to someone, Addie promises to maintain his confidentiality. Paul then tells Addie that he is going away for the weekend with some older teenagers to a

state with a lower drinking age, while telling his parents that he will be at a friend's house in a neighboring town.

Addie is now confronted with a conflict. She has, on the one hand, promised Paul that she will not reveal what he has told her. On the other hand, she knows that Paul's parents trust her with the well-being of their son. It would seem that if Addie is trustworthy to Paul, by keeping his secret, then she is thereby untrustworthy to his parents, by allowing him to do something which they would judge (perhaps rightly so) not to be in his best interests. If Addie breaks her promise to Paul, then it would seem clearly that she is untrustworthy to him. But it is not as clear that if she does not break her promise, then she is untrustworthy to his parents. In trusting Addie to help guard over Paul's well-being, Paul's parents are granting to her discretionary power. They are trusting her to do what is in her best judgment good for Paul. And in this instance, Addie may well judge that it is in Paul's best interest to have someone, such as herself, to whom he can confide his problems with the assurance that they will remain confidential. In this instance, then, it seems that by caring for Paul by keeping her promise to him, Addie is trustworthy to both Paul and his parents.

Imagine however, that Paul confides to Addie, after she has promised not to divulge his confidences to anyone, that he is thinking about suicide. Imagine that Addie judges

Paul to be in real danger--that he could in fact attempt to kill himself. Not confident that alone she can give him the psychological help which he clearly needs, Addie is in a position where she must choose whether to maintain his confidence, and risk him killing himself, or divulge his secret to someone else (perhaps his parents, perhaps someone else). Here the question is--is she untrustworthy to Paul if she does divulge his secret? No. What Paul is asking to have protected in this situation, is not itself worth protecting. So, while revealing his secret may be considered by Paul to make Addie untrustworthy, it does not. Being trustworthy does not entail being required to do everything which a person who trusts us would want us to do. By revealing Paul's secret, (perhaps after first warning him that she intends to do so), Addie may, at the most, be non-trustworthy to Paul, in favor of being trustworthy to his parents (or in fact, it could be argued that she is being trustworthy to him, by keeping him from committing suicide and thereby protecting his interests). Addie must, in this instance, avoid being untrustworthy to Paul's parents. She can do this by failing to be 'trustworthy' to Paul. She is not, however, thereby untrustworthy to Paul.

Similarly, if Paul confided to Addie that he had overheard some other boys from his school (whom he knew) bragging about having raped some girls, Addie would not be untrustworthy if she refused to keep Paul's secret. It may

of course, be best if she could convince Paul himself to do the right thing, but if she could not, and he still insisted that she not reveal to anyone what he had told her, Addie would not be required by trustworthiness to keep his secret. The object entrusted to her, the secret that some boys had raped some girls, is not itself worthy of protection, and so she is not obligated by trust to protect that object. She may be required then, to be 'non-trustworthy' to Paul, but in doing so, she is not untrustworthy to him. She does not take advantage of him or treat him instrumentally, as a means to her own end.

Grant Gillet has argued that in the area of medical confidentiality, it is permissible for doctors to violate a patient's trust if that patient is himself acting in morally corrupt ways. He writes:

Trust is (at least in part) a two-way thing and can only exist between morally sensitive human beings. . . . A basic element of such moral attitudes is the responsiveness of the agents concerned to the moral features of human interactions (Gillet 1989, 125).

Part of being trustworthy then, is knowing to whom to be trustworthy, under what circumstances, and in regards to what objects. Trustworthiness is contextually sensitive. The essence of competence is knowing what is worth safeguarding and under what conditions it should be safeguarded. And being trustworthy requires that the one who would trust us themselves be trustworthy and live up to the demands of morality. To be trustworthy, we need to be

able to respond to people insofar as they are morally sensitive. Gillet writes that we need not be trustworthy to one who:

has moved outside the community of trust, mutual concern and non-maleficence in which moral considerations for the preferences of others have their proper place. It is not that one 'contracts in' to such a community, nor that one in any sense volunteers, but rather one is a *de facto* member of it by virtue of possessing those human sensitivities and vulnerabilities which give moral predicates their meaning and importance. Such weight as one claims for one's own personal privileges and moral principles--such as the demand for confidentiality--is derived from a 'form of life' where the interpersonal transactions which define trust, respect, harm and so on, are in play (Gillet 1989, 125).

The responsibility to be trustworthy can only make sense within a certain set of parameters inside of which people are sensitive to each other and bounded by mutually understood moral concerns. Not helping Paul keep the secret about the boys who raped some girls can then be understood as an attempt to keep him within the moral community which Gillet describes. By not keeping his secret in that case, the friend makes it more difficult for the teenager to stray from the community in which trustworthiness has its rightful place. And this may be one of the most important things we can entrust to our friends: to keep us morally fit and not let us stray from such a community.

Civic friends too, should never be untrustworthy. They need not, however, always be trustworthy. Understanding this can help keep civic friendship from becoming

nationalism. The social trust which in part constitutes civic friendship does not require that a civic friend do everything which her civic friends would want her to do. Persons cannot charge their civic friends with being untrustworthy if they sometimes act out of respect and concern for persons in other nations. Civic friendship, like personal friendship, does not require one to be untrustworthy to others in order to be trustworthy toward one's civic friends.

Being trustworthy does not always mean being perceived by others as such. If trustworthiness requires competency, and I have argued that it does, then it also means that in situations which involve apparent conflicts of interest, some persons may see differently which way our competency should lead us. In the most general sense, to be trustworthy is to be capable and worthy of those relationships which are themselves worthy of our attention and care. To be trustworthy is to be able to function well as a social being, connected to and interdependent upon others. Trust involves sensitivity to the needs of persons and the moral value of those things which they would have us protect, or entrust to us.

§14. When to Trust

Although in some instances we cannot be trustworthy to some persons, trustworthiness is nonetheless something toward which we should always strive, a disposition we should strive to develop and maintain. Trusting, however, is not as simple. While we may teach our children to be trustworthy persons, we also teach them to learn to discriminate between persons, not to trust everyone. This may in fact be one of the more important lessons which we impart to children: whom to trust and under what circumstances. Children may come with basic trust and, Erikson argues this may be necessary to a healthy personality. We want our children to grow up safe in the belief that the world, and in particular the communities in which they live, are generally trustworthy. At the same time, however, we try to teach our children to discriminate between trustworthy and untrustworthy persons. Trust is generally a good, and even necessary thing, but nonetheless we do not think that we should trust everyone. So the question arises: whom and when should we trust?

In Chapter One, I defined trust as follows: *I trust someone volitionally when I am willing to act on my evaluative trust of her.* First, it should be obvious from this definition that we should only trust persons whom we consider trustworthy. It seems, in fact, that to talk of

trusting a person whom we consider untrustworthy makes no sense at all. So half of the answer to the question of when we should trust lies in the answer to the question: who is trustworthy? Whom should we trust evaluatively? The trustworthy person, as discussed above, is the person who bears genuine goodwill for us, is competent to care for us and whom we can expect to act in ways which express his respect for us, even though he could act otherwise. Given this, it follows that we should not trust persons who do not fit this description, i.e., we should not trust persons who are untrustworthy. "We ought to trust those who are worthy of trust. But we ought not to trust those who are not" (Held 1984, 65). And it bears repeating: it is here being definitionally excluded that we could even volitionally trust persons whom we do not evaluatively trust. So then, what makes a person untrustworthy?

First, a person with no genuine goodwill for us cannot be trustworthy. A person motivated solely by self-interest is untrustworthy. While exclusive self-interest is compatible with treating persons well in some superficial sense (that is, with having superficial or instrumental goodwill for others), it is incompatible with **genuine** goodwill, that is, it is incompatible with treating others as ends in themselves, as persons valuable in their own right, not purely as means to ends outside of themselves. So only persons capable of genuine goodwill for others, and

therefore not persons motivated by self-interest **alone**, can be trusted to care for others and what matters to them. Of course, this is not to imply that a person needs to be motivated solely by altruism in order to be trustworthy, but simply that a person must have **some** altruistic motivations in order to be able to have genuine goodwill for others. So, we should remain skeptical of persons who treat us well because they gain some interest from it. The colleague, for example, who treats us well, but we discover only does so in expectation of some benefit to herself, has yet to prove herself trustworthy. (Although neither has she proved herself untrustworthy. It is just that her good treatment of us is not by itself evidence of her genuine goodwill toward us and thereby her trustworthiness.) But the colleague who leads us to believe that he is trustworthy in order to gain our confidence and garner information to use against us at a later date, thereby proves himself untrustworthy, having used us and treated us instrumentally, as a means to his own end.

Second, a person incompetent to care for others must be viewed as untrustworthy. Incompetence can come from a variety of sources. For example, a person can be incompetent by being dishonest. A person who has little skill in discerning the truth of a situation, either about themselves or about others, will likewise have little skill in disclosing the truth to others. A person who either does

not do well in distinguishing what is true from what is false or in recognizing the moral importance of honesty is untrustworthy. The person who lies without compunction is essentially untrustworthy, for to fail to recognize the moral problem of dishonesty further condemns such a person to untrustworthiness.

Self-knowledge, knowing the truth about oneself, is also a characteristic of the morally competent person, especially with regard to unspecified volitional trust. (§3) For someone who does not know himself, in terms of possessing at least some insight into his own motives and desires, etc., can hardly be expected to act competently toward himself, much less other persons.

In the most basic sense, competency can be understood as *moral* competency. Competency is the ability to care well for others, which is itself a moral activity. With regard to unspecified volitional trust, what it is to be competent is to be generally moral, generally capable of entering into morally appropriate relationships with others by being able to perceive the needs of others as well as respond to those needs. With regard to specified volitional trust, the competency is, for the most part, more particular. We trust our car to our mechanic, for example, when we judge that she is competent to take care of cars. And we trust our health to health care professionals when we judge that they have the appropriate training and skill to care for our health.

In unspecified volitional trust, however, the competency is likely to be less particular, more general. So competency can be either general moral competency, as it is in unspecified volitional trust, or some particular skill or knowledge, as it is in specified volitional trust.

And finally, we can only evaluate someone as trustworthy if we can predict that they will act in ways which express their respect for us and concern for our well-being, despite the fact that they could act otherwise. As was discussed in Chapter One, this, the predictive component of evaluative trust, follows from the other components of trustworthiness. (§3) For if we judge that someone has genuine goodwill for us and that she is either generally competent, or competent regarding some specific matter, it follows that we will expect her to behave in ways consistent with that evaluation.

So, we can be said to trust someone volitionally only when we trust them evaluatively. And trusting someone evaluatively is a matter of judging that they have genuine goodwill for us, that they are competent to care for us and what matters to us, and that they will not harm us or take advantage of us despite the fact that they could. This said, however, there is still the question of when we should **act** on our evaluation of their trustworthiness. When should we **act** on our evaluative trust of other persons?

For descriptive purposes, in Chapter One, I defined what it is to trust as being two-fold:

I trust someone when first, I believe that she: has genuine goodwill toward me, is competent to act in the ways which I trust her to act, could act in such a way as to fail to express her respect for me and concern for my well-being, will not act in such a way, and second, I make myself vulnerable to her because I have this set of beliefs about her. In other words, I trust someone to the degree that I make myself vulnerable to her because I believe that she is trustworthy. Or in yet other words, I trust someone volitionally when I am willing to act on my evaluative trust of her.

This is what it is to trust. This describes the circumstances when we make ourselves vulnerable, or choose to remain vulnerable, to another person, whether we are conscious of it or not. It describes those attitudes and actions of persons in relations of trust. Persons in a relation of trust, where genuine goodwill and vulnerability is mutual and reciprocated, must trust each other in the way that this definition describes.

For prescriptive purposes, however, I will describe trust differently. For the purposes of answering the normative question of **when** we should trust, when we should be willing to enter into, or remain in, a relationship of vulnerability with another person based on our evaluation of her as trustworthy, I want to describe trust as a mean. Aristotle tells us that a mean is

between two vices, that which depends on excess and that which depends on defect; and again it is a mean because the vices respectively fall short of or exceed what is right in both passions and actions, while excellence both finds and chooses what is intermediate.

Hence in respect of its substance and the account which states its essence is a mean (Aristotle 1107a2).

Given this description of a mean, I claim that: *Trust is a mean between the defect of cynicism and the excess of naiveté.*¹ Philippa Foot tells us that the virtues "are correctives, each one standing at a point at which there is some temptation to be resisted or deficiency of motivation to be made good" (Foot 1978, 8). The virtues are the mean in both "passions" and "actions." Trust is neither simply a passion, or belief, or simply an action. Volitional trust is based in both a set of beliefs and the action of making oneself vulnerable to those beliefs. Trust as a mean is concerned to find the middle ground between its defect, cynicism, and its excess, naiveté. Aristotle tells us that the virtues, as means, are part of a well-settled character; psychology tells us that trust is a "relatively stable personality characteristic" (Rotter 1980, 1). The trust which is a mean will be the "appropriate trust" to which Baier refers when she writes:

We should not assume that promiscuous trustworthiness is any more a virtue than is indiscriminating distrust. It is appropriate trustworthiness, appropriate trustingness, appropriate encouragement to trust which will be virtues (Baier 1994a, 16).

On the one hand, to have too little trust in other persons, for example, to believe that no one ever does anything except with the intention of gaining some benefit for themselves, or that no one ever has genuine goodwill for

another, is cynicism. "Central to cynicism is the lack of trust that is shown toward the objects of cynicism" (Kleinig 1995, 156). The cynical person is the one who trusts no one, or virtually no one, and is always looking for the ulterior motives in other persons' actions. They are "distrusting or disparaging of the motives of others" (Random House Dictionary).

On the other hand, to have too much belief in the goodwill and competency of others, for example, to take everyone at their word no matter how they may have treated you or others in the past, is naiveté. The naive person believes that everyone is purely good hearted, and that no one would ever intentionally bring harm to another. The naive person is characterized by "having or showing a lack of experience, judgment, or information, credulous" (Random House Dictionary). In turn, the credulous person has a "willingness to believe or trust too readily" (Random House Dictionary). Both cynicism and naiveté are extremes, the one an extreme of defect, and the other an extreme of excess. And both are what Thomas calls "virtue-derailing sentiments" (Thomas 1995, 11). Both cynicism and naiveté are morally problematic dispositions insofar as each undermines cooperation between persons, albeit in different ways. Persons who are either too trusting, i.e., naive, or too distrusting, i.e., cynical, cannot enter into a relation of trust. Yet it is just such a relation, the relation of

trust, which is necessary to persons cooperating with one another and contributing to one another's well-being.

Cynicism is morally corrupting or "virtue-derailing" in that at the least, it fails to affirm the moral worth of other persons, and it may in fact denigrate that worth. "Mistrust can bring out the worst in the mistrusted" (Baier 1994e, 197) The primary value of trust, according to Thomas, is that trust is an important way in which we affirm the moral worth of other persons by expressing the belief that while they could act in ways harmful to us or which take advantage of us, nonetheless, they will not. Thomas writes:

A world where trust abounds affirms our moral worth because the very best measure we could have of the value that individuals attach to objects--be they inanimate or animate, including those capable of self-consciousness--is the way in which individuals freely and regularly choose either to treat or to behave toward the objects in view of a vast array of options in terms of what they can do (Thomas 1990, 251).

Because people can sometimes harm us without harming themselves but choose not to do so, we are morally affirmed by them. And so by trusting persons, by having and expressing the belief that, despite the fact that they could act in ways which fail to express their respect for us or concern for our well-being, they will freely choose not to do so, we express to them that we hold them in high esteem. Volitionally trusting other persons is a means by which we express to them their basic moral worthiness. By trusting

other persons, we display our belief that they are morally competent to care for us and what matters to us. So not to trust, i.e., to be cynical, is to withhold such moral affirmation. Hence, in this way, cynicism is a morally corrupting force.

There is yet another, although perhaps similar, way in which cynicism is morally corrupting. Being trusted and being trustworthy are mutually reinforcing states of affairs. The more we are trustworthy, the more likely we are to be trusted. And likewise, although perhaps less recognized, the more we are trusted, the more likely we are to actually be trustworthy. For example, it can be argued that this is reflected in how we can engender trustworthy behavior in children. We trust children in order to encourage in them trustworthy behavior. Rather than wait for them to prove themselves to us as trustworthy before we act on our trust of them, we guide them into trustworthiness by letting them know that we trust them. In this way, we help them grow into socially and morally responsible attitudes, dispositions and sentiments toward other persons.

To be cynical to children is to convey to them the message that in order to be trusted, they must first display their trustworthiness, i.e., it is to fail to give them the 'benefit of the doubt.' The trusting disposition, as opposed to the cynical one, gives other persons the 'benefit of the doubt.' In this way therefore, it encourages

trustworthy behavior on their parts. The poet Marianne Moore tells us, "As contagion of sickness/makes sickness, contagion/of trust can make trust" (Moore 1944, 136)

Trusting people helps to keep them within the "community of trust, mutual concern and non-maleficence in which moral considerations for the preferences of others have their proper place" (Gillet 1989, 125).

H.J.N. Horsburgh, in an early article on trust, describes the way in which trust and distrust can affect a person's moral capacity. He writes

one of the main ways in which one individual can affect the moral growth of another is by trusting or distrusting him, for each instance of distrust erects barriers to intercourse and co-operation, thereby limiting the scope--and with it the moral opportunities--of the distrusted person's life. Thus the child who is never trusted to bring home the right change when he or she has been sent shopping does not have the same opportunities of developing habits of honesty in the handling of money as the child who is. But it is also true of intercourse between adults. The stunting effects of distrust upon moral development can be seen on every side. Indeed it is no exaggeration to say that trust is to morality what adequate living space is to self-expression: without it there is no possibility of reaching maturity (Horsburgh 1960, 350-351).

Trust, according to Horsburgh, can be morally therapeutic in that it can present persons with opportunities to do right and be good. Distrust on the other hand, closes off those opportunities, and with them avenues to moral growth.

"Distrust is an invitation to accept a dishonest role, and inveterate distrust is such a pressing invitation that many find it difficult to refuse" (Horsburgh 1960, 351). He

concludes that since we ought to offer moral support to one another, and since trust is in this way an important means by which we can offer such support to one another, we ought to trust one another, i.e., we ought not to be cynical.

Dostoyevsky's *Crime and Punishment* offers an insightful, if extreme, illustration of trust's ability to foster trustworthiness. At the end of the novel, after confessing to the crime, Raskolnikov has been sent to Siberia to serve a prison term for the murder of the old woman and her sister. Sonia, his lover, follows him there, living in the prison town in order to be able to visit and be near him. While he is first in prison, Raskolnikov is filled with despair and dejection, unable to morally recover from the heinous crimes he has committed. "He looked on [the other prisoners] and they on him with distrust and hostility" (Dostoyevsky 1866, 553). But Sonia's continuing and ever-present belief in him, her trust in his basic moral worth, eventually rescues him from his despair and dejection.

Her eyes shone with intense happiness; she understood, and she had no doubts at all about it, that he loved her, loved her infinitely, and that the moment she had waited for so long had come at last. They wanted to speak, but could not; tears stood in their eyes. They were both pale and thin; but in those sick and pale faces the dawn of a new future, of a full resurrection to a new life, was already shining. It was love that brought them back to life: the heart of one held inexhaustible sources of life for the heart of the other (Dostoyevsky 1866, 558).

Her love for him, and her deep trust in him which that love arouses, encourages his moral recovery, and hence his trustworthiness. Through her love she offers him the opportunity to recover from his moral failings. Horsburgh describes just this moral transformation as one of the ways in which therapeutic trust can operate. The unfailing devotion of family and friends, which may be better described as faith, can help someone to overcome past moral transgressions which would seem to make them untrustworthy. Horsburgh writes:

there are many cases in which it has brought about remarkable transformation of character, partly because it is a moral challenge to the person who is trusted and partly because there is a tendency to conform with the wishes of such a circle of supporting relatives and friends (Horsburgh 1960, 348).

Trust fosters trustworthiness, and so can morally transform our characters for the better. Cynicism on the other hand fosters untrustworthiness, and so can, and usually does, morally transform our characters for the worse.

Kleinig discusses the ways in which cynicism is prevalent among police. He concludes that this is especially dangerous given the way in which cynicism is antithetical to trust. He writes:

There is a deep and dangerous irony implicit in police cynicism, for one of the major functions of police as social peacekeepers is to safeguard and foster social conditions that will be sustained and characterized by trust. ... And we are all losers thereby, for it is only through the establishment of communities of trust that we are able to flourish as individuals (Kleinig 1995, 157).

Cynicism erodes trust, which is perhaps the most important bond between persons, making possible our collective and individual flourishing. So when the police, those whose role it is to *protect* our communities from the criminal members of it, themselves become cynical, we are all less safe.

Resisting the forces of cynicism can be tiring work. Especially in urban areas, perhaps even more so in New York City, cynicism surrounds us on all sides. On coming home to my apartment building one day, I put my bag of groceries down near the elevator door in the lobby while I retrieved mail from my mailbox, which was located in an alcove several yards down the hallway and out of view from the elevator door. On my return less than half a minute later, a woman was outside her apartment door waiting to reprimand me for what she clearly perceived as foolhardy behavior on my part. She proceeded first to ask me how long I had lived in the building and then lecture me on how our building was full of thieves, waiting to steal bags like mine left unattended. (I do not, by the way, live in a particularly unsafe neighborhood or unsafe building.) I was struck by how cynical she was, how sure that most people with whom we shared our living space were ready and waiting to steal a bag of groceries worth perhaps all of ten dollars. Receiving this cynical lecture on the rapaciousness of my neighbors, I silently steeled myself against an attack of

cynicism: was she about to steal my groceries and the lecture just a cover?

Thomas, in a discussion of the difference in "public trust" (Thomas 1992, 471) accorded to people based on their race, gives an account of a typical encounter facing a black man in the United States:

I was recently walking down a supermarket aisle with a hand-basket full of groceries, one in each hand. A white woman saw me and rushed for her pocketbook, which she had left in her cart. I would have had to put my own groceries down in order to take her pocketbook. No doubt she thought to herself: "He won't fool me with that old basket-in-each-hand trick" (Thomas 1992, 471).

The woman here displays distrust, which at least verges on cynicism, based on the racist belief that all, or at least most black men are thieves or would-be-thieves. No doubt, much cynicism in today's world stems from beliefs based on stereotype-based beliefs about people. Racism, sexism, anti-semitism, and the like, give rise to cynicism and mistrust aimed at particular groups of people. Such beliefs are supported by "cognitive inertia" which can be understood as "not giving due weight to the negative evidence provided by experience" (Good 1988, 41-42). Cognitive inertia, in other words, operates in persons to maintain beliefs, however wrong they may be, simply because it is easier to maintain a belief for which there exists counter-evidence than to revise such a belief. And "cognitive inertia is quite blind. It can support bad habits as well as good ones" (Good 1988, 43). Our willingness to continue to

distrust persons who have given evidence of their trustworthiness is a product of cognitive inertia, and thus one of the "important factors in the maintenance of racial prejudice" (Good 1988, 40). Resisting cynicism then includes being willing to revise our beliefs and resist cognitive inertia.

Gambetta, in his discussion of the Mafia's exploitation of distrust between people, notes that "Sicilian culture encourages people to see conspiracies in random events" (Gambetta 1993, 224). This is consistent with what psychology tells us about cynicism. "As distrust increases, the social fabric disintegrates" (Rotter 1980, 1). Cynicism seems to wear away the very foundations of society.

If it is true that cynicism is morally corrupting, in the sense that it fails to affirm the moral worth of persons--and that likewise it encourages untrustworthiness in people by not extending to them the 'benefit of the doubt'--then there would seem to be a moral presumption in favor of a trusting disposition over a cynical one. Trust engenders trustworthiness, and trustworthiness in turn engenders trust, so not only should we always try to be trustworthy, but there is also a moral presumption that speaks in favor of being trusting, i.e., not cynical. Such trust fosters trustworthiness on the part of the one trusted.

I have characterized a general lack of trust for other persons as cynicism. A complete lack of trust for other persons is exhibited in a lack of mental health. Paranoia is characterized as the belief that

people are persecuting the patient with threats and slander, putting poison in his food, tormenting him with electricity, &c. Not infrequently the patient asks the police for protection; or, in alcoholics, suspicions may appear as to the wife's faithfulness; noises occur in the house; the light and the position of blinds are signs of some secret understanding, &c" (Dictionary of Philosophy and Psychology).

Erikson too recognized the psychopathological nature of the absence of trust. He writes:

In psychopathology the absence of basic trust can best be studied in infantile schizophrenia, while lifelong underlying weakness of such trust is apparent in adult personalities in whom withdrawal into schizoid and depressive states is habitual. The re-establishment of a state of trust has been found to be the basic requirement for therapy in these cases (Erikson 1950, 248-249).

So, not only can cynicism be "virtue derailing", but if it is taken too far, it can actually be derailing of one's mental health. Psychology tells us that persons who do not tend to trust others experience problems including "competitiveness, envy, resentfulness, vindictiveness, and lack of feelings toward others" (Gurtman 1992, 999).

Just as cynicism is morally corrupting or "virtue derailing", so is naiveté. Naiveté, defined as an excess of trust, allows and encourages morally suspicious or questionable states of affairs. As Baier notes:

"Exploitation and conspiracy, as much as justice and fellowship thrive better, in an atmosphere of trust" (Baier 1994b, 95). The specific atmosphere of trust in which such morally corrupt states of affairs as exploitation and conspiracy thrive is an atmosphere of naive trust. "Those who put their trust in those who perpetuate exploitation or domination are not to be admired for their willingness to trust" (Baier 1994e, 184) In this world, naive people are exploited, taken advantage of, abused. Naiveté is a morally corrupting disposition then, because it creates an atmosphere in which people are allowed to take advantage of one another, and freely act toward other persons in untrustworthy ways. Naiveté fosters untrustworthiness, by allowing untrustworthy people to exploit and take advantage of others and benefit from it.

One of the early lessons we teach children, as they grow older and thus are less and less under the constant supervision of their caretakers, is not to be naive. In order to protect children, we teach them to protect themselves by learning to discriminate between persons and learn that not all persons necessarily have their best interest at heart or in mind. One of the more admirable qualities of children--their tendency to believe that everyone is good and friendly--can also be one of their more frightening qualities. Thus we teach them not to talk to strangers, take candy from people they do not know, and the

like. And we do this in order to protect them from exploitation and from persons who would readily take advantage of them.

Naiveté is not only a matter of prudential concern, however; it also contributes to the creation of a morally corrupt atmosphere in which everyone is more likely to be taken advantage of, abused. The man who out of naiveté allows himself to be cheated by the confidence artist, also allows the confidence artist to continue unabated to exploit her next victim. "We must," therefore, "take care that our individual willingness to forgive does not put others in danger" (Baier 1994c, 150). The man, by contrast, who is not naive and therefore not fooled by the confidence artist, helps create an atmosphere in which confidence artists are less free to exploit and abuse in general. Naiveté is to be avoided because it contributes to and cooperates in the existence of exploitation and abuse. Just as there is a moral presumption in favor of a basically trusting disposition over a cynical one, so there is a moral presumption in favor of a basically *discretionary* trusting disposition over a naive disposition.

We should avoid cynicism in order to avoid creating an atmosphere in which we fail to affirm one another morally and in which we fail to encourage trust by failing to encourage one another to be trustworthy. Likewise, we should avoid naiveté in order to avoid creating an

atmosphere in which exploitation and abuse are allowed to thrive. Rather, we should seek the mean between trusting too little and trusting too much; we should practice the mean. "To be trusting, naive, or innocent may *help* others to take advantage of one, it may *contribute* to wrongdoing, and may thus, at various times and places, be a bad rather than a good characteristic" (Held 1984, 65).

In order to illustrate how appropriate trust can be understood as a mean between cynicism and naiveté, consider the following example. Imagine that Sue and Bob, sexually active adults (or at least not virgins), neither of whom have had sex for the previous six months² and neither of whom have been tested for HIV, have been dating for several months. They are mutually attracted to one another and ready to begin an agreed upon monogamous sexual relationship. What would be the mean of trust in this situation? On the one hand, it would be naive for either of them to be willing to have "unsafe" sex with the other without the assurance that the other has tested negative for HIV.³ This would be naive in that it would ignore the rise of HIV among heterosexuals. It would be naive because people are not always honest about their sexual histories, much less aware of whether or not they may be carriers of sexually transmitted diseases. Imagine further that Bob assures Sue that he has never had sex with any high risk women, nor any men. Bob assures Sue that he has no reason

to believe that he could possibly be carrying the HIV, so there is no reason for him to get tested or to practice safe sex. For Sue to accept Bob's assurances would be naive given that people do not always know everything that there is to know about the sexual histories of their partners, that people sometimes have self-interested reasons for not wanting to be tested (for example, perhaps Bob suspects he may be infected with HIV and is afraid to find out whether his suspicions are correct or not). For Sue to have unsafe sex with Bob would also be naive of her in that Bob would not only be asking her to trust him, but, in effect, to trust as well the women with whom Bob has had sex in the past, women whom Sue presumably does not know. It is further to trust the other partners of the women Bob has had sex with, and so on. In general, in an age when sexually transmitted diseases are on the rise and one of them, AIDS, is fatal, to have sex with people who we are not assured are healthy and not carriers of disease, is naive. So for Sue to have sex with Bob under these circumstances would be to display an excess of trust, or naiveté.

On the other hand, imagine that Bob readily agrees to be tested for HIV. But Sue, not satisfied with this, insists on accompanying him to the doctor's office and verifying his test results for herself. Also, although Bob has professed his intention to be monogamous and only have sex with her, Sue insists that they practice safe sex

nonetheless, in order to protect herself from the possibility that Bob is having sex with other women. This attitude on Sue's part would exhibit cynicism. It would be based on the belief that people, Bob in particular (whom she has known for a while, and whom, for the sake of illustration we can take to have given Sue no reason to believe that he is dishonest, unfaithful, or out to harm her), lie and are not to be trusted, no matter the lengths to which they are willing to go to display their trustworthiness (in this case, being tested in the first place). We can well imagine Sue and Bob's relationship breaking down under the strain of this cynicism.

So, not to trust him at all, and not be willing to have unsafe sex with him despite his being tested and expressing his intention to be monogamous, would be cynical. On the other hand, to have unsafe sex with him without his being tested, would be naive.⁴ For Sue then, the mean would be to trust Bob in a way which avoids the excess of trusting him so much that she completely fails to protect her own well-being, and also avoids the defect of trusting him so little that there is hardly any room left for a meaningful moral relationship between them.

As Aristotle tells us, finding the mean is not easy.

He writes:

That moral excellence is a mean, then, and in what sense it is so, and that it is a mean between two vices, the one involving excess, the other deficiency,

and that it is such because its character is to aim at what is intermediate in passions and in actions, has been sufficiently stated. Hence also it is no easy task to be good. For in everything it is no easy task to find the middle, e.g., to find the middle of a circle is not for every one but for him who knows; so, too, any one can get angry--that is easy--or give or spend money; but to do this to the right person, to the right extent, at the right time, with the right motive, and in the right way, that is not for every one, nor is it easy; that is why goodness is both rare and laudable and noble (Aristotle 1109a20).

Finding the mean depends on an entire set of relevant circumstances. To whom, when, where, under what circumstances all matter when determining what constitutes trust as a mean and avoids both cynicism and naiveté. "Trust ... is felt to be something which should be suited to the tactical requirements of each particular moral situation" (Horsburgh 1960, 353). For example, one's location can make a great deal of difference for what counts as trust, cynicism or naiveté. New York City is different, for instance, from rural New Hampshire. In New York City, to use a "club" to prevent one's car from being stolen is not cynical, given how often cars are stolen. Whereas in rural New Hampshire, to use such a device because one believes that one is likely to be victimized by rapacious thieves would be so cynical as to be quite ludicrous. The identity of the agent also clearly determines where the mean of trust lies. For a man walking alone at night, to cross the street because another man was approaching may verge on cynicism, (assuming for the sake of argument that the

approaching man isn't brandishing a weapon, or in some other way appearing particularly threatening), whereas for a woman, it may be advisable, given the prevalence of violence against women. To be cautious when approached by a police officer may be cynical for a white person. But not to be cautious when approached by a police officer would be naive for a young black man, given the prevalence of police harassment and abuse against young black men. At the same time, however, for a woman to take all men to be likely to commit violence against her would be cynical. Likewise, for a young black man to assume that all police officers are racist is also cynical. Who one is, where one is, and all the circumstances in which one must choose whether to trust or not, are relevant as to what counts as cynical as what as naive. Nonetheless, the very temptation to categorize groups as trustworthy or not (as the woman in Thomas's example seems to have done), itself borders on cynicism, and is for that reason to be avoided by exercising appropriate trust.

These illustrations could be reproduced endlessly. They serve to show that finding the mean of trust (or any other mean for that matter), is no easy matter.

To have one's emotions and actions in a mean, says Aristotle, is to feel and manifest each emotion at such times, on such matters, toward such people, for such reasons, and in such ways as are proper (Urmson 1980, 161)

For that reason, there can be no more specific general principle than one such as the following: *Trust in a way which avoids cynicism on the one hand and naiveté on the other.* While a principle to guide us as to when we should trust and when we should not can be no more specific than this, we can know a great deal of what will count as cynical, what as naiveté, and therefore what will count as the mean of trust; it will always depend on specific contextually sensitive factors, which cannot be known in advance. Understanding trust in general then, will have to be done from the ground up; we will first have to be able to make sense out of trust in individual lives, relationships and circumstances before we will be able to make sense of it in any generalized way.

Social trust, and not only personal trust, can be understood as a mean between cynicism and naiveté. Practicing appropriate trust, which navigates between cynicism and naiveté, can foster and maintain the social conditions which themselves promote cooperation between persons. By having genuine goodwill for others and being competent, i.e., being trustworthy, we can help to avoid the social conflicts which themselves are predicated on distrust.

The Prisoner's Dilemma, we are told, starts from the assumption of mutual unconcern.⁵ Robert Axelrod, in the opening sentence of his *The Evolution of Cooperation*, asks

the basic, Hobbesian question he takes as informing his project: "Under what conditions will cooperation emerge in a world of egoists without central authority" (Axelrod 1984, 3)? The conclusion of Axelrod's study, which consisted of a series of computer tournaments designed to test strategies for use in iterated Prisoner's Dilemmas, was that TIT FOR TAT is the best strategy for iterated Prisoner's Dilemmas. TIT FOR TAT's basic qualities are its "niceness" and its "forgiveness" (Axelrod 1984, 42). TIT FOR TAT is a very simple strategy which "starts with a cooperative choice, and thereafter does what the other player did on the previous move" (Axelrod 1984, 31).

TIT FOR TAT can be understood, without a great stretch of the imagination, as practicing appropriate trust (if we can, for the moment, describe such strategies as 'trusting'). It begins by cooperating, and only defects in retaliation to the other player's defection. In other words, TIT FOR TAT "gives its opponents the benefit of the doubt" (Poundstone 1992, 240). It is not cynical. (The strategy which defects on every move no matter how the other player behaves is cynical, and does very poorly.) Neither is it naive. (The strategy which cooperates on every move no matter how the other player behaves is naive, and, being highly exploitable, also does very poorly.) TIT FOR TAT navigates between the two. It is a cooperative strategy.

"Doing well with TIT FOR TAT requires cooperating with it, and this in turn helps TIT FOR TAT" (Axelrod 1984, 53).

Although TIT FOR TAT is the most successful strategy of all strategies which were submitted for Axelrod's tournaments, it is not rational in the sense of straightforward individual rationality's injunction that each and every action be such as to maximize individual utility. In a Prisoner's Dilemma, individual rationality requires defection (§7), yet TIT FOR TAT's first move is to cooperate. It is not in this sense then rational, yet it is not unreasonable either. And perhaps most importantly, by giving the benefit of the doubt, being nice, and practicing forgiveness, TIT FOR TAT manages to yield optimal results for both itself and its opponent. TIT FOR TAT is not rational and it is not competitive, and its results are not collectively disastrous.

The Prisoner's Dilemma, the paradox that results from individual rationality and competition, need not be only predicated on the assumption of self-interest. It is only the **structure** of the interaction which produces the paradox, not necessarily the desires of the players. (§7) That is, the dilemma is equally generated by an altruistic set of preferences as by a self-interested one (Held 1977) and (Parfit 1984, 66).

Trust, however, can negotiate around such dilemmas. Poundstone writes that "the only satisfying solution to the

Prisoner's dilemma is to avoid the prisoner's dilemma" (Poundstone 1992, 278). He also notes that "for a person with a **sufficiently** strong sense of empathy, there is no such thing as a prisoner's dilemma" (Poundstone 1992, 233) (emphasis added). Trust is not altogether altruistic, it need not involve sacrificing one's interests to the truster. But it is not self-interested either; it is constituted by genuine goodwill for others. Appropriately trusting as a policy, as the TIT FOR TAT strategy seems to do, is a good strategy to adopt.

Trust as I have defined it (§3) is not rational; it does not rely on a rational calculation of the costs and benefits to either the truster or trusted. But, while perhaps not always rational, trust is not therefore unreasonable. Held suggests that the lesson the collectively disastrous results which rationality creates in both Prisoner's Dilemmas and Altruist's Dilemma is that we would do well to replace the traditional conception of rationality as espoused by Hobbes and others when dealing with situations of interaction between persons where conflicts of interests between such persons partially constitutes the situation. Rationality demands that persons always pursue their own self-interest, and we would, Held maintains, do well to replace it with the concept of the reasonable. She writes: "I think, that it is *reasonable*, if not demanded by rationality itself, to do that which is in

the interests of others when it is no less in the interests of ourselves" (Held 1977, 736). Being reasonable, and not always presuming that the interests of others can only be furthered at the expense of ourselves, is part of what is required by appropriate trust. Held writes:

Both egoism and altruism are competitive policies, policies which try to maximize an individual value, for self or other. Cooperative policies, in contrast, are concerned with the mutual, collective value which is realized by a given choice relation (Held 1977, 740).

Appropriate trust, by negotiating between cynicism and naiveté, and their disastrous strategies, fosters cooperation. It avoids egoism and altruism and the competition on which they are based. Trust is in this way perhaps, the fundamental virtue. By fostering trust, we can avoid the social dilemmas and the "cognitive inertia" which maintain themselves through cynicism and naiveté. Baier writes:

The moral enterprise is built on the faith that interests can to some extent be reconciled, that flourishing need not always be at the expense of others (Baier 1985a, 143).

Trust supports and maintains this moral enterprise.

Understanding when to trust in Aristotelian terms, as based on a disposition to choose the mean between cynicism and naiveté, is consistent with the understanding offered by Erikson discussed at the beginning of Chapter One. According to Erikson, basic trust must be learned by the infant through appropriate interaction with caretakers in

the first year of life in order for the infant to develop in psychologically healthy ways. But it is not that one must learn to trust *simpliciter*, rather the infant needs to learn to be more influenced by basic trust than by basic mistrust. In order to be psychologically healthy, we need not be altogether trusting, i.e., naive, nor altogether mistrusting, i.e., cynical. Rather, we need to find the middle ground between these two states, that middle ground being 'appropriate' trust. To develop such a disposition to trust in this way, one which avoids the extremes of excess and defect, is recommended psychologically by Erikson. It is after all the "cornerstone of a healthy personality" (Erikson 1959, 58). But it is not only that; it also provides the very foundation for moral relationships between persons. Knowing when to trust, and likewise, knowing when to not trust, is the basic quality of persons capable of entering into relations of trust in which persons can be mutually vulnerable to one another because of their genuine goodwill for each other and competence to care for one another. Such relations of trust allow persons to affirm each other's moral worth, and to contribute to each other's moral growth. We are, Aristotle reminds us again and again, social and moral beings. Precisely because of this, trust is indispensable.

Notes

1. This distinction is similar to the distinction which Horsburgh draws between "innocent trust," "guilty trust" and "therapeutic trust" (Horsburgh 1960).
2. This condition is necessary because the HIV can take up to six months to appear on antibody tests. Given this, a person can be carrying the virus and yet test negative if they are tested within the first six months of having been exposed.
3. It may be more naive for Sue, given the medical evidence that a woman is more likely to be infected by an HIV-positive man, than a man is likely to be infected by a HIV-positive woman. AIDS is a "dead-end" disease in women, meaning that women are more likely to get it than to give it (Padian, Shiboski, and Jewell 1991).
4. No doubt, some people would disagree, claiming that given that HIV causes a disease which is in all instances fatal, it would not be naive of Sue to continue to insist that she and Bob practice safe sex even if he does test negative for the virus. This line of argument however, if carried to its natural conclusion, would imply that **all** people, in **all** circumstances, including married people, should practice safe sex, since there is never any guarantee that people (married or not) who declare themselves to be monogamous or virgins are being honest.
5. William Poundstone, in his biography of John von Neumann, the founder of game theory (of which the study of Prisoner's Dilemmas is a branch), tells us that "Game theory was the brainchild of a cynic" (Poundstone 1992, 39).

REFERENCES

- Adams, Robert M. 1987. *The Virtue of Faith*. Oxford: Oxford University Press.
- . 1993. "The Problem of Total Devotion." in Badhwar (Editor) 1993.
- . 1995. "Moral Faith." in *The Journal of Philosophy* (February 1995): 75-95.
- Antony, Louise M. and Witt, Charlotte (Editors). 1993. *A Mind of One's Own: Essays on Reason and Objectivity*. Boulder CO: Westview Press, Inc.
- Aristotle. *Nicomachean Ethics* (trans. W.D. Ross; revised by J.O. Urmson). in Barnes (Editor). 1984. Vol. 2: 1729-1867.
- Arras, John, and Rhoden, Nancy (Editors). 1989. *Ethical Issues in Modern Medicine* (Third Edition). Mountainview CA: Mayfield Publishing Company.
- Attridge, Derek. "Trusting the Other: Ethics and Politics in J. M. Coetzee's *Age of Iron*" *South Atlantic Quarterly* (Winter 1994): 59-82.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Badhwar, Neera Kapur. 1991. "Why It Is Wrong to be Always Guided by the Best: Consequentialism and Friendship." *Ethics: An International Journal of Social Political and Legal Philosophy* Vol. 101, No. 3, (April 1991): 483-504.
- (Editor). 1993. *Friendship: A Philosophical Reader*. Ithaca, NY: Cornell University Press.
- Baier, Annette. 1985. *Postures of the Mind: Essays on Mind and Morals*. Minneapolis: University of Minnesota Press.
- . 1985a. "Knowing Our Place in the Animal World" in Baier. 1985: 139-156.
- . 1985b. "Secular Faith." in Baier 1985: 292-308.
- . 1988. "Pilgrim's Progress: Review of David Gauthier, *Morals by Agreement*." *Canadian Journal of Philosophy* Vol. 18, No. 2. (June 1988): 315-330.

- . 1994. *Moral Prejudices: Essays on Ethics*. Cambridge: Harvard University Press.
- . 1994a. "What Do Women Want in a Moral Theory?" in Baier 1994: 1-17.
- . 1994b. "Trust and Antitrust" in Baier 1994: 95-129.
- . 1994c. "Trust and Its Vulnerabilities" in Baier 1994: 130-151.
- . 1994d. "Sustaining Trust" in Baier 1994: 152-182.
- . 1994e. "Trusting People" in Baier 1994: 183-202.
- Barnes, Jonathan (Editor). 1984. *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton, NJ: Princeton University Press.
- Baumrin, Bernard. 1989. "Hobbes's Egalitarianism: The Laws of Natural Equality." in M. Bertman and M. Malherbe (Editors) 1989: 119-127.
- Benn, S.I. and Gaus, G.I. (Editors). 1983. *Public and Private in Social Life*. London: Croon Helm
- Bertman, M. and Malherbe, M. (Editors). 1989. *Thomas Hobbes: De La Metaphysique a La Politique*. Colloque Franco-Américain de Nantes. Librairie Philosophique VRIN Paris.
- Blum, Lawrence. 1993. "Friendship as a Moral Phenomenon." in Badhwar (Editor) 1993: 192-210.
- Bok, Sissela. 1978. *Lying: Moral Choice in Private and Public Life*. New York: Vintage Books.
- Bowie, Michaels, and Solomon (Editors). 1992. *Twenty Questions: An Introduction to Philosophy* (Second Edition). New York: Harcourt Brace Jovanovich.
- Braybrooke, David. 1976. "The Insoluble Problem of the Social Contract." *Dialogue* Vol. XV, No. 1: 3-37.
- Chodorow, Nancy. 1978. *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender*. Berkeley: University of California Press.
- Code, Lorraine. 1991. *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca NY: Cornell University Press.

- Cooper, John M. 1990. "Political Animals and Civic Friendship." in Badhwar (Editor). 1993: 301-326.
- Dawkins, Richard. 1989. *The Selfish Gene* (Second Edition). Oxford: Oxford University Press.
- Di Stefano, Christine. 1983. "Masculinity as Ideology in Political Theory: Hobbesian Man Considered." *Women's Studies International Forum* Vol. 6. No. 6: 633-644.
- Dictionary of Philosophy and Psychology* (Vol. II). Gloucester, MA: Peter Smith. Baldwin, James Mark, (Editor). 1960.
- Dostoyevsky, Fyodor. 1866. *Crime and Punishment*. (trans. David Magarshack). Middlesex, England: Penguin Books. (1951)
- Dunn, John. 1974. "The Concept of 'Trust' in the Politics of John Locke" in Rorty, Schneewind, Skinner (Editors). 1974: 270-301.
- . 1995. "Annette C. Baier: Moral Prejudices: Essays on Ethics" review article in *The Journal of Philosophy* (January 1995): 44-47.
- Erikson, Erik. 1950. *Childhood and Society*. New York: W.W. Norton & Company.
- . 1959. *Identity and the Life Cycle*. New York: W.W. Norton & Company.
- Flanagan, Owen and Rorty, Amélie (Editors). 1990. *Identity, Character and Morality*. Cambridge: MIT Press.
- Foot, Philippa. 1978. *Virtues and Vices*. Berkeley: University of California Press.
- Fried, Charles. 1970. *An Anatomy of Values: Problems of Personal and Social Choice*. Cambridge: Harvard University Press.
- Friedman, Marilyn. 1989. "Feminism and Modern Friendship: Dislocating the Community." in Badhwar (Editor). 1993: 285-302.
- Gambetta, Diego (Editor). 1988. *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell.
- . 1988a. "Mafia: The Price of Distrust." in Gambetta (Editor). 1988: 158-175.

- . 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge: Harvard University Press.
- Gauthier, David. 1975. "Reason and Maximization" in Gauthier 1990: 209-233.
- . 1977. "Social Contract as Ideology" in Gauthier 1990: 325-354.
- . 1979. "Thomas Hobbes: Moral Theorist" in Gauthier 1990: 11-23.
- . 1986. *Morals by Agreement*. Oxford: Oxford University Press.
- . 1988. "Hobbes's Social Contract." *Noûs* 22: 71-82.
- . 1990. *Moral Dealing: Contract, Ethics, and Reason*. Cornell: Cornell University Press.
- . 1991. "Why Contractarianism?" in Vallentyne 1991: 13-30.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Cambridge: Harvard University Press.
- Gillet, Grant. 1989. "Aids and Confidentiality." in Arras and Rhoden (Editors). 1989.
- Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.
- Good, David. 1988. "Individuals, Interpersonal Relations, and Trust." in Gambetta (Editor). 1988: 31-48.
- Govier, Trudy. 1992a. "Distrust as a Practical Problem." *Journal of Social Philosophy* (Spring 1992): 52-63.
- . 1992b. "Trust, Distrust and Feminist Theory." *Hypatia* (Winter 1992): 16-33.
- . 1993. "Self-trust, Autonomy and Self-Esteem." *Hypatia* (Winter 1993): 99-120.
- Gurtman, Michael B. 1992. "Trust, Distrust, and Interpersonal Problems: A Circumplex Analysis." *Journal of Personality and Social Psychology* Vol. 62, No. 6: 989-1002.

- Hampshire, Stuart (Editor). 1978. *Public and Private Morality*. Cambridge: Cambridge University Press.
- Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- . 1993. "Feminist Contractarianism." in Antony and Witt (Editors). 1993: 227-255.
- Hardin, Russell. 1990. "Trusting Persons, Trusting Institutions."
- Held Virginia. 1968. "On the meaning of trust." *Ethics: An International Journal of Social, Political and Legal Philosophy* (January 1968): 156-159.
- . 1977. "Rationality and Reasonable Cooperation." *Social Research* (Winter 1977): 708-744.
- . 1984. *Rights and Goods: Justifying Social Action*. New York: The Free Press.
- . 1993. *Feminist Morality: Transforming Culture, Society, and Politics*. Chicago: The University of Chicago Press.
- Herman, Barbara. 1993. *The Practice of Moral Judgement*. Cambridge: Harvard University Press.
- Hobbes, Thomas. 1651a. *Leviathan*. C.B Macpherson (Editor). London: Penguin Books (1985)
- . 1651b. "Philosophical Rudiments Concerning Government and Society" in *The English Works of Thomas Hobbes*. (Edited by Sir W. Molesworth, Darmstadt, 1966).
- Horsburgh, H.J.N. 1960. "The Ethics of Trust." *Philosophical Quarterly* (October 1960): 343-354.
- Hume, David. 1739. *A Treatise of Human Nature*. London: Penguin Books (1984)
- . 1777. *Essays: Moral, Political, and Literary*. (Edited by Eugene F. Miller. Indiannopolis: Liberty Fund 1987)
- Kant, Immanuel. 1775. *Lectures on Ethics*. (trans. by Louis Infield) Indiannapolis: Hackett Publishing Company (1963)

- . 1785. *Foundations of the Metaphysics of Morals*.
(trans. by Lewis White Beck). The Bobbs-Merrill
Company, Inc. (1959)
- Kavka, Gregory S. 1986. *Hobbesian Moral and Political
Theory*. Princeton: Princeton University Press.
- Kelman, Steven. 1990. "Congress and Public Spirit." in
Mansbridge (Editor). 1990: 200-206.
- Kleinig, John. 1978. "Crime and the Concept of Harm."
American Philosophical Quarterly (January 1978): 33-36.
- . 1995. *The Ethics of Policing*. (Forthcoming from
Cambridge University Press; page numbers refer to
typescript version.)
- Lewis, C.S. 1960. "Friendship - The Least Necessary Love."
in Badhwar (Editor). 1990: 39-47.
- Lipset, Seymour Martin and Schneider, William. 1983. *The
Confidence Gap: Business, Labor and the Government in
the Public Mind*. New York: The Free Press.
- Lyons, David. 1965. *Forms and Limits of Utilitarianism*.
Oxford: Oxford University Press.
- Macpherson, C.B. 1968. "Introduction." in Hobbes, Thomas.
Leviathan. 1651. London: Penguin Books. (1985): 9-63.
- . 1973. *Democratic Theory: Essays in Retrieval*.
Oxford: Clarendon Press.
- Mansbridge, Jane J. (Editor). 1990. *Beyond Self-Interest*.
Chicago: University of Chicago Press.
- McAulay, Catharine. 1767. "Loose Remarks on Certain
Propositions to be Found in Mr. Hobbes' Philosophical
Rudiments of Government and Society."
- Mohr, Richard D. 1988. *Justice: A Study of Ethics,
Society, and Law*. New York: Columbia University Press.
- Moore, Marianne. 1944. *Complete Poems*. New York: Penguin
Books, Inc. (1994).
- Nozick, Robert. 1974. *Anarchy, State and Utopia*. New York:
Basic Books.
- Okin, Susan Moller. 1989. *Justice, Gender, and the Family*.
New York: Basic Books.

- Padian, Nancy S.; Shiboski, Stephen C.; Jewell, Nicholas P. 1991. "Female-to-Male Transmission of Human Immunodeficiency Virus." *Journal of the American Medical Association* (25 September 1991): 1664-1667.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Parker, Suzanne L. and Parker, Glenn R. 1993. "Why Do We Trust Our Congressman?" *The Journal of Politics* (May 1993): 442-453.
- Pateman, Carole. 1988. *The Sexual Contract*. Stanford: Stanford University Press.
- Poundstone, William. 1992. *Prisoner's Dilemma: John Von Neumann, Game Theory, and the Puzzle of the Bomb*. New York: Doubleday.
- Random House Dictionary of the English Language: The Unabridged Edition*, The. 1966. New York: Random House. (Stein, Jess, Editor in Chief).
- Raymond, Janice G. 1986. *A Passion for Friends: Toward a Philosophy of Female Affection*. Boston: Beacon Press.
- Rescher, Nicholas. 1990. *A Useful Inheritance: Evolutionary Aspects of the Theory of Knowledge*. Savage MD: Rowman & Littlefield Publishers, Inc.
- Rorty, Amélie (Editor). 1980. *Essays on Aristotle's Ethics*. Berkeley: University of California Press.
- Rorty, Richard; Schneewind, J.B.; Quentin, Skinner (Editors). 1974. *Philosophy in History: Essays in the Histiography of Philosophy*. Cambridge: Cambridge University Press.
- Rotter, Julian B. 1980. "Interpersonal Trust, Trustworthiness, and Gullibility." *American Psychologist* (January 1980): 1-7.
- Sandel, Michael. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Sayre-McCord, Geoffrey. 1991. "Deception and reasons to be moral." in Vallentyne (Editor) 1991: 181-195.
- Schwarzenbach, Sibyl. 1994. "On Civic Friendship." unpublished manuscript

- Smart, J.J.C. 1973. "An Outline of a System of Utilitarian Ethics." in Smart and Williams. 1973: 3-74.
- Smart, J.J.C. and Williams, Bernard. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Stocker, Michael. 1976. "The Schizophrenia of Modern Ethical Theories." *Journal of Philosophy* 73: 453-466.
- . 1993. "Values and Purposes: The Limits of Teleology and the Ends of Friendship." in Badhwar (Editor). 1993: 245-263.
- Thomas, Laurence. 1989. *Living Morally*. Philadelphia: Temple University Press.
- . 1990. "Trust, Affirmation and Moral Character: A Critique of Kantian Morality." in Flanagan and Rorty 1990 (Editors): 235-257.
- . 1992. "Next Life, I'll Be White" in Bowie, Michaels and Solomon (Editors). 1992: 471-472.
- . 1995. "Becoming an Evil Society: The Self and Strangers." (Forthcoming in *Political Theory*, page number refer to typescript version).
- Tong, Rosemarie. 1993. *Feminine and Feminist Ethics*. Belmont, CA: Wadsworth Publishing Company.
- Urmson, J.O. 1980. "Aristotle's Doctrine of the Mean." in Rorty, Amélie (Editor) 1980.
- Vallentyne, Peter. (Editor). 1991. *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*. New York: Cambridge University Press.
- . 1991a. "Gauthier's three projects." in Vallentyne (Editor). 1991: 1-11.
- . 1991b. "Contractarianism and the assumption of mutual concern." in Vallentyne (Editor). 1991: 71-75.
- Walker, Margaret Urban. 1989. "Moral Understandings: Alternative 'Epistemology' for a Feminist Ethics." *Hypatia* (Summer 1989): 15-28.
- Wittgenstein, Ludwig. 1969. *On Certainty*. (trans. and edited by Denis Paul and G.E.M. Anscombe. New York: Harper & Row Publishers).

Wright, J. Eugene, Jr. 1982. *Erikson: Identity and Religion*.
New York: The Seabury Press.

Yi, Byeone-Uk. 1992. "Rationality and the Prisoner's Dilemma
in David Gauthier's *Morals by Agreement*." *Journal of
Philosophy* (September 1992): 484-495.