THE RESPONSIBILITIES OF TRUTHFULNESS:
AN INQUIRY INTO THE ETHICS OF SPEAKING

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHILOSOPHY

MAY 2011

By
Jeremy E. Henkel

Dissertation Committee:
Arindam Chakrabarti, Chairperson
Ronald Bontekoe
Vrinda Dalmiya
Kenneth Kipnis
R. Kelly Aune
We certify that we have read this dissertation and that, in our opinion, it is satisfactory in scope and quality as a dissertation for the degree of Doctor of Philosophy in Philosophy.

DISSERTATION COMMITTEE

______________________________
Chairperson

______________________________

______________________________

______________________________

______________________________
For Mom
Acknowledgements

Not only in writing this dissertation, but in getting to the point where I was able to begin such a daunting task, I have benefitted from more people’s support and encouragement than I can hope to recall here. But I will do my best.

To Arindam Chakrabarti, chairperson of this dissertation: thank you for the mentorship you have provided me throughout my graduate school career. I truly would not be where I am today without your help and guidance. To Ron Bontekoe: thank you for your unyielding support during the last nine years, and for not giving up on me when I didn’t always perform up to the high standards you knew I was capable of. To Vrinda Dalmiya: thank you for agreeing to serve on my dissertation committee without making me beg too severely, and for your incisive comments when I was not as clear in my writing or my thinking as I should have been. To Ken Kipnis: thank you for your patience and guidance when it was unclear how I would avoid the pitfalls you saw me setting for myself. To Kelly Aune: thank you for putting yourself into the unenviable position of joining a dissertation committee when the dissertation was nearly completed.

Thank you also to Holly Phillips, who drilled into me as an undergraduate the importance of writing well by not wasting words, to Jim Tiles, who has contributed as much as anyone to my ability to understand Kant, and to Rama Nath Sharma, for helping me to understand Sanskrit by “seeing through” the language.

Special thanks go to my writing group cohorts: Benjamin Lukey, Amy Donahue, Andrew Lambert, Nicholas Brasovan, and Matthew LoPresti. My writing and thinking have benefitted at every turn from your consistently thoughtful feedback. Thank you also to Thorian Harris for our sporadic ‘weekly’ discussions that helped to keep me on track.

Thank you to Renee Kojima-Itagaki and Jackie Young for frequently saving me from myself and for keeping things running smoothly.

Finally, thank you to my family—all of you—for your unwavering support, even when you didn’t understand what I wanted to do or why I wanted to do it. Whatever I have accomplished, your faith and your confidence in me have made it possible.
Abstract

Analyses of the morality of lying almost universally begin by attempting to define lying and subsequently assessing its ethical status on the basis of the definition provided. This approach is problematic because lying is an inherently ethical concept. As a result, people who attempt to define lying in a value-neutral manner inevitably smuggle in their own ethical intuitions about the concept. As a result, the success of the definitions and analyses provided is wholly dependent on one’s acceptance of the informing intuitions. I take a different approach, following the strategy suggested by natural language philosophy. I first examine the role the concept lying plays in everyday discourse, and then use this analysis to show that lying is best understood as always morally objectionable. This realization provides a starting point for discussing what makes lying blameworthy, and ultimately for deriving a definition of what it is to lie on the basis of that analysis. I argue that lying is wrong because it constitutes a betrayal of trust. The wrongness of lying, and indeed lying itself, is thus more contextually determined than has been generally recognized.

Building on my analysis of lying, I develop an understanding of truthfulness as the central virtue of speaking. Making use of classical Indian philosophy—including insights from the Maḥābhārata, the Upaniṣads, and the Buddhist Nikāyas—I argue that we need to de-couple the notion of truthfulness from that of truth-telling. It is a mistake, I argue, to identify the value-neutral act of truth-telling with the virtue of truthfulness. I then defend my account of truthfulness against the twin objections that it is self-undermining and that it fosters an unacceptable form of paternalism. I conclude by articulating three Responsibilities of Truthfulness that, when fulfilled, help to ensure that one remains trustworthy in what one says.
# Table of Contents

Acknowledgements .......................................................................................................... v  
Abstract ............................................................................................................................ vi  
Chapter 1: What’s Wrong with Lying? .............................................................................. 5  
  Introduction .................................................................................................................. 5  
  1.1: The Folly of Trying to Define “Lie” ...................................................................... 7  
  1.2: The Evaluative Nature of “Lie” ........................................................................... 21  
  1.3: Why Lying is Always Wrong .............................................................................. 35  
  Conclusion: Evaluation Without Definition .............................................................. 44  
Chapter 2: The Sophistication of Kant’s View of Lying .................................................... 48  
  Introduction .................................................................................................................. 48  
  2.1: Kant’s Condemnation of Lying .......................................................................... 49  
    2.1.1: Kant on Truth and Truthfulness ................................................................. 55  
    2.1.2: Kant on Declarations .................................................................................... 61  
  2.2: The Principle of Credulity .................................................................................... 64  
    2.2.1: Credulity and the Skeptic’s Regress ............................................................ 66  
    2.2.2: The Reducibility-Irreducibility Debate ........................................................ 71  
    2.2.3: Credulity and Gullibility ............................................................................. 82  
  2.3: A Problem for Kant .............................................................................................. 87  
  Conclusion: The Importance of Context .................................................................... 97  
Chapter 3: On Trust and Lying in a Moral Sense ............................................................... 100  
  Introduction .................................................................................................................. 100  
  3.1: The Personal Side of Lying .................................................................................. 100  
  3.2: The Importance of Trust ..................................................................................... 106  
  3.3: Reciprocity in Care: Limitations of the Entrusting Model .................................. 119  
  3.4: Trust and True Beliefs ....................................................................................... 128  
  Conclusion: A Recommended Definition of “Lie” ................................................... 141  
Chapter 4: The Importance of Truthfulness ..................................................................... 144  
  Introduction .................................................................................................................. 144  
  4.1: Truth and Truthfulness ....................................................................................... 148  
  4.2: Truth in the Mahābhārata .................................................................................... 155  
  4.3: Truthfulness in the Upaniṣads and Buddhism ..................................................... 173  
  4.4: Hume on Veiled Truths ....................................................................................... 185  
  4.5: Truthfulness and the Ethics of Truth .................................................................. 188  
  Conclusion: Truthfulness as a Social-Epistemic Virtue ............................................ 202  
Chapter 5: In Defense of Truthfulness ............................................................................ 208  
  Introduction .................................................................................................................. 208  
  5.1: Consenting to be Deceived ................................................................................ 209  
  5.2: Paternalism and Truthfulness .......................................................................... 214  
  5.3: Paternalism and Truthfulness in Medical Ethics ................................................. 218  
  Conclusion: The Responsibilities of Truthfulness ...................................................... 240  
Bibliography ....................................................................................................................... 256
Introduction: Lying and the Ethics of Speaking

In this dissertation I attempt to remedy—or rather, begin to remedy—what I take to be a significant oversight on the part of many ethicists and philosophers of language. Given how important ethical questions are to people, whether they consider themselves philosophers or not, and given how central the use of language is to how we interact with one another, it has long surprised me how little is said in the philosophical literature—and particularly in the analytic tradition—about the ethics of speaking. The mid-twentieth century heyday of emotivism and prescriptivism found meta-ethicists providing detailed accounts of the language of ethics and morality,¹ but the ethics of language use has been left largely undiscussed in the western tradition—and this even after the recognition of speech acts, by Austin and others, as voluntary actions par excellence. Questions about what we ought to do are of course central to much philosophical inquiry; and such language-use issues as meaning, intention, interpretation, and knowledge-gathering from others’ words occupy a position of privilege in philosophy of language and epistemology. But excepting discussions of lying, the intersection of ethics and the philosophy of language is, it seems to me, a woefully under-explored area.

Of course, the exception of lying is a substantial one. Many thinkers have explored the tangle of issues surrounding lying—what it means to lie; when, if ever, lying is ethically justifiable; the nature of the transgression that lying constitutes when it is blameworthy. Thinking on the issue tends to fall into two major camps. Those who are inclined toward a deontological ethical stance tend to see lying as a violation of duty, and thus as always wrong: the moral law dictates that one should tell only the truth (which is

¹ See, for instance, R. M. Hare’s The Language of Morals and Charles Stevenson’s Ethics and Language.
not to be confused with the very different demand to tell the whole truth). Those who incline more toward consequentialist reasoning tend to view lying as sometimes justified, because beneficial: some circumstances are such that the greatest good can only be brought about by lying to someone, and in such circumstances lying is the right thing to do. I want to carve a middle path between these two approaches. I think the deontologist perspective is correct that lying is always wrong, but I think the consequentialist perspective is correct in locating the blameworthiness of lying in the actual consequences of the act rather than in the abstract realm of rights and duties.

The mistake made in most discussions of lying is one of beginning from a definition of what lying is, and attempting to derive ethical conclusions from the description of lying that is taken to be operative. While this is exactly the correct approach with many concepts, it does not work with lying because the issue of lying is an irreducibly ethical one. It is revealing that two of the three central issues surrounding lying that I mentioned above focus on its ethical status. And with any morally weighted concept, such as lying, it is better to embrace the morally weighted nature of that concept and limit its application accordingly than to either deny the moral import of the concept or to attempt to separate the concept from its moral considerations. To identify an utterance as a lie, or an utterer as a liar, is inescapably to cast the entire situation in which the utterance occurs in a negative light.

With regard to lying, the intervention I aim to make with my project is into the extension of the concept—into what qualifies as a lie. On the basis of standard usage, which indicates that lying is a primarily and irreducibly ethical concern, I advocate a shift
in the definition with which we operate—both in everyday usage and in academic
treatments of the subject.

But this is not just a dissertation about lying; it is a dissertation about the ethics of
speaking, broadly construed. The analysis of lying with which I begin reveals that
trustworthiness is the central ethical consideration both in lying and in speaking more
generally. This itself is a somewhat novel realization. I take it to be uncontroversial that
truth is the primary norm of belief: not only do we strive to have true beliefs, but falsity is
a fatal objection to belief. To believe some proposition X is to believe it to be true that
X. Thus, taking X to be false precludes believing that X, at least provided that one
recognizes that the X to be believed and the X that one takes to be false are in fact the
same X. It is widely held that truth is a norm of assertion in the same way as (and largely
because) it is a norm of belief. But this is a mistake. Trust, I argue, is a more basic
ethical good than is truth, and points us in a more fruitful direction in the search for a
norm of assertion. We value true beliefs, and thus true assertions, because they
.generally and for the most part) reward trust.

Recognition that at the core of the concern for truth is a desire to trust allows us to
expand our understanding of what truth itself consists in. Twentieth-century Western
analytic philosophy has tended to take statements or propositions as the primary bearers
of truth. On this understanding, such uses of ‘true’ as found in the notions of a spouse, a
friend, or a tuning fork being true constitute an expansion of the concept—a metaphorical
usage, perhaps. But the focus on trust and trustworthiness reveals that these ‘non-
standard’ uses of the concept true to mean genuine, reliable, and trustworthy may well
indicate the primary meaning, and the notion of a true proposition the extended usage.
The idea of truthfulness as trustworthiness in speaking, or more simply as trustworthy speaking, thus begins to take hold. But truthfulness in this sense must be decoupled from truth-telling, precisely because and insofar as truth-telling is not always trustworthy telling. Truthfulness understood as trustworthy telling rather than as truth-telling has the additional benefit of enabling us to expand our inquiry from the rather narrow concern of the ethics of assertion to the much more broad topic of the ethics of speaking. Questions and commands, for example, cannot be assessed in terms of truth or falsity. This is why truth can be taken (even if mistakenly) as a norm of assertion but not as a norm of speaking. But truthfulness can and should be considered in assessing not just assertions but all utterances. A question that harbors dubious assumptions or that frames an issue in a misleading way, or an exclamation that feigns an emotive response that one does not actually feel, can thus be considered problematic not in some vague and undefined sense, but in the very specific sense of being untruthful. Exaggerations, on the other hand, as well as metaphorical and other sorts of figurative language, even though literally untrue, can still qualify as truthful based on the contribution such utterances make to understanding—contributions that frequently cannot be made, or cannot be made as well, by sticking to the literal, factual truth. Consider the difference between all that I have written on truthfulness and the eight lines Emily Dickinson wrote on the topic:

Tell all the truth but tell it slant –
Success in Circuit lies
Too bright for our infirm Delight
The Truth’s superb surprise
As Lightning to the Children eased
With explanation kind
The Truth must dazzle gradually
Or every man be blind –
Chapter 1: What’s Wrong with Lying?

Introduction

Any philosophical account of lying must address two basic questions about the topic: first, what is it to lie? And second, when, if ever, is it morally permissible to lie? The answers that anyone provides to these two questions are bound to be related; further, the first question seems clearly to be conceptually prior to the second. In his 2006 article “The Definition of Lying,” Thomas Carson states this intuition clearly: “Conceptual questions about the nature of lying are prior to questions about the moral status of lying. Any theory about the moral status of lying presupposes an account of what lying is.”¹

In Section 1 of this chapter I will examine three recent attempts by philosophers to define ‘lie’—all of which, I will argue, are unsuccessful. This failure is not because those trying have not been up to the task of defining ‘lie’, but rather because they have misconceived the project. I will argue that the philosophers considered in this section all have legitimate disagreements about the nature and definition of lying, but no way to resolve them. This is because any attempt to define ‘lie’ by listing necessary and sufficient conditions for set membership is question-begging. Attempts to define it in this way necessarily rely on the intuitions of the person doing the defining—and with regard to lying, as we will see, it is precisely those intuitions that are in question. On the basis of their divergent intuitions, people advocate more or less divergent definitions of ‘lie’. In order to come to an agreement, they would have to agree in advance on whether certain problem cases should qualify as lies, rather than relying on their individual definitions to resolve said cases, because it is their intuitions about what should qualify as

¹ Carson, “The Definition of Lying” 284
a lie that ground their preference for one definition or another. Yet, somewhat ironically, one of the few things that they all appear to agree on is a belief that underlies their whole approach: that the definition of ‘lie’ must be prior to any evaluation of lying. They can be forgiven for taking this to be an uncontentroversial starting point. After all, without a definition, how can we ever come to an agreement on what it means to lie? And if we cannot agree on what it means to lie, then how can we ever hope to agree on whether lying is always wrong or sometimes permissible—and if sometimes permissible, when? But I think this belief is mistaken, and that insofar as it underlies any attempt to characterize lying it also undermines that attempt. It is this belief, and the corresponding practice of beginning with the descriptive question when dealing with lying, that makes quarrels about the definition of ‘lie’ seemingly intractable.

Since our intuitions do not always agree about what qualifies as a lie, the definition of ‘lie’ is a bad place from which to begin reasoning about the concept. In Section 2 I show that it is not just feasible but in fact preferable to begin instead with the moral question. I begin by providing a conceptual framework to show why the approaches considered in Section 1 were bound to fail. To do so, I make use of prototype semantics, a branch of cognitive theory according to which participation in a concept is not all-or-nothing, defined by necessary and sufficient conditions, but rather admits of degrees. An understanding of prototype semantics reveals that the lack of a watertight definition of ‘lie’ that unambiguously resolves all problem cases is precisely what one should expect. I then argue that we can instead look to everyday usage to ground our reasoning about the concept. One thing such usage reveals is that lying is inherently a
morally weighted concept. This is important, I argue, because its moral element renders attempts to use, or define, ‘lie’ in a purely descriptive way implausible.

In Section 3 I consider, in response to the realization that ‘lie’ is primarily evaluative, the relative strengths of two competing views about the morality of lying: that lying is always wrong, and that lying is sometimes permissible. While at first glance this issue appears to be just another case of conflicting intuitions, thereby rendering any approach beginning with the moral question just as fruitless as those beginning with the descriptive question, I provide a novel argument in favor of the view that lying is always wrong.

1.1 The Folly of Trying to Define ‘Lie’

In his 1966 article “Lying” Frederick Siegler seeks to clarify the concept of lying by examining six features that he takes to be present in a typical case of lying. He cautions against the temptation to understand this process as an attempt to delineate the necessary and sufficient conditions for lying, preferring instead an approach that appears influenced by Wittgenstein’s theory of family resemblances: “One might be inclined mistakenly to believe that what is embodied in the typical case of lying represents necessary conditions for any and all cases of lying, but I shall try to show some connections between each of the features and various locutions which make up the cluster surrounding the concept of lying.”² Siegler identifies six features that are present in the typical case of lying, as he understands the concept. He then considers each of these features with respect to one or more problem cases, in an attempt to determine both whether the problem cases qualify as lies and the extent to which each feature is present.

² Siegler 128
in atypical cases of lying. Condition (1), that the liar must say something, is the only condition that he takes to be necessary for all cases of lying. This enables a differentiation between lying and non-linguistic forms of deception.

The primary contribution that Siegler’s analysis provides, aside from the recognition that seeking necessary and sufficient conditions for lying would be a mistake, is his suggestion to divorce the notion of lying from that of telling a lie. A lie, he argues, is a token of speech, whereas lying is an action—merely telling a lie is not sufficient for lying. He says, “the noun which labels what was said (a lie) may be applied when the correlate verb which labels what was done (lie) ought not to be applied.”3 Perhaps more surprisingly, on Siegler’s account telling a lie is not necessary for lying, either. His considerations of conditions (2)—that the liar must intend to deceive—and (3)—that the liar must say something false—in particular lead to these conclusions. When considering the extent to which the intent to deceive is necessary for an utterance to be a lie, he considers the example of someone who is coerced into saying something false to someone who is thereby deceived. Siegler questions whether it is appropriate to say that the speaker intended to deceive the listener. Since the speaker is merely the conduit for the intention of a third party, Siegler argues that it would be odd to say that the speaker intended to deceive the listener; at the same time, it also seems odd to say that someone who utters a false sentence with the expectation that it will deceive the listener does not intend to deceive. Siegler concludes from this that the coerced deceiver tells a lie—because what she says is intended to deceive, and she is aware of this intention—but does not lie—because she is not the author of the intent to deceive.

3 Siegler 130
He uses the same analysis to resolve the problematic case of the deluded deceiver. Of a deluded deceiver—one who attempts to lie but, due to the possession of a false belief, inadvertently says something that is true—Siegler says that it would be odd to say that he told a lie since, *ex hypothesi*, what he said was true, but that it would also be odd to say that he told the truth, since telling the truth seems to imply that one has not lied, “And after all if a man did this regularly we would not want to exonerate him from being a liar simply because he is also a fool.”4 His solution to the deluded deceiver is the inverse of that to the coerced deceiver: the deluded deceiver has *lied*, but has not *told a lie*.

Siegler’s project illustrates what I take to be two fundamental problems with beginning an analysis of lying with the descriptive rather than the moral question—problems that the others I consider in this chapter will run into as well. The first can be called the *problem of various intuitions*. Despite Siegler’s claim that he is not looking for necessary and sufficient conditions for lying, that no single definition can be expected to cover all cases, his analysis of the problem cases relies on his own linguistic intuitions—intuitions that are far from universally accepted, and thus stand in need of independent support. This becomes clear in his analysis of the coerced deceiver. Given that there is an intention to deceive (even if said intention is not the speaker’s5), the question becomes whether this is a case of lying. And whether the coerced deceiver is said to lie will depend on the definition of lying that one adopts. For most, making a statement that one knows to be false, with the intent to deceive the hearer, would qualify as lying. But

---

4 Siegler 130
5 I disagree with Siegler when he says that the coerced deceiver does not intend to deceive. The coerced deceiver may *unwillingly* deceive, but she does not *unwittingly* deceive. This seems to make it clear not only that she did deceive, but also that she did intend to deceive—albeit reluctantly.
apparently not for Siegler, or at least not in this case. And what are his reasons? No
more and no less than, “It seems correct to say that A told the lie to B, but it does not
seem quite right to say that A lied to B.” 6 Here, as elsewhere throughout the article, his
claims that a certain locution “does not seem right” in a certain context are based on
nothing more than his own intuitions. And since he offers no evidence in support of
these claims, it appears that he is simply counting on his audience having the same
intuitions as him. Certainly many will agree with some or all of his intuitions; but it is
equally certain that many will report conflicting intuitions. Siegler has reasoned
admirably from certain intuitions regarding lying, but he has not provided for the
 possibility of reasoning about those intuitions. Since he has not actually argued in favor
of these particular intuitions, whatever characterization of lying follows from them will
be no more convincing than are the intuitions it is based on.

The second problem illustrated by Siegler’s project is what I call the problem of
moral priority. In short, the problem of moral priority hinges on the fact that lying is an
irreducibly moral issue, and attempts to define it by first bracketing or denying the moral
aspect inevitably lead to the moral being brought back in surreptitiously, usually in the
form of unannounced intuitions. It can be best explained by asking a question: why does
it matter whether the coerced deceiver lied or not? Should the moral evaluation of the act
hinge on whether one decides that it merits the appellation ‘lie’? Or rather, should
whether it warrants the appellation depend on an assessment of the act’s blameworthiness
(or lack thereof)? Is there a reason beyond moral assessment that we are concerned with
whether to call a particular utterance a lie, or a particular act one of lying? In fact, raising

---

6 Siegler 130
this issue helps to pin down the intuitions Siegler appears to be relying on when deciding whether a particular locution seems appropriate: his moral intuitions. Siegler, it seems, is loath to say the coerced deceiver *lies* because lying is a morally charged notion and he does not think that someone who deceives under duress should have to bear the blame that it carries. By accepting that the coerced deceiver *tells a lie* he is able to salvage the blameworthiness that he takes to be present in the situation and place it firmly on the noun rather than on the verb—and thus on the utterance rather than on the utterer.

The same analysis is relevant to the deluded deceiver example. It is clear that the deluded deceiver deceives, since he misrepresents his own beliefs, but he fails to deceive with regard to the external world. Siegler says that the intention to deceive (both with regard to his beliefs and to the external world) that the deluded deceiver demonstrates is sufficient to call what he does a *lie*; other philosophers have held the opposite position—that the deluded deceiver attempts, but fails, to lie. But, again, why does it matter whether we call this a lie or not? The answer that Siegler hints at in this case supports the analysis just provided of the coerced deceiver as well: his assertion that “we would not want to exonerate the liar” reveals that he thinks the deluded deceiver should not be let off the hook. His solution to say that the deluded deceiver *lies* even though not *telling a lie* works for him because in this case he relies on the verb to do the moral work that the noun does in the case of the coerced deceiver.

This points as well toward what is valuable, but also dangerous, about the gap that Siegler introduces between the verbal and the nominal forms of ‘lie’. The differentiation is valuable because it allows one to attribute parts of the concept to situations without over-attributing parts that are inappropriate. To this extent, he is actually anticipating
prototype semantics, which we will discuss in Section 2. The danger with the distinction, though, is that there is no rule for how to differentiate. Assignments of both the nominal and the verbal form of ‘lie’ thus become ad hoc, based on one’s intuitions. So while the distinction allows for a more nuanced analysis than is typically on offer with the two packaged, its conclusions are no more authoritative than are those that derive from straightforwardly stipulative definitions of lying—they are just as subject to the problem of various intuitions.

Roderick Chisholm and Thomas Feehan’s 1977 article “The Intent to Deceive” takes a view of lying that differs markedly from Siegler’s, but without engaging Siegler’s account to any significant extent. And because of this, they make largely the same mistakes that Siegler makes. They begin their essay by effectively announcing that they are going to fall victim to the problem of moral priority:

We are here concerned with what it is to intend to deceive and with what it is to lie. We will attempt to set forth a detailed analysis of these concepts and we will use this analysis to answer a number of philosophical questions about the intent to deceive. We are not here concerned with ethical or moral questions. And therefore we will not discuss the conditions, if any, under which a lie or other act of intended deception may be justified or excused.\(^7\)

Their intentions here are understandable, even if ill-advised: the moral question is commonly taken to be the stickier issue, especially if approached without a concrete answer to the descriptive question. But they are no more able to keep the moral issue at bay than is Siegler. Instead of using the problem cases to help them understand the application of the concept of lying, they begin by providing a definition of “lie,” which they base on a detailed analysis of the concepts deception and lie. They then apply this

\(^7\) Chisholm and Feehan 143
definition to questions about problem cases, categorizing them explicitly as lies or non-lies. For Chisholm and Feehan lying requires deception, but it requires a very specific sort of deception. First, and rather uncontroversially, it must be linguistic. More specifically, though, it must involve stating something in such a way and in such a context as to be taken seriously by the audience. They explain the concern with being taken seriously in terms of assertion, where “the concept of assertion is essentially normative. We can explicate it only by reference to justification. And the justification in question is epistemic, the type of justification that is implied by knowledge and evidence.”\(^8\) This leads to the following definition of assertion:

\[
L \text{ asserts } p \text{ to } D =_{df} L \text{ states } p \text{ to } D \text{ and does so under conditions which, he believes, justify } D \text{ in believing that he, } L, \text{ not only accepts } p, \text{ but also intends to contribute causally to } D \text{'s believing that he, } L, \text{ accepts } p. \text{ }^9
\]

And from here, they proceed to provide their definition of a lie:

\[
L \text{ lies to } D =_{df} \text{There is a proposition } p \text{ such that (i) either } L \text{ believes that } p \text{ is not true or } L \text{ believes that } p \text{ is false and (ii) } L \text{ asserts } p \text{ to } D. \text{ }^{10}
\]

Their account of lying relies on a particular notion of assertion, such that the lack of an actual assertion absolves one from the accusation of lying (though not necessarily from the accusation of being deceptive). But proposing assertion as a necessary condition for lying does not provide an answer to the question of why it matters. If we ask why it matters whether \(L\)’s utterance involved an assertion or not, the answer is that it determines whether we can say that \(L\) lied or not. But if we are not concerned with ethical issues, as Chisholm and Feehan claim not to be, then when we ask why it matters whether \(L\) lied or not, no satisfactory answer is forthcoming. If lying is not either more

---

\(^8\) Chisholm and Feehan 151
\(^9\) Chisholm and Feehan 152
\(^10\) Chisholm and Feehan 152
or less blameworthy than other sorts of deception, then what does it matter whether $L$’s utterance involved an assertion? Attempting to differentiate lies from non-lies without explicit reference to the moral issue threatens to yield a concept of *lie* that fails to do any work because it is undifferentiable from other forms of linguistic deception.

Based on the definition of lie that they provide, Chisholm and Feehan agree with Siegler that the intention to say something false, and thereby to deceive, is a necessary condition for lying, but they differ from him in holding that the incidental truth of an utterance that is intended to deceive is irrelevant to considerations of whether someone has told a lie. Because they see the intent to deceive as more central than actual falsity to the concept *lie*, they think it clear that the deluded deceiver lies. And since they do not follow Siegler in divorcing the verb and the noun, they must (at least implicitly) take the deluded deceiver both to lie and to tell a lie.

What is perhaps most interesting about this account is not so much *where* Chisholm and Feehan disagree with Siegler, but rather *how* they do it. While at some points they do agree with, and even rely on, Siegler’s reasoning, where they disagree they do not argue against his claims; they simply ignore them. This is not, presumably, because they are irresponsible scholars, but rather because an argument here would be inappropriate; they are just reporting conflicting intuitions, even if they do not always clearly recognize that this is what they are doing. But insofar as they are merely reporting intuitions, especially if they are conflicting intuitions, they confirm that both they and Siegler have fallen victim to the problem of various intuitions.

---

11 Even though Siegler eschews talk of necessary and sufficient conditions, his unwillingness to say that the coerced deceiver lies, on the basis of his claim that she is not the one who intends to deceive, reveals that he views the intent to deceive as *something like* a necessary condition for lying, even if not for telling a lie.
Much the same thing occurs in Thomas Carson’s 2006 attempt to define ‘lie’. He begins his analysis by insisting that the deluded deceiver does not lie: “Showing that a statement is true is always sufficient to counter the accusation that one has told a lie.”12 At least Carson admits that this part of his account amounts to no more than an expression of his own intuitions:

My linguistic intuitions tell me that a lie must be a false statement, and that, therefore, what I say in this case is not a lie. I intend to lie in this case but I don’t. Others report conflicting intuitions about this case and the question of whether a lie must be a false statement. … To the extent that it rests on disputed intuitions, my claim that a lie must be a false statement is open to question.13

When he revisits the issue of lying in his 2010 book Lying and Deception, Carson makes more of the problem of various intuitions. For instance, he acknowledges the point I have been arguing, that “[a]ppeals to moral intuitions are not adequate to justify judgments about controversial moral questions.”14 But given this recognition, the only explanation for the approach he takes in the book, which does not differ appreciably from that in the article already discussed, must be that he does not take lying to be a controversial moral issue. He says,

I do not claim to be defending surprising results or giving answers to difficult cases that involve serious moral dilemmas. Rather, I am trying to answer many practical moral questions without appealing to disputed moral intuitions…. I am answering skeptical, egoistic, and relativist challenges to conventional moral beliefs about issues of lying and deception and vindicating the core of most people’s common sense moral beliefs about lying and deception.15

---

12 “The Definition of Lying” 284
13 “The Definition of Lying” 285
14 Carson, Lying and Deception 5
15 Lying and Deception 165
Three issues arise out of this description of his project. First is that, as we have just seen, Carson explicitly does appeal to disputed intuitions, so that what he says in this passage is not entirely accurate. Carson might respond that he does not appeal to disputed moral intuitions, as the question whether a lie must be a false statement is a descriptive rather than a moral question. But the question whether one has lied or not is clearly a moral question. And to decide this moral question by appeal to non-moral intuitions that are also in dispute is not really to resolve the issue at all. Appealing to disputed non-moral intuitions is no more justifiable than is appealing to disputed moral intuitions.

The second issue is the concern that most people’s common sense moral beliefs about lying and deception might not warrant vindication. A belief does not warrant vindication just because a majority, or even a plurality, of people from a particular population or sub-population believe it. It might be the case that most people’s beliefs about marriage include that it should be between one man and one woman. This would not mean that philosophers should be in the habit of attempting to argue in support of that belief. The marketplace of ideas is not a democracy, it is a meritocracy; a philosopher’s obligation, much like a scientist’s, is very often to argue in favor of unpopular views because that is what the available evidence dictates.

The third issue is that Carson is, his claims to the contrary notwithstanding, defending surprising results. Most prominently, he claims that the intent to deceive is not a necessary condition for an utterance to qualify as a lie. He provides two examples that he thinks conclusively demonstrate that one can lie without intending to deceive. The first case he considers is that of an eyewitness to a crime who is called to court to testify as a witness, but who falsely testifies that he did not see the defendant commit the crime
because he fears for his well-being if he tells the truth. The witness in this case, Carson says, may even be hopeful that his testimony does not convince the jury and that the defendant is convicted despite the false testimony. But even if the jury is deceived, Carson insists, this deception is incidental—since the intention behind the false testimony is self-preservation rather than the deception of the jury, the witness cannot be said to intend to deceive the jury.\textsuperscript{16}

In case this version of the example is unconvincing, Carson provides a modified version—one wherein the witness knows that the whole crime, including his own witnessing of it, was caught on camera, and that the video is shown to the jury. In this case, Carson argues, the witness clearly cannot intend to, or even reasonably believe that he will, deceive anyone either that what he says is true or even that he believes what he says. But the fearful witness still lies, Carson insists, by virtue of saying something that he knows to be false while under oath.

The fact that the witness is under oath is important here, for Carson does not take it that merely saying something that one knows to be false is sufficient for lying. If one speaks sarcastically, for instance, no lie is told. The relevant issue for lying, he says, is not the intention or desire that others believe what one says, but rather simply being on record as having made the assertion. This explains why he takes his second example to also be a case of lying without the intent to deceive. The second case is an example of what Roy Sorensen has recently called a “bald-faced lie”. Carson tells the story of a

\textsuperscript{16} I disagree with Carson’s analysis of this case. I think it can be analyzed in much the same way I analyze Siegler’s coerced deceiver: since the deception is not unwitting, it does make sense to say that the speaker intends to deceive the audience, even if he simultaneously hopes that he fails at his project. Implicit in Carson’s argument is the idea that any action can be described by only one intention. But this idea is implausible. If I intend to save myself by deceiving the jury, then I intend both to save myself and to deceive the jury.
college student who has been caught cheating on an exam, but who knows that the Dean (who is overly afraid of lawsuits from unhappy students) will not punish for cheating any student who does not admit guilt. If the Dean has absolute confidence in the professor who caught the student cheating, and the student knows this, then the student can have no pretensions to convince the Dean of his innocence—as Carson says, the student may even take a malicious pleasure in knowing that the Dean knows he’s lying—but this does not change the fact that the student is lying. The bald-faced lying cheater is interested in being on record as having declared his innocence, not in convincing people of his innocence; in this he is like the fearful witness.

The idea of being on record as having declared something leads Carson to focus on the notion of warranting as the crucial component of lying. In order to lie, he says, one must knowingly warrant the truth of something that she knows to be false. One warrants the truth of something when “one promises or guarantees, either explicitly or implicitly, that what one says is true.” He allows that the notion of a promise is not entirely parallel to that of warranting, but maintains that it provides a helpful analogy. He also says that in most linguistic communities, “there is a presumption that the warranty of truth is in force in any situation. Convention dictates that one warrants the truth of one’s statements in the absence of special contexts, special signals, or cues to the contrary.”

Carson is correct that, if the two cases he presents are in fact cases of lying without the intent to deceive, then most philosophers’ definitions of lying (and certainly

---

17 “The Definition of Lying” 294
18 “The Definition of Lying” 294
those that we’ve considered in this chapter) are in need of some amount of revision or refinement. To consider just Chisholm and Feehan’s definition, Carson’s examples appear to undermine their reliance on assertion. Remember that for Chisholm and Feehan, to assert $p$ is, *inter alia*, to state $p$ under circumstances that one believes justify another in believing that one accepts $p$. But it is precisely this requirement that is not satisfied in the cases of the fearful witness and the cheating student. The question, though, is whether Chisholm and Feehan would accept these as examples of lying. And the best place to turn for the answer to this question is their treatment of the seventh problem case that they consider. This is the case wherein one person says something false to a knowing accomplice with the intention of deceiving a third person, who is listening in on the conversation.¹⁹ They argue that the deceiver here has lied neither to the accomplice nor to the eavesdropper, because to neither of them has he *asserted* what he stated. What he says to the accomplice is clearly not an assertion according to their definition, as he does not take himself to be giving the accomplice any justification for believing that he accepts what he says. But precisely because his utterance does not directly address the eavesdropper, he cannot be said to be asserting anything to the eavesdropper either. Recall that the second condition for one to assert $p$ to $D$ is to state $p$ under circumstances that one believes justify $D$ in believing that one intends to cause $D$ to believe that one accepts $p$. On this account it is possible to deceive, even to intentionally deceive, an eavesdropper, but it is not possible to lie to an eavesdropper.

The case of the eavesdropper is clearly not precisely the same as the cases that Carson provides, but Chisholm and Feehan’s treatment of it casts some doubt on the idea

---

¹⁹ Chisholm and Feehan 156
that they would accept Carson’s cases as true counterexamples. Precisely to the extent that the fearful witness and the cheating student think that they will not be believed in their claims, they are not asserting anything—and thus, for Chisholm and Feehan, they cannot be said to be lying. At most, we could say that the witness and the cheater are speaking as if they were asserting, but that a necessary condition for assertion is missing. In Austin’s terminology, the would-be assertion is infelicitous; in Searle’s, it is defective. This would be much like the case of a person attempting to promise another person something that the other does not want—one can speak as if to promise, but if the recipient does not want or would not like the thing promised, then no promise is made. Carson’s dispute with Chisholm and Feehan, then, is qualitatively the same as Chisholm and Feehan’s dispute with Siegler—it is simply a matter of competing intuitions vying for supremacy.

It is sometimes said that one person’s modus ponens is another’s modus tollens, and this seems to be particularly applicable to the current situation. One could say that in order to be a lie, any utterance must have properties x, y, and z; A lacks property z; therefore, A is not a lie. But one could, with equal validity, say: A is a lie; A lacks property z; therefore, an utterance need not have property z in order to be a lie. As Carson admits, and the other philosophers we have considered have demonstrated (whether intentionally or unintentionally), which of these arguments one adopts depends to a very large extent on one’s own intuitions about lying. And our intuitions about lying are more a matter of assertion than argument. Carson says that

Chisholm and Feehan’s definition has the very odd and unacceptable result that a person who is notoriously dishonest couldn’t tell lies to those he knows distrust him. Their definition implies that it is self-contradictory
to say that I lie when I know that others know that I am lying (and thus are not justified in believing that I believe (accept) what I say). He thinks it a virtue of his account that it avoids this consequence. I must admit that my intuitions favor Chisholm and Feehan’s definition on this matter. It does not seem to me at all counterintuitive, let alone unacceptable, that a serial liar would eventually be unsuccessful in her continued attempts to lie because her tendency becomes common knowledge—and that such a wolf-crier could thus be said to end up only attempting to lie, with such would-be lies repeatedly misfiring because of the lack of trust in her relationships. But I do not want to get dragged any further into this debate, as it is bound to be largely fruitless. The problem with debates about the proper application of concepts that cannot be defined with the precision found in Euclidian geometry is that the paradigm cases provide surprisingly little help in dealing with non-paradigm cases. In such cases, we are often left to consult our own intuitions; and if our intuitions disagree, then there is seemingly little we can do but wash our hands of the matter, agree to disagree, and call the issue an open question.

1.2 The Evaluative Nature of ‘Lie’

I once got into a debate with one of my philosophy professors over whether the Pope is a bachelor. He insisted that the Pope, being an unmarried male of marriageable age, clearly fits the definition; I maintained that, the possibility of marriage being unavailable to the Pope, it would be inappropriate to call the Pope a bachelor and thus, everyone being either a bachelor or not a bachelor, clearly the Pope is not a bachelor. My...
professor and I were both wrong in our reasoning, but our debate illustrates an important insight of prototype semantics: many of the categories we operate with presume an overly simplified context for their application, and consequently when we try to apply those concepts to the vagaries and complexities of the real world we find that it is not always entirely clear whether or to what extent a given case falls under a particular concept. As Eve Sweetser puts it, “Lexical categories can have better or worse members, or partial members.”²² And with particularly troubling or novel cases, the fact that people’s intuitions differ may simply be because there is no fact of the matter. It may not be true that what the case provides either is or is not a member of the category. And, unless there is some legitimate authoritative body that can stipulatively resolve the issue, there may be no prospect for the resolution of conflicting intuitions.²³

The concept bachelor is based on a simplified worldview in which men want to get married, mostly marry at or around a certain age, and remain married throughout their lives. Given this worldview, it is understandable that a male of the relevant age who is not married would be seen as an aberration, and further that all males of the relevant age who were not married would be seen as similarly aberrant, thus warranting their own category. But the concept bachelor, presuming a simplified world that makes possible this unambiguous category-membership criterion, is ill-equipped to handle the complexity of the real world. If a bachelor is just an unmarried male of marriageable age, then an adult homosexual male living with a committed life-partner is a bachelor; but this is not a consequence anticipated by the simplified worldview. This understanding

²² Sweetser 43.
²³ Referring to a dictionary will not solve the problem either, at least for the most novel cases: dictionaries describe word usage within a particular linguistic community, they do not prescribe it.
explains why it is so difficult to say with a great deal of confidence whether a 35-year-old twice-divorced widower—or the Pope—is a bachelor. They are not prototypical cases, but neither are they clearly ruled out by comparison to the prototypical case.

Sweetser argues that the concept *lie* is based on an oversimplified worldview in much the same way *bachelor* is, and that a complete understanding of the concept must be based on an examination of the folk theories that comprise it. She seeks to ground theoretically the experimental results of Linda Coleman and Paul Kay, which show that the usage of the concept *lie* is best explained by prototype theory. Sweetser describes their findings:

Clear central cases of lies occur when all of Coleman and Kay’s proposed conditions are fulfilled; namely, (a) speaker believes statement is false; (b) speaker intends to deceive hearer by making the statement; and (c) the statement is false in fact. Conversely, a statement fulfilling none of a-c is a clear nonlie. But when only one or two of a-c hold, speakers are frequently confused and find it difficult to categorize an action as lie or nonlie. Further, these conditions (unlike checklist-features) differ in weight, (a) being strongest and (c) weakest in influencing speakers’ categorization of acts as lies.24

Just as *bachelor* presumes an oversimplified world regarding issues related to marriage, Sweetser points out that *lie* presumes an oversimplified world regarding issues related to human interaction and language use. In the case of *lie*, the presumption is that language is used to inform—and not, for instance, to tell jokes or to maintain social harmony. Sweetser identifies two more general principles of social interaction that ground this presumption about language. They are:

(1) Try to help, not harm and (2) Knowledge is beneficial. Together, the two principles yield the result that knowledge (since it is beneficial) is part

---

24 Sweetser 43-44
of a general goal of helping others. Thus, in cases in which (2) is true, (1) translates at least partly as (3) Try to inform others.25

She then presents an analysis that explains how a folk understanding of language as informational would understand conditions (a), (b), and (c) as equivalent—thereby demonstrating that the concept lie is just as straightforward in its own (oversimplified) context as the concept bachelor is in its.

Before moving on to that analysis, though, it is worth pausing to point out that one consequence of what Sweetser has said up to this point is that in those real-world cases wherein knowledge is not beneficial, there can be no presumption that the maxim to inform others holds. In fact, in cases in which knowledge would be harmful, the helpfulness maxim could even entail attempting to misinform others. This issue will become important in Chapters 3 through 5.

Sweetser’s analysis of the folk understanding of language as informational begins with an examination of our cultural model of knowledge. She is not interested in abstruse philosophical accounts of knowledge, but rather in “our cultural idea of what counts as knowledge, since this is what underlies our understanding of lies and truths in discourse.”26 She points out that we are not able, nor do we expect to be able, to prove all our beliefs either logically or by firsthand experience; yet we do not normally take this inability to be an objection to claims of knowledge. “Whatever our rules of practical everyday inference are like,” she says, “we trust them, in the default case. Thus, belief is normally taken as having adequate justification, and hence as equivalent to knowledge,

---

25 Sweetser 45
26 Sweetser 46
which would entail truth.”

Given the folk understanding of language that has been so far characterized, someone who is speaking can reasonably be understood as attempting to inform her audience. And given the cultural model of knowledge that Sweetser has just described, someone who is attempting to inform her audience is not just stating what she believes, but is actually telling the truth. It follows from this that if you say $X$, I can infer not only that you believe that $X$, but also that you know that $X$; and since knowledge entails truth, within the cultural model I can infer from your saying $X$ that $X$ is true.

Sweetser then concludes:

Logically (outside our model), or statistically, this conclusion is rubbish. But as a folk model of language by which we all operate from day to day, it makes good sense – in fact, it seems doubtful that we could ever live our lives questioning the truth of every statement presented to us. We question truth if we fear that our simplified discourse-world is too far from reality: when our source might be ill-informed (a broken link between belief and justification), naïve (breaking the entailment between justification/evidence and truth), or might want to deceive us (invalidating our assumption that folks are out to help, and so wish to inform correctly).

This analysis explains why the prototypical case of lying is what it is. The three possible breakdowns that she describes—source ill-informed, naïve, or deceptive—amount to the same thing within the simplified worldview: if $X$ is false, then you do not believe $X$; which means that if you said $X$, then you lied. In the simplified worldview, either you tell the truth or you lie. There is no middle ground, and there are no ambiguous cases.

“Thus, we can define lie as a false statement, if we assume the statement occurs in a prototypical (informational) speech setting. This definition is elegant and would also

---

27 Sweetser 47
28 Sweetser 48
help explain why native speakers tend to define *lie* as a false statement.”

Sweetser is correct, but she could actually make the stronger claim that we can define *lie* as a false statement *only if* we assume the statement occurs in a prototypical speech setting. If it does not occur in a prototypical speech setting, then the purpose may not be to inform, which means the possibility is immediately opened up that at least some false statements are non-lies—and also that not all lies are false. And since the prototypical speech setting is based on a drastically oversimplified view of language use, we clearly cannot often, if ever, assume that a given statement occurs in a prototypical speech setting. This is why the elegant definition does not work in the real world.

But the correct response to this realization is not to declare, “We need a better, more nuanced, or more sophisticated definition!” The inclination behind this demand is on the right track, but if the question asked is, “What, exactly, is a lie?”—that is, if we strive simply to provide a definition of *lie* that will work in the unsimplified world of reality—then we will run afoul of the *modus ponens/modus tollens* problem all over again. As we saw in Section 1, stipulative definitions of lying based on one’s intuitions about what does or does not count as a lie tend to disable, rather than enable, productive discourse. Because of this, the most responsible approach to understanding lying is to suspend judgment about those characteristics that are not universally accepted as being necessary characteristics of lies. Thus, at the beginning of this inquiry, we must leave open the possibility that the deluded deceiver, the fearful witness, and the cheating student are all lying. Other so-called problem cases that have not yet been considered must also be included. We must not, for instance, rule out the possibility that someone

---

29 Sweetser 50
who intentionally deceives while carefully choosing his words so as to not utter a falsehood is lying. Consider, in this capacity, Molière’s play Tartuffe. When the title character is accused of attempting to seduce Elmire, his benefactor’s wife, he responds to the accusation with complete honesty:

Yes, Brother, I’m a wicked man, I fear:
A wretched sinner, all depraved and twisted.
The greatest villain that has ever existed.
My life’s one heap of crimes, which grows each minute;
There’s naught but foulness and corruption in it;
And I perceive that Heaven, outraged by me,
Has chosen this occasion to mortify me.
Charge me with any deed you wish to name;
I’ll not defend myself, but take the blame.
Believe what you are told, and drive Tartuffe
Like some base criminal from beneath your roof;
Yes, drive me hence, and with a parting curse:
I shan’t protest, for I deserve far worse.  

What Tartuffe says in this passage is true, if overstated, but he says it in such a way as to ensure that the one person who he most needs to keep in the dark regarding its truth will not believe a word of it. It is definitely a deceptive utterance, but by virtue of the fact that it is both true and believed to be true by the utterer, it does not fit very many people’s definition of a lie. I am not (yet) saying that we should call Tartuffe’s speech a lie, but I am saying that we should not rule it out simply because no falsehoods (or perceived falsehoods) were uttered. By the same token, we must leave open the possibility that one could lie through conversational implicature. Disagreements about whether

---

30 Molière, Tartuffe III, vi
31 I suspect that I will lose some of my readers on this point. If you find it absolutely implausible to think that Tartuffe has lied here, I ask you whether this feeling is based on a preconceived definition of lying, and whether that definition is based on anything more than intuitions—either your own or someone else’s that you have adopted as your own.
32 “Conversational implicature” is a term coined by Paul Grice to describe something that is not logically implied, but is nonetheless conventionally suggested, by what a person says. For instance, if I tell you that I used to be addicted to gambling, you will likely draw the conclusion that I am no longer addicted—as that
misleading use of conversational implicatures qualifies as a lie are very likely, as Sweetser suggests, related to “genuine disagreement about the degree to which a conversational implicature constitutes a ‘statement’ and hence makes the speaker responsible for having said it.” But since it is a matter of serious disagreement, we must not presume to know the answer before exploring the concept’s use.

The one descriptive characteristic of lying that seems to be universally (or universally enough for our purposes) accepted is that lying must involve an utterance. But even this warrants a clarification that is not often made in discussions of lying: namely, that silence—or, more specifically, refraining from speaking—can in some circumstances qualify as an utterance. That is, an utterance can be made without uttering anything at all, in what might be termed an “unspeaking utterance.” In a context wherein one can reasonably be expected to speak—say, in the middle of a conversation, when asked a direct question, and otherwise normal circumstances obtain—choosing not to say anything is just as much a use of conventional signs to communicate meaning as is speaking. A husband who is asked point-blank by his wife whether he has ever cheated on her may use silence to convey offense that she would even ask the question, or he may use silence to convey that he has in fact cheated on her (interpretation of unspeaking utterances is no less varied—and consequently no less difficult—than is interpretation of standard utterances), but in saying nothing he is most definitely conveying something. His silence can thus plausibly be considered an utterance. Something like this happens in

---

is typically what is meant by the locution “used to”—even though it does not necessarily follow from what I said. There is nothing logically inconsistent in me following up my statement that I used to be addicted to gambling by saying, “Indeed, I still am.” I will discuss conversational implicatures in more detail in Chapter 4.

33 Sweetser 60; see also Adler, “Lying, Deceiving, or Falsely Implicating.”
the movie *All the President's Men*. In one scene, the character Carl Bernstein, a reporter for the *Washington Post*, needs confirmation from one of his sources that the story they are about to run is accurate. The problem is that his source refuses to say anything, even off the record, about one of the key people in the story. Bernstein says that he is going to count to ten, and asks the source to hang up before he gets to ten if there is anything wrong with the story. When he gets to ten the source is still on the line. Because the source indicated an understanding of the gambit before Bernstein started counting, his silence coupled with his choice not to hang up clearly counts as the use of conventional signs to communicate meaning, and for our purposes therefore qualifies as an unspeaking utterance.

Much more frequently recognized, and equally important, is the recognition that “linguistic” must be understood to include conventional signs beyond spoken language. So nodding would be linguistic, but limping (say, to indicate an injury) would not; writing would, of course, be linguistic, but painting normally would not be. Throughout this essay I will call all linguistic acts “utterances” regardless of the medium, and I will call the purveyor of the utterance the utterer, and the recipient the audience.

Because there is so little agreement to be found regarding the proper description of the concept *lie*, the examination to be conducted into the use of the concept should not be into *where* the concept gets used—what gets classified as a lie and what as a non-lie—but rather *how* the concept gets used. Or, to put it in an anti-Socratic but pro-Wittgensteinian manner: “What is a lie?” is not the right question to ask; “How is the cluster of terms surrounding the concept *lie* used?” is far more productive. If we want to understand more about the concept, we should look to usage: how a word is used
throughout a linguistic community will tell us more about what a word means than will one philosopher’s (or one hundred philosophers’) intuitions. A failure either to understand or to consider how the word is used leads to the wrong questions being asked, and thus to unhelpful and unsatisfactory answers.

‘Liar’ is used as a term of reproach, and not simply as a way to describe someone’s utterance as false. I can no more call someone who has spoken falsely ‘liar’ and be surprised if he expresses offense (or shame) than I can call someone of African descent ‘nigger’ and be surprised if he expresses offense. Granted, at least part of what makes it apposite to call someone a liar has something to do with that person’s having spoken falsely, but such description is not all—or even most—of what the concept lie conveys. ‘Lie’ and the cluster of terms surrounding it are used as a form of moral condemnation; saying someone has lied is not just an assertion, it is an accusation.\(^\text{34}\)

Consider the following anecdote. In August of 2004 Paul Krugman and Bill O’Reilly appeared together on the CNBC show *Tim Russert*. At one point O’Reilly criticizes Krugman on the basis of things he claims Krugman had said in a series of articles the latter wrote in 2002. Krugman responds by saying, “I’m sorry, that’s a lie.” In response to this O’Reilly interrupts his own discussion of the veracity of his claims to point at Krugman and say, “Don’t call me a liar, pal.”\(^\text{35}\) O’Reilly clearly seems to take the accusation of lying to be not simply a statement about the veracity of his claims, but

\(^{34}\text{In Lying and Deception Carson makes the same point with regard to the notion of deception. He claims that the words “deceive” and “deception” are most typically used in evaluatively critical ways, and thus that it makes little sense to speak of unintentional deception, which would typically be blameless (p. 47). I find it interesting that he chooses to take this approach with deception but not with lying, as the assessment of lying as blameworthy has greater claim to universality than does that of deception. Deception can be described purely descriptively, as the (intentional) conveyance of a false belief; lying cannot.}\)

\(^{35}\text{Video of the discussion can be found at http://www.youtube.com/watch?v=MUOFTPbuxWA; a transcript is available at http://poorandstupid.com/2004_08_08_chronArchive.asp#109207098954281902.}\)
rather an attack on his character. Whereas elsewhere throughout the discussion O’Reilly consistently refers to “Mr. Krugman,” at this one point he reverts to the blatantly derisive “pal,” as if to indicate that claiming someone has lied is in itself disrespectful enough to warrant disrespectful treatment in return. The issue at stake in this exchange is ostensibly the truth or falsity of (some of) O’Reilly’s claims, but this part of O’Reilly’s response focuses on what he takes himself to have been called, rather than on the characteristics of what he said; it is the use of “lie” that gives rise to a shift from speech-act evaluation to character-evaluation. And O’Reilly’s reaction is not entirely unreasonable: it is quite possible that Krugman’s word choice was intentional, and that the word choice was intended as a (perhaps thinly veiled) statement about O’Reilly’s character. Certainly, if Krugman had opted to say, “I’m sorry, that’s not true,” instead of using the word lie, it is

---

36 This restriction of scope is important to note because of the dispositional nature of being a liar. Even if Krugman has claimed that in this case O’Reilly has lied, he has demonstrably not said that O’Reilly is a liar. It is an important question—and one that it is unfortunately beyond the scope of this essay to address—how much or how often one must lie in order to properly be called a liar. One need only commit one murder, or one rape, in order to be a murderer or a rapist. But it would be extreme to the point of absurdity to insist that anyone who has ever told a lie is a liar, on the supposed ground that only liars lie. A person who has lied only once is not a liar, any more than a person who has told the truth only once is truthful. To use an Aristotelian example, to say of someone that she is courageous is not to say either that she is courageous in every moment or that she has never acted in a manner inconsistent with courage. It is, rather, to say that she demonstrates a consistent tendency (or disposition) to act in a courageous manner when presented with the opportunity—and that she either seeks out, or at least does not actively avoid, such opportunities. Certainly there are moments regarding which the issue of courage does not apply. One need not sleep in a courageous manner in order to be appropriately described as courageous. (Though it is possible to sleep courageously. Socrates’ ability to sleep soundly before his impending execution might qualify as an example.) Further, one can be courageous despite having occasionally acted in a cowardly manner. Dispositional attributes such as being courageous or being a liar either apply or do not apply to a person because of that person’s overall body of work—their actions over (some significant portion of) a lifetime—rather than individual events in isolation.

The pedagogical implications of the dispositional nature of lying are worth exploring as well. If one does not want one’s children to become habitual liars, then what sort of restrictions should be placed on individual cases of lying? If the restrictions are too strict, they serve paradoxically as an encouragement to give up: if I believe that once I have told a lie I am forever a liar, then what motivation is there for me to avoid embracing my newfound nature and acting as a liar? At the same time, if the rules are too lax, one risks raising habitual liars anyway. A certain willingness to pardon transgressions is necessary, but just how much willingness is appropriate is difficult to assess. It is important to note, though, that the command “Do not ever lie!” is markedly different from the command “Do not be a liar!”—in the same way that the command “Do not ever surf!” differs from the command “Do not be a surfer!”
hard to imagine that O’Reilly would have responded with similar vehemence. Had he wanted to for rhetorical reasons, he would have had to bring in the concept of lying himself—“Are you calling me a liar, pal?” The same moral outrage just does not come across with, “Are you saying what I said was false, pal?”

The idea that moral blameworthiness is inherent to the notion of lying is supported by the results of recent psychology experiments regarding lying. In one experiment, four- and five-year-old children were read a story about a misinformed protagonist who accidentally tells a falsehood when attempting to tell the truth.\(^\text{37}\) The children were more likely to say that the protagonist deserved to be punished if they were asked whether the protagonist lied before they were asked about punishment than if they were asked after. If the children took lying to be a morally neutral concept, we would not expect this result. In another study, children ranging from preschool to fifth grade not only disapproved more of lying than of truth-telling, but also disapproved more of doing something bad and lying about it than of doing something bad alone—which indicates that their disapproval of lying was not simply a carryover effect from their disapproval of the misdeed.\(^\text{38}\)

Studies of children’s evaluation of lies are of limited value in trying to understand how adults reason about the issue, however. Unfortunately, substantially fewer such experiments have been conducted on adults, so the evidence is more limited. One relevant study examined Chinese and Canadian adults’ evaluations of “lies” told in prosocial and antisocial contexts.\(^\text{39}\) The results of this experiment indicate that Canadian

\(^{37}\) Wimmer, et.al.
\(^{38}\) Bussey
\(^{39}\) Fu, et.al.
adults are significantly more likely than are their Chinese counterparts to identify as a lie the utterance of someone who falsely denies having done a good deed. Moreover, while all groups rated more prototypical antisocial lies as worse than these prosocial lies, those Chinese who identified the prosocial falsehoods as lies evaluated them less positively than did those who identified them as non-lies. The Canadians, who were consistent in calling prosocial falsehoods lies, rated them more negatively than did either of the Chinese groups. There is no way to determine from this experiment what, if any, causal relationship there is between lie-attribution and moral assessment, but even the presence of a correlation provides support for the general claim that lie is a morally weighted concept, conveying some amount of moral condemnation.

But just saying that lie is a morally weighted concept does not capture everything that everyday usage reveals about it. The evaluative element of a morally weighted concept can be either primary or secondary. Cowardice is an example of the former. Though there is undeniably a descriptive element in the notion, if I call you a coward I do not just claim this description and only incidentally suggest moral blameworthiness—I explicitly and unequivocally assign such blameworthiness to you. Contrast this with irrationality, which is also a morally weighted concept but which is only secondarily evaluative. If I call you irrational I am explicitly committed to the descriptive claim that some action, behavior, or decision of yours fails to conform to a certain standard, but I only incidentally imply that this is a blameworthy failure.\(^{40}\) Because irrationality is secondarily evaluative, a purely descriptive use of the concept, wherein one denies that

\(^{40}\) That blameworthiness is implied by calling a particular decision irrational is more easily seen by contrasting it with calling something non-rational, which is purely descriptive and thus does not imply blameworthiness.
the evaluative element is in play, is at least plausible. But with cowardice a purely
descriptive use is not plausible. Just because you act out of fear does not mean that I can
call you a coward, then defend myself against your accusations of insult by saying that
your action fits the definition. Blameworthiness is essential to the concept. If the action
in question possesses the descriptive characteristics of cowardice but lacks the moral
ones—if it is an act of fear, but not an act of excessive fear, perhaps—then it is not an act
of cowardice. And to claim otherwise is to misapprehend the role cowardice plays in the
English language.

Lying, like cowardice or rape, is a primarily evaluative notion. It carries
descriptive content—though, as we have seen, precisely what that content is is a matter of
some dispute—but ‘lie’ and ‘liar’ are primarily terms of criticism. And, like cowardice
or rape, its being primarily evaluative has implications for the plausibility of any attempt
to use the related terms in non-evaluative ways. If Krugman were to respond with
surprise to O’Reilly’s indignation, defending his claim by stating that what O’Reilly said
did in fact fit the descriptive definition of a lie, then we should be unsurprised if O’Reilly
remained unpacified. Lying is subject to the same analysis as was just provided for the
notion of cowardice. To say that some utterance is a lie, yet not blameworthy, is
inconsistent with lying’s being a primarily evaluative concept. The utterance may have
the descriptive characteristics of a lie, but because it lacks the blameworthiness there is
something implausible about calling it a lie. Because ‘lie’ does not clearly describe any
particular type of utterance—a lie may or may not be false, and it may or may not be
intended to deceive, depending on who you ask—if it also may or may not be
blameworthy, then it is an empty concept: it does no work. But clearly ‘lie’ does some
work. And that work is to point out some sort of moral culpability on the part of the utterer of the lie.

1.3 Why Lying is *Always* Wrong

The idea that lying is categorically wrong has a long and distinguished pedigree. Perhaps the two most prominent exponents of this view are Saint Augustine and Immanuel Kant. Augustine identified eight types of lie, differentiated according to their motives and their consequences, and insisted that each of them violated God’s will and was thus morally inexcusable:

Now it is evident that speech was given to man, not that men might therewith deceive one another, but that one man might make known his thoughts to another. To use speech, then, for the purpose of deception, and not for its appointed end, is a sin. *Nor are we to suppose that there is any lie that is not a sin, because it is sometimes possible, by telling a lie, to do service to another.*

In the modern era, Kant is well known for advancing a similarly absolutist condemnation of lying. In his essay “On a Supposed Right to Lie From Philanthropy” he rejects the idea that telling the truth is a duty only to one who has a right to the truth, and argues instead that one must not lie even to protect an innocent friend from an unjust fate. Kant insists that a lie “always harms another, even if not another individual, nevertheless humanity generally,” and that “To be *truthful* (honest) in all declarations is…a sacred command of reason prescribing unconditionally, one not to be restricted by any conveniences.”

---

41 Augustine, *The Enchiridion*, XXII (emphasis added)
42 Kant, “On a Supposed Right to Lie from Philanthropy” (hereafter SRL) 8:426
43 SRL 8:427
But however distinguished the pedigree, the view that lies are without exception wrong strikes many as careless and simplistic. Even Augustine admitted that some lies were worse than others—and, as Sissela Bok points out, no sooner had Augustine’s views come to hold sway in the Catholic Church than people were looking for ways around them.⁴⁴ And Kant is roundly criticized, even by those who otherwise consider themselves Kantians, for going overboard in his prohibition of lying, for failing to recognize that other duties can in fact conflict with the duty for truthfulness. These critiques tend to view Kant as similar to Alceste, the lead character in Molière’s *The Misanthrope*, who provides a clear illustration of the absurdly anti-social and self-righteous behavior that can occur if one takes the prohibition of lying too far. Alceste becomes morally outraged at every falsehood he encounters, regardless of how well-intentioned or inconsequential it may be. Even falsehoods told out of politeness he takes to be an abomination: “I want us to be men and say what we really mean in all circumstances. Let what we have in our hearts be apparent in our words; let it be our hearts that speak, and let us not allow our feelings to be concealed under a mask of empty compliments.”⁴⁵ Like Alceste, Kant seems to be overeager in condemning all lies. After all, it seems obvious that legitimate extenuating circumstances sometimes obtain which render lying, if not morally obligatory, then at least morally permissible. And a failure to recognize this reality can itself be seen as a moral failing—a failing that can lead, for instance, to the belief that it is okay to allow one’s innocent friend to be killed because one refuses to lie to the would-be killer.

⁴⁴ Bok 34-39. Bok describes three common responses to the universal prohibition of lying, each of which brought with it its own difficulties: “to allow for the pardoning of some lies; to claim that some deceptive statements are not falsehoods...; and finally to claim that certain falsehoods do not *count as lies*” (36). ⁴⁵ Molière, *The Misanthrope*, I, i
While the idea that some lies are morally permissible seems to be at least as defensible as the idea that lies are always wrong, without some sort of equivocation the two ideas clearly cannot both be correct. The greater ability of the former, more nuanced position to be responsive to the complexities of human social life leads many to choose it as the more plausible position to take with regard to lying. To claim that lying is always wrong, regardless of circumstance, comes off as too simplistic to be feasible. But the idea that the absolutist view is more simplistic than the view that some lies are permissible depends on a particular way of framing the issue. When the question posed is whether lying is always wrong or not always wrong, the smart money would appear to be on the less extreme option. But we can pose the question somewhat differently: is lying an evaluative or a descriptive concept? To ignore the fact that claiming someone has lied is taken as an affront in a way that claiming they have spoken a falsehood is not—that is, to reason about lying as if it is a merely descriptive concept—is at least oversimplistic and arguably irresponsible. To reason about intentionally false assertions, declarations, or warranted statements (the problem of characterizing lies rears its head as soon as we begin down this path) and draw conclusions about lies is akin to reasoning about homicide and drawing conclusions about murder: some, perhaps many, of those conclusions will be true, but the reasoning is not valid. As we saw in Section 2, unless we divorce the concept from its use, we simply cannot divorce the notion of wrongfulness from the concept lie.

The question whether all lies are wrong or some lies are permissible appears to be another case of the modus ponens/modus tollens problem presented with regard to defining lies. One could argue that A is a lie, and A is not wrong, therefore lying is not
always wrong; or one could argue that lying is wrong, and \( A \) is not wrong, therefore \( A \) is not a lie. This aspect of the problem has been recognized by many who have grappled with the issue of lying. According to Carson, which side one takes is immaterial. While arguing, contra Kant, that it does make sense to speak of someone having a right to the truth, he states his position that a deceptive falsehood uttered to one who is undeserving of the truth is a permissible lie, but a lie nevertheless. He then says that he has “no substantive moral disagreement with” those who take such an utterance to be a non-lie: “We both agree that it is morally permissible to speak falsely to the man and mislead him. Our only disagreement is about whether we should count this example as a case of lying.”

Alasdair MacIntyre, in the 1994 Tanner Lectures, agrees with Carson that the difference is small, but disagrees with him regarding its importance. He argues that we should not underestimate the import of what appears to be a mere semantic difference:

Some adherents of these two contrasting and generally rival traditions may in fact disagree about very little of moral substance. For among some of those for whom lying is altogether prohibited, the definition of a lie is such as to exclude just those cases that some adherents of the other tradition treat as permissible or required lies. But it would be a mistake to conclude from these cases that the differences between the two traditions are unimportant…

The differences between the two views of lying are important, he says, because they represent competing views about just what is wrong with lying, leading to incompatible answers to three questions: how to define lie, what exactly the nature of the offense is in lying, and what sorts of arguments are to be advanced in favor of the answers to the first two questions. Those who advocate absolute condemnation, he points out, tend to view

\[46\] “The Definition of Lying” 288
lying as “an offense against truth,” while the others take a consequentialist perspective, seeing lying “as an offense against credibility and trust.”

Carson and MacIntyre both make important points. Most of us, I suspect, would agree that it is okay to deceive someone if such deception is the only way to prevent that person from bringing evil plans to fruition. But there are two ways to respond to this realization. The first option, which Carson takes, is to accept that some lies are morally permissible. It is the existence of so-called prosocial lies (which can include white lies) that lends credibility to this position. And it is precisely this sort of case that those who prefer the second option—defining cases of permissible linguistic deception as non-lies—must be ready to respond to. Augustine and Aquinas attempted to do this by admitting that some lies were less bad than others (but not that some lies were permissible). Kant has been subject to much criticism for apparently failing to make even this concession. In the next chapter I will provide a reading of Kant that shows the received understanding of him to be a caricature of his actual (quite subtle) view.

One concern with the second option that neither Carson nor MacIntyre points out is the question of whether the act of defining the permissible cases as non-lies can be anything other than *ad hoc*. Not all stipulative definitions are problematic, but if the only reason for defining “lie” in a particular way is to avoid admitting that some lies can be permissible, then we are back to asserting intuitions rather than reasoning about the issues. In the remainder of this section, I will argue that there are independent reasons for choosing the second option. Specifically, an absolute rejection of lying is more faithful

---

48 Mill and Kant 317
to the reality of how the concept *lie* is used; moreover, a failure to recognize this has negative repercussions in the realm of human interaction.

MacIntyre is correct that the difference between the two options is more significant than Carson allows, but not entirely for the reasons that he gives. In my view, the difference is important because the two positions reflect two different starting points for reasoning about lying—only one of which accurately reflects how the concept is used in everyday speech. As the framing of the *modus ponens/modus tollens* problem indicates, the claim that not all lies are wrong is not a premise pulled out of thin air, but rather the conclusion to an argument that begins with a particular starting point: the assumption that *lie* can be defined without regard for the moral element that is in fact central to the concept. As should be clear from Section 1, attempts to treat the definition of *lie* as conceptually prior to moral assessment are misbegotten; the moral element always sneaks back in under the guise of one or more “descriptive” characteristics. But regardless of this limitation, the attempt to define lie independently of its moral connotations inevitably leads to a notion that is more or less tied up with falsity—falsity of utterance, of belief, of representation, or of something else, but always falsity. But if falsity is taken to be the salient feature of lying, then once the (at least *prima facie*) moral blameworthiness of lying is recognized—if nothing else, because of the negative experience associated with having been the victim (notice the word) of one or more lies—it is only natural to draw a more than incidental connection between blameworthiness and falsity. Falsity thus comes to be understood as morally blameworthy—precisely the position that Alceste takes. In response to the realization (which Alceste apparently never comes to) that not all falsehoods are blameworthy, a
caveat is added: falsity can be okay, but requires additional justification in order to be so. But since falsity and lying are conceptually bound at this point, the same caveat is thought to be needed for lying. Lying thus comes to be understood as presumptively blameworthy, but justified in certain (as yet undefined) circumstances. Alceste’s (and by extension, on this reading, Kant’s) problem is thus understood as a failure to recognize that, since falsity is not always blameworthy, not all lies are blameworthy. But this is not the only way to assess Alceste’s mistake. I submit that Alceste’s mistake lies not in believing that all lies are wrong, but rather in believing that all falsehoods are lies—and, thus, that all falsehoods are wrong.

One consequence of the moralizing of falsity just described is that truth and falsity come to be differentiated on moral grounds. Once you identify falsity as a natural source of blameworthiness, it is easy to begin identifying truth as a natural source of praiseworthiness. Thus while falsity is taken to be suspect, always requiring additional justification, truth is placed under no such constraints. Quite the contrary, in fact: truth comes to be considered not just permissible, but laudable, extra considerations being required to justify withholding the truth. If considering an utterance to be a lie negatively affects moral evaluation of that utterance—as the experiments discussed above suggest—then this could have the effect of reducing our esteem of pro-social lies, simply by virtue of the fact that we call them lies. The experiment by Fu, et.al. provides some confirmation that this is in fact what happens. The concern is that calling pro-social falsehoods ‘lies’ could cause them to be construed primarily as falsehoods, and only secondarily (if at all) as prosocial behaviors. Such a construal could lead not only to the
condemnation of falsity, but also to an unrestrained demand for truth, even at the expense of social good (including, but not limited to, social harmony).

If falsity overrides social benefit in moral considerations, then truth overrides social harm in the same considerations. To the extent that it is considered wrong to tell a falsehood even to help somebody, it must be considered right to tell the truth even if it harms somebody. That is to say, truth comes to be taken as not just instrumentally but intrinsically good, and falsity as intrinsically evil. If falsity for the sake of social benefit is merely tolerated but not commended, then social harm for the sake of truth will be tolerated rather than condemned. But the idea that truth and falsity should either be defining characteristics or else override what would be competing considerations of what qualifies as praiseworthy or blameworthy behavior is precisely the idea that the portrayal of Alceste refutes. Such a view enables people to use truth as a weapon with which to do great harm, then as a shield behind which to hide from criticism: “I was only telling the truth!” Truth does not deserve a free pass, nor should falsity be always suspect. But this is precisely the treatment that is invited when the descriptive question regarding lying is taken to be prior to the moral question.

But if we begin instead with the moral question, which can be answered uncontroversially by the moral intuitions that are revealed through language use, the proscription against lying is just as universal as is the proscription against being a coward, a rapist, or a murderer: it is universal because that is the nature of the concept. I hope by now to have shown not only that this is how the concept is used, but also that this is how the concept ought to be used.
A recent example of how the concern I have expressed in this section plays out can be seen in Bernard Williams’ discussion of lying in his book *Truth and Truthfulness*. He acknowledges that there are times when the right thing to do is to deceive someone, even if such deception requires outright lying. He also says that, at least some of the time, it would be inappropriate to feel any sense of guilt or remorse regarding having told the lie. But he also allows for the possibility of cases wherein lying is justified but a sense of guilt or shame regarding the lie is appropriate. The example he uses is of an elderly woman whose child has experienced an unfortunate turn of events and who is not in a position to do anything about it. To tell her the truth, he stipulates, would cause nothing but pointless pain. He says that “The lie to the old lady may have been rightly told out of kindness, but we have reason to regret it, to feel bad about it, to think that something has been lost, because our relations to her were in general structured by a sense that we deserved the truth from each other.” ⁴⁹ Williams criticizes Adam Smith for thinking that such a sense of remorse would be appropriate for a gentleman who breaks a promise that was extorted out of him, stating that such a view must be “a hangover from an age before the modern world [Smith] did so much to inaugurate.” ⁵⁰ But Williams’s view seems to me no less parochial than Smith’s. The situation Williams describes is underdetermined, naturally, but if the relationship between myself and another is one wherein we deserve the truth from each other, then there are clearly other things that we deserve from each other as well. Included among these other things is compassion. What is lamentable in this story is the child’s misfortune, the woman’s inability to help,

⁴⁹ Williams 115
⁵⁰ Williams 116
(perhaps) her inability to cope well with the news; but given that these are the actual circumstances, the act of deception just amounts to an act of compassion. So, in this case, to lament the deception is to lament being compassionate. And there is no reason to lament that.

**Conclusion: Evaluation without Definitions**

At this point, an objection to my account demands articulation. Even granting that I am correct in identifying *lie* as a morally weighted concept, does it necessarily follow that the descriptive question cannot be answered independently of the evaluative question? Wouldn’t it be simpler to provide a definition of lying by detailing the descriptive characteristics and then adding, “and it must be considered bad by society”? Such a divide-and-define approach is intuitively quite plausible. After all, it seems obvious that one must know what one is evaluating before evaluating it. It makes little sense to say, “I don’t know what an apple is, but that’s a bad apple” or “I don’t know whether he is a lawyer, but I know that he is a bad lawyer.” Given this, if I have not yet defined “lie,” how can I possibly go about claiming that every lie is bad?51 This seems to be a guiding concern for Carson as well, leading him to take the approach he does. He says, “It is not helpful to assess the moral status of deception or any other class of actions without first clearly delineating that class of actions.”52

While it is true that one cannot evaluate an object without criteria on which to base an evaluation, and it is also true that such criteria are exceedingly difficult to come by without a descriptive definition as a starting point, this objection misses the point of

---

51 I thank Arindam Chakrabarti for the articulation of this objection.
52 *Lying and Deception* 49
what I have said regarding the concept *lie*. I am not making claims about individual *tokens* of lying, but rather about lies as an utterance *type*. It is precisely because not all apples or lawyers are bad that the descriptive definitions are necessary prior to the evaluation of any particular apple or lawyer. It is just as problematic with a lie as with any other object to say, “I don’t know if that’s an X, but I know that it’s a bad X.” But this is not the sort of claim I have made with regard to lying. It would be closer (though still not quite accurate) to attribute to me a statement of the form, “I don’t know what an X is, but I know that all X’s are bad.” That is to say, I have not made any claims about bad X’s—only about X’s; and that claim is: all X’s are bad.

To see why this new formulation is more plausible than the first one, consider Alanna, an anthropologist trying to learn about the concept *slorg* in an unfamiliar culture. Alanna observes that the terms surrounding the concept *slorg* are always used in evaluatively negative ways: slorging is derided, adults tell stories to their children about the dangers of slorging, and so on. Alanna could very easily conclude that slorging is a negative thing, something that she should avoid doing (or at least avoid doing openly) while among this culture, even without knowing what it is to slorg. Given her current state of knowledge, Alanna could quite plausibly file a preliminary report in which she says, “I have no idea what it means to slorg, but I know that it’s bad to slorg.” Granted, without knowing what it is to slorg, Alanna could never know whether she agrees with her host culture’s disdain of slorging, but she could still know that the culture takes slorging to be a bad thing. And she could also know that a question asking whether a particular act of slorging was good or bad would be inappropriate, and perhaps
incoherent, within that culture. To see just how inappropriate, imagine if Alanna were later to find out that “slorg” referred to what in English is called “rape.”

Because I am examining the concept *lie* rather than individual cases of lying, my analysis of lying is more akin to Alanna’s analysis of slorging than to analyses either of apples or lawyers. It makes sense to ask whether an apple, or a lawyer, or a professor, is good or bad, because the concepts are purely descriptive. But to ask of a rape whether it was a good or a bad rape reveals either extreme perversity or else a failure to understand the role the concept (and thus the term) plays in our language. And in this case what is true of *rape*, I have argued, is true of *lie*. Contra Carson, then, assessing the moral status of a class of actions can sometimes help with the task of delineation—as it does with lying.

Nevertheless, this objection does point to an important concern regarding my project, one that will remain with us throughout the next several chapters. The divide-and-define strategy that I reject is a sort of descriptivism, any form of which I argue cannot provide us with a comprehensive understanding of primarily evaluative concepts such as *lie*. But despite my focus on the evaluative question, the response must not be to advocate for a straightforward prescriptivism as a means of understanding the concept. Such an approach would be limited to describing lies as “Boo utterances” and would thus fail to differentiate between lies and other forms of morally blameworthy utterance, such as insults. The response must instead be a sort of quasi-prescriptivism. Other, descriptive elements of lying will need to be imported, and to the extent that this happens there will be a pull toward the descriptivism that I have rejected. An inquiry into lying, as with any primarily evaluative concept, thus involves a persistent tension between the
divergent pulls of descriptivism and prescriptivism. My goal is to maintain a balanced
approach between these two extremes as we examine just what, exactly, is wrong with
lying.
Chapter 2: The Sophistication of Kant’s View of Lying

Introduction

In Chapter 1, I argued that lying cannot be fully understood unless we acknowledge the priority of the moral question over the descriptive question, and that this prioritization shows an absolute prohibition of lying to be preferable to the more moderate account that considers some lies to be morally permissible. In this chapter we will consider perhaps the most familiar absolute condemnation of lying in Western Philosophy: that of Immanuel Kant. In Section 1 I will show that Kant has less trouble defending an absolute condemnation of lying than standard objectors think he does. Making use of two distinctions that he draws—between an utterance and a declaration, and between truth-telling and truthfulness—I will show that the class of utterances that Kant takes to qualify as lies is much smaller than it is ordinarily taken to be. As a result, he intends to prohibit less than he is normally taken to prohibit when he insists on an absolute avoidance of lying.

This is not to say that Kant’s account is free from difficulty. Consistent with his moral philosophy in general, Kant looks to explain the blameworthiness of lying by identifying a logical inconsistency in the act. One of his mistakes, I will argue, is an over-reliance on this approach: he attempts to elucidate what is wrong with lying by focusing exclusively on what he takes to be the intrinsic properties of the utterance rather than on the context in which the utterance is made. The context of an utterance gives rise to relevant communicational expectations of rights and duties, and is integral to understanding what makes a lie a lie. Kant’s explication of a lie as an “untruthful declaration” requires him to reject the Principle of Credulity, which he elsewhere
endorses. But why should we accept the Principle of Credulity instead of Kant’s account of lying? In Section 2 of this chapter I will show how the debates in the philosophy of language surrounding the nature of testimonial knowledge demonstrate the necessity of accepting the Principle of Credulity. I will then return in Section 3 to show that Kant’s account of untruthful declarations commits him to a violation of this principle.

2.1 Kant’s Condemnation of Lying

Given Kant’s view of morality as lawlike, no more subject to exceptions than the laws of physics are,¹ it may seem obvious that he would take the prohibition of lying to be exceptionless. But it is not as obvious as it may at first seem, since Kant also makes a distinction between perfect and imperfect duties, and imperfect duties do in fact admit of a certain amount of latitude in their fulfillment.² For instance, the duty to develop one’s talents is an imperfect duty because there is room for individual choice regarding which talents to develop and how far to develop them. And if at a given moment one happens not to be cultivating one’s talents, this does not necessarily constitute a violation of one’s duty. Kant explains that “ethics, because of the latitude it allows in its imperfect duties, unavoidably leads to questions that call upon judgment to decide how a maxim is to be

¹ Of course this is not to imply that Kant did not see the distinction between a natural law and a normative law. To say that a natural law is exceptionless is to say that it cannot be broken; to say that a normative law is exceptionless is to say that it may not be broken. Both are exceptionless in the sense that they apply universally. Such laws can be formulated, “An object will always…” (natural law) or “An agent must always…” (moral law). Each is exceptionless because it does not admit of “except in cases of” clauses.

² Kant uses different language at different points when discussing the categorization of duties. He sometimes refers to perfect and imperfect duties, sometimes to narrow (or strict) and wide duties, and sometimes to duties of right and duties of virtue. While his choice of terms varies, he appears to be consistent in the logic underlying his distinctions. In the *Groundwork* he identifies perfect duties with strict and narrow duties, and opposes them to wide duties (4:424). In the *Metaphysics of Morals* he equates wide duties with duties of virtue, as opposed to duties of right (6:390-395). I am adopting the perfect/imperfect characterization simply because it seems to me the most common among both Kant and Kantians.
applied in particular cases…”"³ To the extent that there is room for judgment regarding not just how but even when to fulfill an imperfect duty, imperfect duties might be thought sometimes to admit of exceptions. But Kant would insist that this is a mistake. The moral law can never give rise to duties that conflict with one another,⁴ even if it can and does give rise to duties that constrain one another’s application. He explains that “a wide duty is not to be taken as permission to make exceptions to the maxim of actions but only as permission to limit one maxim of duty by another (e.g., love of one’s neighbor in general by love of one’s parents)”⁵ So, in the interest of being faithful to Kant, the question whether the obligation not to lie admits of exceptions should be reformulated as the question whether Kant takes the duty not to lie to be a perfect or an imperfect duty.

The answer that Kant gives to this question is unambiguous: the duty not to lie is a perfect duty. The clearest illustration of this occurs in the Metaphysics of Morals. While issues of lying and truthfulness appear throughout the work, Kant places the section titled “On Lying” within his discussion of perfect duties to oneself. This identification of lying as a violation of a perfect duty is consistent with Kant’s treatment of the false promise in the Groundwork of the Metaphysics of Morals. He takes any false promise to be based on a non-universalizable maxim, and thus clearly in violation of the categorical imperative. But it is not obvious to everyone that Kant’s purpose with the false promise is to condemn lying, precisely. James Ellington, for instance, in his translation of the Groundwork, includes an editorial note to Kant’s analysis of the false promise at 4:422 which reveals that he (Ellington) takes it to be an issue of promise-

³ Kant, Metaphysics of Morals (hereafter MS) 6:411
⁴ See MS 6:224
⁵ MS 6:390
keeping rather than lying *per se.*\(^6\) But, at least if we take his ethics lectures to be representative of his views, Kant clearly does take the false promise to be a lie.\(^7\) In the lectures, Kant differentiates between a promise made with no intention to keep it (which he calls a lying promise) and a promise that is made sincerely, but is subsequently broken: the former, he says, is correctly understood as a lie, while the latter is a breach of faith but not a lie.\(^8\)

While Kant is consistent in treating lying as a violation of a perfect duty, there is nevertheless a noteworthy difference between his discussions of lying in the *Groundwork* and in the *Metaphysics of Morals.* In the *Groundwork* Kant clearly takes the false promise to be a violation of a perfect duty to others, whereas in the *Metaphysics of Morals* he treats lying as a violation of a perfect duty to oneself. This is not an insignificant shift. In a recent article, James Edwin Mahon has argued that Kant actually advocates two separate duties: one is a perfect duty to others not to lie to others, and the other is a perfect duty to oneself not to lie to oneself or to others.\(^9\) In Section 3 of this chapter I will argue that Kant’s formulations of both duties are problematic, because he attempts to identify the blameworthiness of lying by focusing on the properties of the utterance rather than on the relationship between the utterer and the audience. But to the extent that the question is whether lying could ever be morally permissible for Kant, the

---


\(^{7}\) Certainly what we find in the lectures—which are in fact notes taken by, and probably distributed among, students attending Kant’s classes—is less reliable as a reflection of Kant’s thought than are his own published writings. However, there is a consistency among the ideas as they are presented in the lecture notes, as well as between the lecture notes and Kant’s published writings, that supports the notion that the lecture notes are largely accurate reflections of Kant’s thinking at the times the notes were taken. For more information, see J. B. Schneewind’s Introduction to the *Lectures on Ethics.*

\(^{8}\) Kant, *Lectures on Ethics* (hereafter LE) 27:449

\(^{9}\) Mahon, “Kant and the Perfect Duty to Others Not to Lie”
important issue is whether the duty not to lie is perfect or imperfect, not whether it is a duty to oneself or to others.

Given that Kant takes lying to be a violation of a perfect duty, the remainder of the present section will examine what precisely he takes a categorical prohibition of lying to prohibit. Kant uses the false promise as a paradigm example of lying. But promises, like questions and exclamations, are not truth-apt. That is, a promise has no truth-value: it is not appropriately assessed as either true or false. This is important because the pre-reflective folk definition of “lie” as a false statement (as discussed in Chapter 1) has difficulty dealing with utterances that one might want to call lies, but that are not truth-apt. At the very least, a definition of lie that hinges on truth and falsity requires an addendum to account for infelicitous speech acts such as false promises. And since the false promise is a paradigm example of lying for Kant, it follows that his working definition of lie must differ from the pre-reflective one that people (at least in contemporary contexts) are likely to bring to their reading of Kant. If we respond to this realization by asking, “So what is Kant’s definition of ‘lie’?” we receive an answer that unabashedly puts the moral question first: lies “are strictly untruths that are contrary to duty.”\textsuperscript{10} For Kant, a lie is by definition a violation of one’s duty, and thus immoral. This is what enables him to say that there is no such thing as a white lie: “A white lie is often a contradictio in adjecto, like pretended tipsiness, it is untruth that breaches no obligation, and is thus properly no lie.”\textsuperscript{11}

\textsuperscript{10} LE 27:61, emphasis in original
\textsuperscript{11} LE 27:62, emphasis in original
From this it should be clear that, for Kant, the duty not to lie does not amount to a duty to tell the truth. If it did, then any uttering of an untruth would breach an obligation. This realization is important for understanding the definition Kant provides of lying in “On a Supposed Right to Lie,” where he says that a lie is “an intentionally untruthful declaration.” Because of the subtlety of the shift from “untruth” to “untruthful,” this definition can appear more descriptive and less prescriptive than that in his lectures. This shift is easily overlooked, but it indicates that the avoidance of lying is not simply a matter of telling the truth. This is a shift that Kant must make in order to hold the view, as he rightly does, that lying is a moral issue. That he defines lie as an untruthful declaration, rather than an untruthful utterance, is also significant. As we will see, it is not untruthfulness per se that Kant identifies as immoral, but only untruthfulness when making a declaration. One of the main purposes of this section is to clarify Kant’s use of “untruthful” and “declaration” as technical terms. Before doing that, though, we need to revisit an objection that was raised in the previous chapter.

In Chapter 1 I addressed the common-sense objection that an absolute condemnation of lying, such as Kant’s, is too extreme, failing to consider the nuanced nature of human life. Alceste, in Molière’s Le Misanthrope, is taken to represent a reductio ad absurdum of the absolutist position. I argued that the extremism objection rests on a particular way of framing the issue, and that a successful re-framing can shift the burden of proof onto those who would advocate a non-absolutist approach. I also argued that there are independent reasons for adopting the absolutist approach. At most, however, my argument only shows that an absolutist approach is preferable. It certainly

12 Kant, “On a Suposed Right to Lie from Philanthropy” (hereafter SRL) 8:426
does not establish the legitimacy of any particular absolutist approach, let alone Kant’s. The absurdity of Alceste’s position is no illusion: he is in error, just not the error that was initially supposed. As I said, Alceste’s error lies not in taking all lies to be immoral, but rather in taking all falsehoods to be lies. I argue that Kant does not commit Alceste’s error, because he recognizes that the duty to be truthful is not the same as the duty to tell the truth. It is essential to be clear on this distinction, because a correct understanding of Kant’s account hinges on it. An obligation to be truthful is an obligation not to assert what one does not believe. An obligation to tell the truth, on the other hand, is more like an obligation to be both perfectly informed and perfectly candid.

In the *Metaphysics of Morals* Kant does recognize a duty to be candid, but he also recognizes a duty to be reticent.\(^{13}\) While we have a moral duty to point out the faults of our friends in order to help them improve,\(^ {14}\) to do the same with regard to those who are not our friends Kant calls defamation: “The intentional spreading (propalatio) of something that detracts from another’s honor—even if it is not a matter of public justice, and even if what is said is true—diminishes respect for humanity as such… It is, therefore, a duty of virtue…to throw the veil of benevolence over their faults, not merely by softening our judgments but also by keeping these judgments to ourselves…”\(^ {15}\) Kant’s discussion of dissimulation in the *Lectures* shows that we have a duty to be reticent with regard to our own faults as well, and for similar reasons.\(^ {16}\) The duty to be candid and the duty to be reticent are clearly in tension with one another: to the extent that one is reticent, one is necessarily not candid; and to the extent that one is candid, one

\(^{13}\) MS 6:433, fn
\(^{14}\) MS 6:470
\(^{15}\) MS 6:466
\(^{16}\) LE 27:444-445
is necessarily not reticent.\(^\text{17}\) The endorsement of both duties amounts to an acknowledgment that each is subject to limitations—that is, that the duty to be candid and the duty to be reticent are both imperfect duties. But while the duty to be candid and the duty to be reticent are in tension with one another, neither is in tension with the duty not to lie, which for Kant is the duty not to make untruthful declarations.

### 2.1.1 Kant on Truth and Truthfulness

As I have indicated, the key to Kant’s understanding, not just of what a lie is, but also of why lying is wrong, is found in his understanding of “untruthful” and “declaration”. The difference between truth and truthfulness comes out most clearly in what Kant says about the latter in the *Doctrine of Virtue*: “Truthfulness [Wahrheitsliebe] in one’s declarations is also called honesty [Ehrlichkeit] and, if the declarations are promises, sincerity [Redlichkeit]; but, more generally, truthfulness is called rectitude [Aufrichtigkeit].”\(^\text{18}\) The fact that he elucidates truthfulness by reference to honesty, sincerity, and rectitude is noteworthy not just for what he includes, but also for what he leaves out. Conspicuously absent from Kant’s characterization of truthfulness is any mention of accuracy, of one’s utterance being true (\textit{wahr}). This is because the duty to be truthful must not be understood as a duty to tell the truth. If it were, then contingent matters of fact could override one’s intentions and turn a statement uttered in sincerity

---

\(^\text{17}\) One could attempt to dissolve the tension between candour and reticence by including a conditional element in one’s definition of candour, such that the duty to be candid is the duty to tell the truth even when unpleasant \textit{when one is speaking at all}. This way, an utterer could remain silent (and thus reticent) while simultaneously fulfilling the duty for candour. But this is to stipulatively define candour in a way that is in tension with common usage, and certainly with Kant’s treatment of the term. If my friend knows that my spouse is cheating on me but does not tell me, either because I “never asked” or because the matter “never came up,” my friend cannot on the normal use of the term be said to have been candid with me regarding this issue. In the same vein, Kant says that it is a duty to point out one’s friends’ flaws—simply avoiding the issue or remaining silent on the matter would constitute a violation of such a duty.

\(^\text{18}\) MS 6:429
into a lie simply by virtue of the speaker’s being deceived or otherwise misinformed. This consequence is not just counterintuitive, it is fundamentally at odds with Kant’s understanding of morality.

Since Kant takes truth to be independent of the will of the utterer, he maintains that we cannot have any duties with regard to speaking the truth. But while Kant is an externalist with regard to truth, he is an internalist with regard to truthfulness: “One cannot always stand by the truth of what one says to oneself or to another (for one can be mistaken); however, one can and must stand by the truthfulness of one’s declaration or confession, because one has immediate consciousness of this.”19 The moral issue thus concerns not actual falsity, which is an empirical matter, but rather intended falsity, which is a matter of will, as all moral issues must be for Kant. A speaker who is mistaken, and who sincerely expresses her mistaken belief, does not lie.20 The truthful speaker is one whose utterance accurately reflects her beliefs, regardless of whether those beliefs are themselves accurate. In fact, a speaker can be truthful even if her statements have no truth-value at all. Mahon makes the point that statements about future states of affairs and fictional hypotheticals—he uses the example, “Spiderman would beat the Incredible Hulk in a fight”—have no truth-value (unless one is a strict determinist, or a modal realist, respectively). “Every statement that a person makes, however, is either

---

19 On the Miscarriage of all Philosophical Trials in Theodicy 8:267; cited in James Edward Mahon, “Kant on Lies, Candour, and Reticence” (hereafter Candour) 106.
20 This is not to say that a mistaken speaker is necessarily blameless. She may be at fault for not knowing something that she ought to know, or for speaking as an authority when she is not one, or for violating a confidence. The point here is merely that, to the extent that the mistaken speaker is blameworthy, it is not because she has lied: she hasn’t.
truthful or untruthful. Every statement that a person makes has, as it were, a truthfulness-value.”

Mahon’s point helps to clarify the distinction Kant draws between truth and truthfulness, but when he goes on to define truthfulness as “the correspondence between a person’s statement and what she believes to be true” he misses part of the point of Kant’s distinction. Truthfulness for Kant is not just about beliefs: it is also about intentions. Statements are not the only type of utterance that can be assessed in terms of truthfulness. At the very least, promises qualify as well. I can tell you that Spiderman would beat the Incredible Hulk in a fight when I do not believe it, and I can promise to pay you back next week if you lend me fifty dollars today when I have no intention of doing so. The first involves a failure of my words to reflect my beliefs, while the second involves a failure of my words to reflect my intentions. But for both cases, we could say that my utterance fails to reflect my thoughts while purporting to do so—and thus both utterances are untruthful, regardless of whether they are truth-apt.

---

21 Candour 105
22 Candour 105
23 While Kant pre-dates speech-act theory, it seems apparent from his writings that he did not take promises to be truth-apt. If we interpret a promise as a statement rather than a performative utterance, the question of truthfulness becomes misplaced. For then, truthfulness in an utterance of the form, “I promise…” would consist in the speaker’s taking herself to be promising, not in the speaker’s intention to fulfill the promise made. This is clearly at odds with Kant’s treatment of the false promise in the Groundwork and in the Metaphysics of Morals. Thus it seems reasonable to conclude that Kant would have been amenable to something resembling a performative account of promising, and would have denied that promises are truth-apt.
24 One might respond here on Mahon’s behalf that this refinement is unnecessary, because if I promise that I will pay you back next week when I don’t believe that I will pay you back next week, my words do in fact fail to reflect my beliefs. But as long as I take myself to be promising in uttering the words, “I promise,” then my words do accurately reflect my beliefs, regardless of whether I intend to fulfill the promise or not. There is a subtle grammatical difference between promising that something and promising to do something, but I see no compelling reason to believe the difference in surface grammar here reflects a difference in the actual speech act that is performed in promising.
An understanding of the distinction between truth and truthfulness is essential to understanding what Kant says in his essay “On a Supposed Right to Lie From Philanthropy.” In this essay Kant notoriously refuses to allow for the permissibility of lying even to prevent a would-be murderer from learning the whereabouts of one’s friend, the prospective victim. Kant’s willingness to accept this consequence of the absolutist position is what leads many to believe that he does commit Alceste’s error. But Kant’s position is much more nuanced than this objection gives him credit for. The essay, published in 1797, is primarily a response to an objection to Kant’s ethics voiced by Benjamin Constant—an objection that rings true with many even today. If a person has no right to a particular truth, Constant argues, it follows that nobody can have an obligation to share this particular truth with that person. “To tell the truth is therefore a duty,” Constant concludes, “but only to one who has a right to the truth. But no one has a right to a truth that harms others.”\(^{25}\) Kant responds to Constant by saying that the idea of having a right to the truth is incoherent. Since (objective) truth, for Kant, is independent of people, it makes no sense to talk of having a right to the truth. His point is that having a right to truth means having a right that what one says be true—that for any assertion I make, I have a right for that assertion to come out true. But this would give rise to the absurd consequence that if I were to say, for instance, “I am officially recognized as the President of the United States,” then I would have a right to be officially recognized as the President of the United States.\(^{26}\)

\(^{25}\) SRL 8:425

\(^{26}\) I recognize that, in providing this argument, Kant seems to be missing the point of Constant’s objection. At this point I only intend to discuss Kant’s argument; for a discussion of its failure as a response see below, page 93.
While the notion of a right to truth is senseless, Kant acknowledges that we do have a right to truthfulness (of a sort): “One must instead say one has a right to his own truthfulness (veracitas), that is, to the subjective truth in his person.”

While it may sound odd to speak of having a right to one’s own truthfulness, such a notion makes a good deal of sense within the Kantian framework. If I have a perfect duty to myself to be truthful, a corollary of this is that I have a right to my own truthfulness. As a rational being endowed with the capacity to communicate my thoughts via language, in using language to communicate the contrary of my thoughts I treat myself qua object (homo phenomenon) as a mere means—a “speaking machine,” in Kant’s words—for bringing about the ends that I hold qua subject (homo noumenon).

Regardless of either intended or actual consequences, in lying one treats oneself as a mere means, and thus “violates the dignity of humanity in his own person.” Kant goes on:

By a lie a human being throws away and, as it were, annihilates his dignity as a human being. A human being who does not himself believe what he tells another…has even less worth than if he were a mere thing; for a thing, because it is something real and given, has the property of being serviceable so that another can put it to some use.

This is why it makes sense to speak of having a right to one’s own truthfulness: lying is a violation of one’s duty to oneself, and thus a violation of one’s own right to be treated as an end in oneself. Kant says that “To be truthful (honest) in all declarations is therefore a sacred command of reason prescribing unconditionally, one not to be restricted by any conveniences.”

---

27 SRL 8:426
28 See MS 6:430
29 MS 6:429
30 SRL 8:427
It must be remembered, though, that even the requirement to be truthful in all declarations is not a prohibition against ever saying what one does not believe.

According to Kant it is okay for one’s utterances not to match one’s thoughts, but it is not okay for an utterance to misrepresent one’s thoughts. The violation of one’s duty to oneself lies not in merely being untruthful, which is simply a matter of one’s words failing to reflect one’s thoughts, but rather in using such unreflective words explicitly to mislead others with regard to one’s thoughts. Jokes and social niceties that are untrue are thus not lies because, while untruthful, they are not meant to be taken as reflecting one’s thoughts; nobody is meant to believe them. But the permissibility of untruthfulness is not limited to such seeming trivialities. Kant even allows that the use of language for the express purpose of deception can be permissible: “I can make a pretence, and give expression to something, from which the other may deduce what I want him to; but he has no right to infer from my utterance a declaration of intent, and in that case I have told him no lie, for I never declared that I was opening my mind to him…” Kant illustrates this point by saying that one does not lie in misinforming a mugger regarding the whereabouts of one’s wallet, because the mugger “has no right whatever to demand the truth from” his victim. But he goes on to say that, if the victim announces an intention to speak truthfully, misinforming the mugger does count as a lie. This wrinkle clarifies the force of Kant’s response to Constant in “On a Supposed Right to Lie”: while the would-be murderer has no right to expect or demand truthfulness from me, if I tell him that I am going to speak my mind and yet proceed to mislead him, I lie to him and thus

---

32 LE 27:446-447, emphasis added
33 LE 27:447
violate my own right to truthfulness in myself. I am not, though, *obliged* to tell the
would-be murderer *anything*, unless I take that obligation upon myself. Even if someone
has no right to demand or expect truthfulness from me, if I offer to speak truthfully I am
then bound to do so. So “Not every untruth is a lie; it is so only if there is an express
declaration of my willingness to inform the other of my thought.”34

2.1.2 Kant on Declarations

The idea of informing another of one’s thoughts is what differentiates declarations
from other utterances for Kant. This is the second distinction that is essential for
understanding Kant’s condemnation of lying. Mahon uses the notion of *inviting* to
elucidate the difference between declarations and other utterances: “A person makes a
declaration when she makes a statement, and invites someone to believe that statement to
be true.”35 This elucidation is helpful, but Mahon’s focus on statements is overly narrow
here, just as it was in his discussion of truthfulness. He uses ‘statement’ to mean,
roughly, a truth-apt utterance, and accordingly seems to equate ‘declaration’ with
‘assertion’. But this cannot be what Kant means by ‘declaration,’ since he explicitly
identifies promises as a type of declaration.36 For Kant a lie—and thus a declaration—
need *not* be a statement; it need only be an utterance that purports to reflect one’s
thoughts, and in so doing invites someone to take said utterance as truthful.

34 LE 27:448. It is worth noting that the discussion of the mugger hinges on truth and untruth, rather than
truthfulness and untruthfulness. This discrepancy can, I suspect, be chalked up to the fact that the mugger
example comes from the 1784 lectures—the time when Kant was working on the *Groundwork*, and fully
thirteen years before he published “On a Supposed Right to Lie” and the *Metaphysics of Morals*, in which
the importance of truthfulness as distinct from truth reaches its full development in Kant. Reading
“truthfulness” in for “truth” does not, it seems to me, alter the sense of what Kant seeks to accomplish with
the example.
35 Candour 102
36 MS 6:429
This discussion of declarations recalls to mind the consideration of warranting, of wanting to be on record as having said something, that Carson brings up in his discussion of lying. Kant only discusses cases where the liar actually intends for the lie to be believed, but there is room in his account for Carson’s fearful witness or the bald-faced liar. You can invite someone to believe something even if you have good reason to believe the invitation will be declined, in much the same way that you can invite reclusive relatives to a family gathering without the expectation that they will show up. The bald-faced liar and the fearful witness can both be understood as making invitations that they expect or hope will be refused, and thus as making declarations.

While inviting and warranting are both useful heuristics for understanding what Kant has in mind when he discusses declarations, a closer parallel can be found with the notion of avowal. A declaration, as Kant uses the term, is an utterance that claims a commitment to a belief, an intention, or a course of action. Making a declaration is in this way similar to swearing an oath or taking a vow. A declaration is an utterance that avows truthfulness. To make an untruthful declaration, then, is self-contradictory (and thus non-universalizable) because it involves avowing that one’s utterance contains precisely that which it does not contain: it amounts to saying, “I believe what I do not believe,” or “I intend what I do not intend.” Given this understanding, it becomes apparent just how appropriate it is that Kant use the false promise to illustrate what is wrong with lying: every lie can be understood as a false promise. Kant sees an implicit promise within any declaration, and this promise is that one will communicate one’s

---

37 I owe the use of this term to Arindam Chakrabarti, who first suggested it to me.
actual thoughts. A maxim to speak in such a way as to misrepresent one’s thoughts fails the universalizability test in precisely the way the false promise does in the *Groundwork*.

Harry Frankfurt has made the point that lying always involves a twofold deception—about the world and about one’s beliefs:

Thus someone who lies about how much money he has in his pocket both gives an account of the amount of money in his pocket and conveys that he believes this account. If the lie works, then its victim is twice deceived, having one false belief about what is in the liar’s pocket and another false belief about what is in the liar’s mind.38

To Kant, the relevant consideration in lying is the second of these deceptions. Though it frequently does, lying need not involve any misrepresentation of the external world. However, lying necessarily involves misrepresenting one’s internal world—or at least that part of one’s internal world that includes one’s thoughts. And since, according to Kant, others’ access to one’s internal world is always mediated by what one chooses to communicate,39 people’s ability to engage effectively with one another hinges on the truthfulness of our declarations.

This discussion has shown that Kant restricts the scope of his universal condemnation of lying by drawing a distinction between declarations and other utterances, and insisting that only untruthful declarations count as lies. In this way he can escape the charge of extremism while remaining an absolutist. Anything that is not a declaration, regardless of untruth, untruthfulness, or even intent to deceive, is not a lie. The fact that lying always involves the very specific sort of misrepresentation that he attributes to it provides insight into precisely what Kant thinks is wrong with lying:

---

38 Frankfurt, *On Bullshit*, 13
39 There are reasons to doubt the accuracy of this claim. But, as we shall see, Kant takes language to be a necessary condition for the communication of (at least some of) one’s thoughts.
“Every lie is objectionable and deserving of contempt, for once we declare that we are
telling the other our thoughts, and fail to do it, we have broken the pactum, and acted
contrary to the right of humanity.”

At this point, I hope to have made clear Kant’s understanding of truthfulness and
to have provided a cursory description of what he means by “declaration.” In order to
clarify his understanding of declarations further, however, it will be necessary first to
address some epistemological issues surrounding the philosophy of language.
Specifically, the issue is what must be the case in order for declarations to be a possible
source of knowledge. In the next section, I will argue that a Principle of Credulity such
as that articulated by Thomas Reid must hold in order for declarations to do the work that
Kant needs them to do. In Section 3 I will show how this leads to a problem for Kant’s
account.

2.2 The Principle of Credulity

It is a commonplace, acknowledged both by philosophers and as a matter of
common sense, that people can and often do gain knowledge as a result of others’
utterances. Hume, one of the few in the Western canon to discuss the matter at some
length, \(^{41}\) acknowledges that “[T]here is no species of reasoning more common, more
useful, and even necessary to human life, than that which is derived from the testimony

---

\(^{40}\) LE 27:448

\(^{41}\) In his insightful examination of testimony, C.A.J. Coady comments on Western epistemology’s failure to
deal sufficiently with the issues surrounding testimony: “either it has ignored testimony altogether or it has
been cursory and dismissive” (6). Arindam Chakrabarti begins the preface to Knowing From Words with a
similar observation: “Although knowledge and language have both been at the centre of the twentieth
century’s philosophical preoccupation, neither epistemologists nor philosophers of language, on an
average, have taken more than a side-long look at the pervasive use of language in transmitting, sharing and
preserving knowledge” (vii).
of men, and the reports of eye-witnesses and spectators.”

Kant acknowledges as much as well, if in a more understated manner: “[A]part from speculation, I have only two ways of enlarging my store of information: by experience, and by testimony. But now since I cannot experience everything myself, if the reports of others were to be false tidings, the desire for knowledge could not be satisfied.”

Gaṅgeśa, the philosopher who gave rise to the Navya Nyāya school in India, points out that even attempting to deny that an utterance can be knowledge-generating embroils one in contradiction, at least if the denial takes the form of an utterance. If the utterance is correct, then it will be impotent, and thus pointless to make; on the other hand, if the audience comes to know, as a result of this denial, that words cannot generate knowledge, then the denial is false—at least on this occasion an utterance was knowledge-generating. Of course, the contradiction can be avoided if the denial is taken to merely raise a concern, the legitimacy of which is subsequently verified by means of perception or inference. Gaṅgeśa recognizes this and provides more sophisticated arguments, the details of which need not concern us here, for the knowledge-generating power of utterances.

While it is uncontroversial that language use can give rise to knowledge, exactly how it does so is a matter of some dispute. The crux of the dispute regards whether such language-based knowing, or testimonial knowledge, is a unique mode of knowledge or is rather reducible to some other means of knowing, such as perception or inference. Let us call the former the Irreducibility Thesis and the latter the Reducibility Thesis. A version of the Reducibility Thesis, one holding that knowing from words involves some form of

---

42 Hume, An Enquiry Concerning Human Understanding, Section X, part 1
43 LE 27:448
inference, is the dominant (if generally implicit) view within the Western philosophical tradition. The Irreducibility Thesis has fewer advocates in the Western tradition, but is relatively more common among Indian philosophical systems. The purpose of this discussion is not to take sides on the reducibility-irreducibility issue, but rather to show that a version of the Principle of Credulity is instrumental to both sides. To state it more clearly: my thesis is that a default position of credulity with regard to utterances is presupposed by any claim that language can be epistemically fruitful, whether testimonial knowledge is taken to be reducible or irreducible. This will become clear with a thorough examination of the reducibility-irreducibility debate. But first, we must first clarify exactly what the Principle of Credulity involves.

2.2.1 Credulity and the Skeptic’s Regress

Thomas Reid articulates two principles that he takes to be essential not just for language to be informative, but for language to exist at all.\textsuperscript{44} The first is the Principle of Veracity, which he describes as “a propensity to speak truth, and to use the signs of language, so as to convey our real sentiments.” He explains that people are naturally disposed to say what they believe, and that they say the contrary of it only when under an additional influence. The second principle is the Principle of Credulity, which amounts to a tendency to believe what one is told. This latter principle, he says, “is unlimited in children, until they meet with instances of deceit and falsehood: and it retains a very considerable degree of strength through life.” Bernard Williams aptly sums up the Principle of Veracity when he says, “[I]n the first instance and in the simplest cases, we

\textsuperscript{44} Reid, \textit{An Inquiry into the Human Mind}, Chapter 6, Section XXIV (All citations from Reid are taken from this Section.)
are disposed spontaneously to come out with what we believe.” The Principle of Credulity can be similarly summarized: “In the first instance and in the simplest cases, we are disposed spontaneously to believe what others come out with.”

That people do in fact tend to believe what they are told is both a consequence of this principle and a testament to its presence. It is worth noting that the Principle of Credulity is not a claim about knowledge, but rather about belief. On its own, the Principle of Credulity takes no stand on the issue of whether testimonial beliefs can constitute knowledge, let alone whether any such knowledge would be reducible or sui generis.

Reid explains that the Principle of Credulity is essential in order for language to be informative because Hume’s problem of induction applies to any attempt to infer a speaker’s present meaning or sincerity from previous experiences:

Have they promised that they will never impose upon us by equivocation or falsehood? No, they have not. And, if they had, this would not solve the difficulty; for such promise must be expressed by words, or by other signs; and before we can rely upon it, we must be assured, that they put the usual meaning upon the signs which express that promise. No man of common sense ever thought of taking a man’s own word for his honesty; and it is evident that we take his veracity for granted, when we lay any stress upon his word or promise.

Reid here is pointing toward what I call the skeptic’s regress. To assume that a particular speaker (or group of speakers) is untrustworthy is to foreclose the possibility of useful linguistic exchange with the objects of that assumption. If I assume that all Cretans are liars, then there is nothing any Cretan can do to convince me otherwise.46

45 Williams 75
46 It is not just that no Cretan can say anything to make me believe him. As we will see below, even non-linguistic behavior will fail to quell my skepticism.
To see the power of the skeptic’s regress, consider the case of Maggie and Molly. Maggie is trying to gain information about Molly’s secret hideout and has resorted to torture. Maggie is very skilled at inflicting pain on her victims, and she knows that a victim of torture is likely to say anything necessary just to escape the suffering. Knowing this, she is skeptical of anything Molly says, suspecting that it is just a ploy to get the pain to stop. So if Molly says that the hideout is at 221B Baker Street, Maggie will think Molly is lying. And if Molly says, “I swear, that’s where it is. I’m telling you the truth,” this will do nothing to mollify Maggie’s suspicion. Maggie and Molly are both trapped: as long as Maggie thinks Molly is a liar, Molly can do nothing to convince Maggie that she is telling the truth. To assume that one’s interlocutor is untrustworthy is to poison the well of discourse. An assumption of untrustworthiness undermines the ability of language to be informative because if I assume that you are disingenuous, then there is nothing you can say or do to convince me otherwise. This is the skeptic’s regress. If I am consistent in my doubting, then your most sincere protestations of honesty and truthfulness will meet with nothing more than further doubt—you cannot give me anything that I can recognize as a reason to believe you.

The example of Maggie and Molly illustrates the thrust of the skeptic’s regress, but it has a crucial limitation. While the discussion so far has focused on particular conversations between individuals, the Principle of Credulity operates on a more theoretical level. There is nothing intrinsically problematic with disbelieving the utterances of a particular individual. To assume that someone is untruthful may be to poison a particular well, but we really ought not to drink from some wells. Disbelieving a person is, in some cases, a matter of prudence. Maggie is probably correct to disbelieve
whatever testimony she extracts from Molly by means of torture. Further, though disbelieving an entire group of people, or an individual on the basis of group membership, amounts to prejudice, it does not eliminate the possibility of communication per se. One violates the Principle of Credulity when one assumes as a matter of course that language use is untrustworthy, that truthfulness must be independently established in every case. The skeptic’s regress threatens here because there is nothing anyone can say or do to overcome the doubt. Even checking utterances against the world, as a sort of verification, will not help. Verifying that an utterance is factually accurate will do little to alleviate skepticism, because the truth of a particular utterance or set of utterances can always be explained away: the speaker could be deceptive but misinformed, or could be telling the truth merely to foster the trust necessary to pull off a more significant long-term deception. Once doubt has taken hold, it is notoriously difficult to exterminate. Any desire to know on the part of the skeptical hearer will remain unsatisfied after hearing the words of a distrusted utterer: one will always feel the need to verify the utterer’s claims independently. In such a context, language becomes utterly impotent as a means of knowledge.

Reid also makes the insightful point that the skeptic’s regress cannot be avoided simply by declaring oneself an agnostic rather than a skeptic with regard to utterances. The only recourse, he argues, is actual credulity:

[I]f nature had left the mind of the hearer in equilibrio, without any inclination to the side of belief more than to that of disbelief, we should take no man’s word until we had positive evidence that he spoke truth. His testimony would, in this case, have no more authority than his dreams; which may be true or false, but no man is disposed to believe them, on this account, that they were dreamed. It is evident, that, in the matter of testimony, the balance of human judgment is by nature inclined to the side
of belief; and turns to that side of itself, when there is nothing put into the opposite scale. If it was not so, no proposition that is uttered in discourse would be believed, until it was examined and tried by reason; and most men would be unable to find reasons for believing the thousandth part of what is told them. Such distrust and incredulity would deprive us of the greatest benefits of society, and place us in a worse condition than that of savages.47

In fact, as Reid himself notes, the problems with a pervasive skepticism such as we are discussing here go beyond the impotence of language to the impossibility of language. As Chakrabarti indicates,48 an argument can be made following Donald Davidson that a person who fails to take as true much of what is said would be unable to interpret, and thus understand, any use of language at all. To such a person, what we take as utterances would be meaningless sounds, marks, or physical movements. A community of such individuals would have no language at all—and probably would not be much of a community either.

The Principle of Credulity is thus a necessary condition for the possibility of generating beliefs, and thus knowledge, as a result of an utterance. But not all who discuss the epistemology of testimony readily acknowledge this necessity. One objection that is regularly voiced against the Principle of Credulity is that it amounts to a defense of gullibility. This objection comes primarily from those who advance the Reducibility Thesis. To understand why they take credulity to give rise to gullibility, and how to respond to their concerns, we will need to examine the reducibility-irreducibility debate in some detail.

47 The reference to savages here can lead to a misunderstanding of Reid’s point. His intent is not to establish that “savages” are untrustworthy, and so worse than “us” normal humans, but rather to point out that if humans had no convention for trusting each other they would be unable to learn from one another, and would thus be as epistemically impoverished as non-human animals are typically taken to be.

48 Chakrabarti, “Knowledge from Trusted Tellings and its Preventers” (hereafter Trusted Tellings) 33
2.2.2 The Reducibility-Irreducibility Debate

The Reducibility Thesis holds that knowledge from words, when it can be attained at all, can be explained in terms of one or more other means of knowing. By far the most common version of the Reducibility Thesis attempts to reduce testimonial knowledge to a type of inference. Such a view can seem intuitively plausible. Since there is no necessary connection between S’s saying that P and the P’s being true (S may be joking, misinformed, deceptive, or otherwise insincere), it seems the only way that H can come to know that P from S’s saying that P is with the help of one or more mediating premises, such as that S is knowledgeable and sincere. In short, common sense seems to hold that testimonial belief (and thus even the possibility of knowledge) involves two steps (at least): the audience first understands what the utterer means, and then, on the basis of an assessment of the utterer as trustworthy, believes the content of the utterance. But if a chain of reasoning is necessary in order to conclude the truth of some claim about the world from someone’s making the claim, especially if that chain of reasoning includes a detour through the utterer’s thoughts, beliefs, or intentions, then many potential pitfalls present themselves in the move from utterance to knowledge. This is the sort of concern that guides the skepticism with which Descartes opens his *Meditations*. It has also informed much of twentieth-century Anglo-American philosophy of language, leading to a rather poor regard for testimony as an independent means of knowledge.

As intuitively appealing as such a view may initially seem, it leads to consequences that are more than a little counterintuitive. Consider, for instance, the view advanced by Locke: “words, in their primary or immediate signification, stand for
nothing but the ideas in the mind of him that uses them…. [A]nd the end of speech is, that those sounds, as marks, may make known his ideas to the hearer.” Chakrabarti and Matilal point out an absurdity of the Lockean view, saying that, “upon that theory, all our conversation about politics, travel or sports should be construed like our confessions to the psychotherapist.” On the view here being attributed to Locke, any assertion that P, while perhaps apparently about the world, contains an implicit “I believe” that renders the assertion actually about the speaker’s mental state. But this is in tension with what we typically take ourselves to be doing when using language. If I believe that Omaha is the capital city of Nebraska, and express this belief by saying “Omaha is the capital city of Nebraska,” it is appropriate for someone to tell me that I am mistaken. But if my utterance is taken to be a statement about my beliefs, rather than about the capital city of Nebraska, it would be inappropriate to tell me that I am mistaken, for I am not mistaken about my belief that Omaha is the capital city of Nebraska. It is appropriate to correct me because my utterance is an expression of my belief, not a report of it.

There is a difference, in intent and in subject matter, between telling someone, “Your car is on fire,” and telling someone, “I believe your car is on fire.” The fact that “I believe,” when appended to a sentence, softens or reduces one’s commitment to the truth of the main clause is significant. If I say to you, “Your car is on fire,” and you respond as if my utterance is about my beliefs rather than about the world, you will likely not act with the urgency that the facts demand. In such a case, I am likely to respond by saying,

---

49 Locke, Book III, chapter 2, paragraph 2; emphasis in original
50 Chakrabarti and Matilal, “Introduction” 4
“You don’t understand. I don’t care if you think I believe that your car is on fire, so long as you believe it, because your car is on fire.”

It must be noted that advocates of the Reducibility Thesis are not constrained to the Euro-American tradition. The Cārvākas held a similarly skeptical view in classical India: “According to the Cārvākas,” Satishchandra Chatterjee writes,

[T]here is no logical ground or justification for our believing in anything simply on the statement of another person. If it were so, we shall have to believe in many absurd and fictitious objects about which any fool may tell us. If, however, śabda or testimony be constituted by the statement of a trustworthy person, it is only a case of inference from the character of a man to the truth of his assertion.

Dignāga, and those who follow him in the Buddhist tradition, also hold that knowledge from words (śabda in Sanskrit) involves inference, though the Buddhist arguments differ markedly from those in the West and of the Cārvākas—and are less subject to the sorts of concerns just expressed regarding the Lockean view. According to Dharmakīrti, for instance, the object of perception (pratyakṣa) is the bare particular (svalakṣaṇa), while the object of inference (anumāna) is the concept (sāmānyalakṣaṇa). Roughly speaking, my coming to know that there is a fire on the hill by means of perception involves a cognition of the actual fire, while coming to know the fire by means of inference (say, from seeing smoke emanating from a particular part of the hill) involves cognition of the concept fire, but not of the actual fire that is on the hill. Similarly, knowledge that there

---

51 Assuming, of course, that I am sincere in my utterance. But note that if I am trying to play a prank on you, perhaps because it is April Fool’s Day, your failure to take my utterance as about the world will lead to frustration on my part. Which is to say, attempts at linguistic deception are (typically) predicated on the fact that we usually take utterances to be about the world rather than about our mental states. We can use words to deceive people about the world only because we typically use words to tell people about the world.

52 Chatterjee 319

53 Actually, it involves direct perceptual experience of svalakṣaṇas that give rise to cognitions of properties such as heat and bright, and which we subsequently cognize as fire. But for the purposes of the discussion here, it suffices to talk of direct perceptual experience of fire.
is fire on the hill from someone’s utterance involves cognition of the concept rather than the particular. Dharmakīrti also holds that each means of cognition has a unique type of object, similar to the way in which each sense modality has a unique type of object of cognition—color for sight, sound for hearing, etc. Thus testimony, which cognizes the same type of object that inference does (the sāmānyalakṣaṇa), must be a form of inference. The upshot of this approach is that it is possible to hold that testimonial knowledge is inferential in nature without taking apparent statements about the world to be disguised statements about beliefs. But the relative merits of the Buddhist and Lockean understandings of testimonial inference need not be further discussed here, as they are beyond the scope of our present concerns.

In contrast to the Reducibility Thesis, the Irreducibility Thesis has few advocates in the Western tradition. Thomas Reid is the most prominent; more recently, C. A. J. Coady has advocated for a version of the Irreducibility Thesis. By far the most fully articulated version of the Irreducibility Thesis, though, is that of the Nyāya school of Indian philosophy. According to Nyāya, understanding what is said typically involves direct, unmediated assent to, or belief in, what is said—and such belief at least sometimes constitutes knowledge. In contrast to the Lockean view described above, the Nyāya view treats utterances as signifying objects primarily. As Chakrabarti says, “Nyāya regards the distrustless reception of information as the primary cognitive attitude to meaningful, nonfictional, sincere nonfigurative speech.” This means that, according to Nyāya, we

---

54 The following discussion of the Nyāya account of śabda is based on Jonardon Ganeri’s treatment of the subject in Artha: Meaning, primarily Chapters 4 and 5 (hereafter Artha), and Arindam Chakrabarti’s discussions in “Telling as Letting Know” and “Knowledge from Trusted Tellings and its Preventers”.

55 Trusted Tellings 2
use language first and foremost to communicate about the world, and only by extension to communicate about our ideas:

Thus, understanding an utterance of ‘Rāma is cooking rice’, made sincerely by a competent speaker, consists in the hearer’s direct assent to the proposition that Rāma is cooking rice. Understanding issues sometimes in a belief not about what the speaker said but in what was said.\(^{56}\)

And this unmediated belief in what was said, when what was said is true and comes from a sincere and competent speaker, qualifies as knowledge.

Of course Naiyāyikas are aware that utterances are not always true. They know as well as anyone that understanding does not always entail belief, let alone knowledge. If H has reason to believe that S is incompetent or insincere, or if what S says is incredible to H, then H will withhold assent. This caveat risks making the Nyāya account sound very much as if it involves the same inference from the trustworthiness of the speaker that was discussed as characteristic of the Reducibility Thesis. The similarity is both undeniable and informative, but must not be taken to indicate more than it does.

Nyāya identifies certain defeating conditions that prevent knowledge from words, whereas Lockeans identify certain necessary conditions that allow knowledge from words. That the defeating conditions that the Nyāya tradition identifies as preventing knowledge are similar to the conditions that advocates of the Reducibility Thesis identify as necessary in order for the inference to go through shows that the two sides are aware of (many of) the same potential hazards. What is important, though, is the difference in response they provide to the hazards. According to the Nyāya account, effort is involved in withholding (or withdrawing) assent rather than in granting it. The difference may be

\(^{56}\textit{Artha} 74\)
subtle, but the consequences are substantial. To see this, imagine you are driving a car, and your goal is to get through the intersection ahead. According to the Lockean account, you need to put your foot down on the accelerator in order to achieve your goal, for the car will not clear the intersection on its own. On the Nyāya view, on the other hand, the car is already moving and you need only refrain from engaging the brake to achieve your goal.

The Reducibility Thesis holds that one’s interlocutor must not only be trustworthy, but must be known to be trustworthy, in order for testimony-based knowledge to be possible. Thus, if and when trustworthiness cannot be independently established, knowledge cannot be obtained (the car does not go anywhere); and when knowledge can be obtained, it is based on an inference from the trustworthiness of the utterer. The Irreducibility Thesis, by contrast, holds that not being aware that one’s interlocutor is untrustworthy is sufficient to generate belief in what is said, and that the reliability of one’s interlocutor can mean that such a testimony-generated belief constitutes knowledge (the car goes through without any effort). It is not necessary that one trust the utterer; only that one not distrust. “Appeal to the general or occasion-specific reliability, sincerity, seriousness, and so on of the speaker or any coherence criteria or pragmatic success comes at a subsequent stage when, according to Nyāya, you are not justifying or giving evidence for what you know but justifying or giving evidence for the claim that you know it.”57

It is important to be clear on the difference between a reliance on trust and on a lack of distrust, because much rides on this seemingly trivial distinction. I mentioned

57 Trusted Tellings 8-9
above that Coady has recently advanced a version of the Irreducibility Thesis. His understanding has much in common with the Nyāya view described here, but there is one difference that is worthy of notice. In his analysis, Coady draws a close connection between believing an utterance and trusting the utterer. “When we believe testimony,” he says, “we believe what is said because we trust the witness.” On this point Coady is, I think, demonstrably wrong. Certainly there is a correlation between trust and belief, but Coady is not justified in positing the causal relationship that he does. It is possible to believe an utterance without actively trusting the utterer, for two reasons. First, for some utterances the issue of trusting or distrusting the utterer does not even arise. When reading an instruction manual, following a recipe, or obeying a computer prompt, for instance, one frequently follows the command without considering the source. As Chakrabarti says,

> When I read *commands* (in a manual or on a computer screen), I know *that* I am supposed to remove the lid or press the ‘start’ key. Quite straightforwardly I (the addressee) am the subject or qualificand and ‘being supposed to do this or that’ is the predicate or qualifier. Thus the awareness generated by the imperatives doesn’t have to bring in the speaker (there might be none) into the content.

Chakrabarti makes his point in arguing for the Irreducibility Thesis, on the grounds that correctly understanding an utterance need not involve any reference to, or consideration of, the utterer’s communication-intention. But it also helps make the point that one need not trust the source of an utterance in order to trust the utterance itself, whether that trust comes in the form of belief, obedience, or an answer to a question. In the vast majority

---

58 Coady 46  
59 Trusted Tellings 15
of cases, what is needed for telling to give rise to beliefs is not trust, but rather a lack of distrust.

Further, and here I go beyond what Chakrabarti says, belief is possible even in those instances wherein the audience actively distrusts the utterer. You can believe much of what people say even if you think them irredeemable liars, because much of what anyone says is trivial to the point of banality. You may not believe them when it comes to issues on which much of great importance depends, but you can believe the little, inconsequential things. More to the point, if a person says something that is strongly (even if perhaps not logically) implied by some of your beliefs, then you are likely to believe what is said regardless of whether you trust the person or not. Even Maggie, skilled torturer that she is, can believe Molly when the latter says she wants the pain to stop, even if she does not believe anything Molly says to get the pain to stop.

An additional argument in favor of a focus on the absence of distrust instead of the presence of trust is that it allows theories to be consistent with their accounts of other modes of epistemic gain. If knowledge from the testimony of another must involve an inference from the trustworthiness of the other, then it would seem to follow that knowledge from the testimony of one’s senses (that is, perception) must similarly involve an inference from the trustworthiness of one’s sense-organ(s). But Reducibility and Irreducibility theorists alike readily accept that perception is not reducible to inference—that we do not consciously infer from the trustworthiness of our sensory apparati that what we perceive is true, we simply believe it. And here, what is true of perception is also true of testimony:
We do not have to establish the separate trustworthiness of the source; it is enough if no evidence to the contrary is already available. So, the question is not at all whether testimony is a source of knowledge. It is simply dishonest to doubt that. If fallibility of our sensory equipment does not tell against the reliability of perception, and the general skepticism about validity of inductive inferences can be met by a naturalistic or pragmatic response, it will be sheer prejudicial use of double standards to refuse the title of knowledge to testimonial true beliefs on the ground that all understandable utterances could be generally false.\textsuperscript{50}

The fact that people believe their senses unless they are aware of a reason not to is the perceptual analogy of the Principle of Credulity.

According to Nyāya epistemology, testimony is similar to perception in that it yields knowledge unless one or more defeating conditions obtains. Ganeri lists three types of defect (\textit{dośa}) that can prevent a perceptual experience from yielding knowledge: environmental, pathological, and cognitive. Environmental defects include such things as bad lighting and great distance. I may fail to see something while looking directly at (or toward) it because the lighting is insufficient or because the object is too far away to be seen. Pathological defects involve problems with the sensory apparatus rather than the external conditions. If I have cataracts I will not see things as clearly as if my eyes were functioning normally. Cognitive defects involve errors in the way one thinks about or conceives of the input provided by one’s sensory apparatus. Inattentiveness or distraction can cause one not to attend to sensory input at all, while emotional states such as anger or jealousy can color how we interpret such input.

The Nyāya account of testimony acknowledges that it is subject to the same sorts of defect as is perception. Hearing (or reading) a sentence does not lead ineluctably to knowledge any more than looking at an object does. An untrustworthy (misinformed or

\textsuperscript{50} Trusted Tellings 20
A deceptive utterer is an environmental defect, akin to poor lighting. A poor or mistaken grasp of the language being used constitutes a pathological defect, similar to cataracts. And a hearer who is untrusting, or who has incorrect beliefs regarding the utterer’s trustworthiness—such as someone who fails to recognize that a sentence is spoken in jest or as part of a play, or the townsfolk in the story of the boy who cried wolf—is recognized to have a cognitive defect on a par with someone who misinterprets visual input because of fear or anger. Further, assent to what one is told, just as to what one perceives, is defeated only if one is aware that one or more defects is present in the process. People believe what their eyes tell them, unless and until they have reason to believe their eyes to be untrustworthy. Thus, while the presence of a perceptual defect may prevent beliefs thus acquired from constituting knowledge, only awareness of such a defect can prevent the generation of belief therefrom. Similarly, people believe what others tell them unless and until they have reason to believe those particular others to be untrustworthy. And equally similarly, the presence of a defect in the utterer may prevent the audience from gaining knowledge, but unless the defect is suspected, it will not prevent the generation of testimonial belief.

---

61 Interestingly, if an utterer is both misinformed and deceptive the defects may cancel each other out, the net result being an accurate utterance. Chakrabarti compares the case of such a deluded deceiver to the situation of a person who is near-sighted, but who wears corrective lenses. The distortion of the lenses cancels out the distortion of the eyes, yielding a net result of accurate perceptual knowledge. Similarly, one could argue that believing a deluded deceiver can at least generate knowledge and not merely fortuitously true belief. Indeed, the twelfth- to thirteenth-century Indian philosopher Gaṅgeśa, founder of Navya-Nyāya, raises precisely this point in his opus Tattvacintamāṇī. Bimal Matilal explains: “In this case, the hearer has a belief which is true and has it derived from, as far as he knows, a reliable method, and hence we have to accept that she has knowledge. This only means that knowledge by testimony does not depend always upon the intention of the testifier, but rather on what she testifies to, and how she does it” (“Understanding, Knowing, and Justification” 365-366). See also “Trusted Tellings”.

62 The “particular” here is important. Just as it would be a hasty generalization to disbelieve one’s sense of smell after realizing one’s eyes are in error, so too would it be a mistake to disbelieve all people—or even a group of people—because one has found an individual to be untrustworthy.
Reid also recognizes the analogy between perception and testimony. But he identifies an apparent disanalogy between the two that he takes to be important. He notes that testimony admits of intentional deception in a way that perception does not. According to Reid, “Men sometimes lead us into mistakes, when we perfectly understand their language, by speaking lies. But nature never misleads us in this way: her language is always true; and it is only by misinterpreting it that we fall into error.”

According to Reid, then, any failure to gain knowledge from perception is a failure of interpretation, and the fault lies solely with the agent; but even perfect interpretation of an utterance cannot guarantee knowledge because the utterer may be misinformed, deceptive, or otherwise insincere. In short, the important difference for Reid between perception and testimony is the presence of volition in the origin of testimony and its absence in the origin of perception. This disanalogy may well seem important, but the Nyāya account of defeating conditions that we have already discussed provides a ready response that shows this to be a distinction without a difference. Deceptiveness, ignorance, or misinformedness on the part of an utterer are simply species of environmental defects in the process of coming to know, on a par with bad lighting or environmental factors that can give rise to illusions. Perceptual illusions such as a stick that appears bent in the water can be understood as environmental defects, and can qualify as a form of perceptual deception in a manner analogous to a deceptive utterance.

Coady, who has advocated a non-reductionist view of testimony in the tradition of Thomas Reid, also recognizes the parallel between perception and testimony, and the

---

63 Reid 6.XXIV
64 At least some illusions, it should be noted, are more properly understood as cognitive defects rather than environmental ones. Even so, the analogy between perception and testimony is unharmed by the possibility of intentionally deceptive utterances.
concomitant need for both to be equally recognized as *sui generis* means of knowledge-acquisition. He sums up the point well:

If we can have cases of direct knowledge in perception (as when I see a red apple close at hand in full daylight), then surely we can have cases of direct knowledge in testimony, as when I am told by a normally sighted, disinterested, and non-malevolent friend that there is a red apple on the table in the next room. It is no objection to this to cite the fact that I would not know unless the witness were visually competent, disinterested, and non-malevolent, since it is equally true that I would not know by my own perception unless similar conditions, such as normal lighting and properly functioning eyes, were fulfilled.\(^{65}\)

### 2.2.3 Credulity and Gullibility

From what has already been said, it should be clear that the Irreducibility Thesis openly accepts the Principle of Credulity. In this section I am going to argue that the Reducibility Thesis, despite the objections by many of its advocates that credulity amounts to gullibility, relies on the Principle of Credulity just as much as the Irreducibility Thesis does.

In his article, “Testimony, Justification and Coherence” Keith Lehrer gives voice to the gullibility objection by pointing out that the Principle of Credulity (which he calls the “principle of testimony”) seems too good to be true. Such a principle, he says, if adopted, “would appear to suffice as the basic principle of epistemology.”\(^{66}\) His point is that any apparent problem of epistemology can be solved all too easily once you assume the truth of whatever your interlocutor says:

Suppose that we wish to solve some traditional problem, the problem of other minds or the problem of perception, for example. If I ask another whether my perceptual beliefs are true when she perceives the same things I do…she will assure me that my belief is true. If, moreover, I am justified in believing what she tells me, I can justify my perceptual beliefs

\(^{65}\) Coady 123-124  
\(^{66}\) Lehrer 51
in this way. And there is the solution of the problem of the external world. Ask somebody, and they will tell you. Similarly, ask somebody whether they have thoughts, feelings, and sensations, and when they tell you that they do, then you are justified in believing what they tell you. And there is the solution of the problem of other minds. The solution to all other epistemological problems will be no more taxing given the principle of testimony and a sufficiently loquacious discussion partner. All this appears too wonderful to believe.  

According to Lehrer, then, accepting the Principle of Credulity lands one in the inverse of the skeptic’s regress—call it the believer’s egress: from a default position of credulity, it is too easy to escape from doubt, because one can always find confirmation of what one is inclined to believe, and thus it is not possible to genuinely doubt anything.

Elizabeth Fricker expresses similar concerns about the Principle of Credulity. Fricker is engaged in the larger project of arguing against the Irreducibility Thesis, and she sees something like the Principle of Credulity as an important plank in the irreducibility platform. Fricker does not refer to the Principle of Credulity by name; instead she argues against what she calls the Presumptive Right (PR) thesis:

On any occasion of testimony, the hearer has the epistemic right to assume, without evidence, that the speaker is trustworthy, i.e. that what she says will be true, unless there are special circumstances which defeat this presumption. (Thus she has the epistemic right to believe the speaker’s assertion, unless such defeating conditions obtain.)

Fricker’s PR thesis sounds very much like the Principle of Credulity that underlies the Irreducibility Thesis. Fricker seeks to undermine the Irreducibility Thesis by establishing the falsity of a claim (she calls it the “Negative Claim”) that she takes to be necessary for the PR thesis to hold. This claim is that one generally cannot independently establish the trustworthiness of a particular speaker. Since she does acknowledge the possibility of

---

67 Lehrer 51-52
68 E. Fricker 125
testimonial knowledge, she recognizes that “if this Negative Claim is true, then knowledge can regularly be gained through testimony only if there is no need for independent confirmation of the trustworthiness of speakers; that is, if the PR thesis holds.”

The claim that the Principle of Credulity is in fact a Principle of Gullibility is a strong one, and if true it deals a serious blow to both the Principle of Credulity and the Irreducibility Thesis. However, in the course of Fricker’s discussion it becomes apparent that she has set up a straw man. While the PR thesis is a (very strong) version of the Principle of Credulity, it is not the version that I, or anyone I am aware of—including Reid, Coady, and the Naiyāykās—advocate. It is a consequence of the PR thesis as Fricker understands it, for instance, that the fact that something has been said provides “fully adequate” justification for belief. Fricker’s concern with the PR thesis is that she sees it as “a dispensation from the requirement to monitor or assess the speaker for trustworthiness….” She objects to the alleged right to believe blindly, without any critical reflection or epistemic activity; it is her belief that such a right follows from the PR thesis that leads her to call it “an epistemic charter for the gullible.” But this is not an accurate depiction of the views held by those who actually advocate the Principle of Credulity. Certainly it is not the view held by Nyāya, which we have already examined,

---

69 E. Fricker 125-126  
70 E. Fricker 127  
71 E. Fricker 127  
72 E. Fricker 145
or by Thomas Reid, the most prominent advocate of the Principle of Credulity in the Western tradition.\textsuperscript{73}

Central to the gullibility objection is a concern with trust. Fricker argues that the most plausible version of the PR thesis focuses on trustworthiness first, and belief only derivatively. She sees the plausible version, which requires that one assess the speaker for trustworthiness before assenting to what is said, as “a much more limited affair”\textsuperscript{74} than the stronger version that prescribes belief without any epistemic activity whatever. But in being rendered plausible, she contends, the new version actually loses its status as a PR thesis: “Such an attenuated PR is not a PR at all: it is not a dispensation from epistemic activity.”\textsuperscript{75} Again, though, it should be clear that the irreducibility theorists discussed here do not see the principle of credulity as a “dispensation from epistemic activity.” Fricker is simply begging the question against the Principle of Credulity. In reality, the defeasibility of the presumption in favor of credulity and the focus on belief rather than trustworthiness are not mutually exclusive. Belief in what is said, without inferring from considerations of the trustworthiness of the speaker, can be legitimate provided that the belief is subject to defeating conditions. Certainly Nyāya can at least be read this way, and I see no convincing reason that Coady and Reid cannot also be so read, or at least modified to be in accordance.

\textsuperscript{73} The recent film \textit{The Invention of Lying} depicts an alternate reality in which humans never developed the ability to utter falsehoods. When one individual finds himself able to “say that which is not,” everyone else is powerless to detect his lies. Everything he says, no matter how absurd, is accepted at face value by his interlocutors—even if it flatly contradicts what he said the moment before. This utter, naïve trust amounts to an illustration of the view of language that Jonathan Swift describes the \textit{Houyhnhnms} as holding in \textit{Gulliver’s Travels} (See Part IV, Chapter IV). Those who have misgivings about the Principle of Credulity, Lehrer and Fricker included, tend to do so because they see it as understanding language in this overly simplistic way. But such a situation is so obviously absurd that it should not serve as a critique of the Principle of Credulity, but rather as a \textit{reductio} of their depiction of it.

\textsuperscript{74} E. Fricker 143

\textsuperscript{75} E. Fricker 143
The crux of Fricker’s argument against the PR thesis lies in her assertion that it is at least sometimes possible to establish, without recourse to testimony, the trustworthiness of a speaker with regard to a particular utterance. In the course of presenting her argument, she presents what she calls the default-position thesis:

[T]he hearer is obliged, always, to assess the speaker for trustworthiness; but within this exercise, the hypothesis of trustworthiness has special status in that it is the default position – it is to be ascribed, in the absence of positive signs of its opposite.76

What is interesting here is that in ascribing special status to the hypothesis of trustworthiness, Fricker concedes the Principle of Credulity as Reid and Nyāya intend it. Fricker apparently recognizes the threat of the skeptic’s regress, and understands that a default position of credulity is necessary in order for the possibility to exist of establishing the trustworthiness of a particular speaker at a particular time.

Further evidence that reducibilists and irreducibilists alike must accept the Principle of Credulity can be found in another point on which Fricker (mistakenly) takes the two to disagree. “[O]n my account, but not on a PR thesis, the hearer must always be monitoring the speaker critically. This is a matter of the actual engagement of a counterfactual sensitivity: it is true throughout of the hearer that if there were any signs of untrustworthiness, she would pick them up.”77 Fricker takes this to be a point of distinction, but it is not clear what actual view she is really distinguishing herself from. Ganeri, in addressing precisely this point by Fricker, points out that the Naiyāyikas have a similar mechanism at their disposal: the anupalabdhi-hetu, or argument from non-observation. A fully competent hearer will be susceptible to the presence of defeating

76 E. Fricker 144
77 E. Fricker 154
conditions, even while not actively searching for them; the hearer can be performing a competence check “behind the scenes,” as it were. Such a hearer, Ganeri says, “would, in a perfectly legitimate sense, be open to the possibility that the speaker is lying. However, in the normal case, when the speaker is not lying, she will not possess any information or beliefs about the speaker’s credentials, nor should she need to in order to discharge her epistemic obligations.”

2.3 A Problem for Kant

This detour through the issues surrounding the Principle of Credulity is important for our discussion in this chapter because Kant subscribes to the belief that testimony is epistemically fruitful. In itself, that Kant holds this position is unremarkable. What is remarkable, though, is the fact that, based on many of his comments regarding lying, Kant appears to reject the Principle of Credulity. He seems committed instead to an initial position of agnosticism with regard to the truth or falsity of utterances. Kant’s account of declarations is incompatible with the Principle of Credulity. In order to maintain his account, then, he must abandon the Principle of Credulity, and with it the possibility of language being knowledge-generating. But this is too steep a price to pay, as it traps him in the skeptic’s regress.

I noted in the previous section that Reid takes testimony and perception to differ insofar as the source of testimony can be ill-informed or intentionally deceptive, whereas untrue beliefs generated from perception are always a matter of misunderstanding on the part of the perceiver. Concern with this apparent difference seems to be at work in Kant’s thinking as well, resulting in his own misgivings with regard to the Principle of

78 Artha 110
Credulity. This is easiest to see from Kant’s discussion of the cases that he thinks amount to permissible deception. As we have seen, Kant thinks that it is unproblematic for me to deceive my mugger, or to mislead my neighbors into thinking I am going on a trip, because an interlocutor “has no right to infer…a declaration of intent” from someone’s utterance. He also says that an untruthful utterance is a lie “only if there is an express declaration of my willingness to inform the other of my thought.” Both claims are in direct opposition to the Principle of Credulity, under which at least my neighbor could be described as having legitimate grounds to believe what I say simply because I say it—whether that belief is understood as resulting from an inference or as an immediate cognition resulting from the utterance itself. (The principle of credulity can also account for the fact that my mugger has good reason not to believe what I say.)

Kant seems to be as concerned about gullibility as Fricker is, but he does not acknowledge the need for a default-position in favor of trustworthiness as she does. The question, then, is whether we can pull something analogous to the default-position out of the account Kant does provide. The most promising possibility lies in his discussion of declarations. In treating untruthful declarations as tantamount to false promises, and therefore as always immoral, Kant seems to rely on a version of the Principle of Credulity (or Fricker’s default-position thesis) being in play for declarations, but not for utterances more generally. Clearly Kant does not take all utterances to be declarations, for this would amount to claiming that there is in every utterance an implicit promise that what is said will reflect one’s thoughts. If this were the case, then any utterance would automatically generate an obligation to communicate one’s actual thoughts, on pain of contradiction. But this cannot be what Kant means, because there would then be no point
to his focus on declarations. If Kant took every utterance to contain an implicit promise of truthfulness, then he would see *every* untruthful utterance, rather than merely those that include an “express declaration” of the intent to inform, as contravening duty. The whole point of Kant’s focus on declarations is to avoid this absurd position. In order for Kant’s account to work it must be the case that at least some utterances are not declarations. This being the case, the question becomes how to determine which utterances warrant the title “declaration”. Put another way, when, according to Kant’s account, is credulity warranted? To state the conclusion in advance: there is no satisfactory way to answer this question because, as Reid’s discussion of the Principle of Credulity makes clear, a position of agnosticism such as Kant advocates precludes the establishment of a context of credulity.

If I say that I am going to declare my thoughts, then, according to Kant, I must do so. But based on what Kant says, this obligation can only arise if the claim that I am going to declare my thoughts is itself a declaration, because only with declarations is the Principle of Credulity in play. And it is not clear how the Principle of Credulity could ever be taken to be in play when starting from a point of agnosticism as Kant wants to. If I have no obligation to tell you what I think, then how can telling you that I am going to tell you what I think change that? If my claim that I am making a declaration is to be taken as just like any other ordinary utterance, then the audience will, based on the Principle of Agnosticism, not know how to take it, and thus will not take it to be truthful.79 But this would lead to the skeptic’s regress in the same way a general

---

79 This is not to say that they will necessarily take it to be untruthful. A consistent agnostic, unsure about the trustworthiness (or lack thereof) of an utterance, will either seek independent verification of one or
principle of suspicion does. So instead my claim that I am making a declaration must itself be taken to be truthful. But this is to start from a position of credulity rather than agnosticism, which is contrary to the original premise.

One possible way out of this difficulty for Kant is to argue that claiming one is making a declaration itself conveys the obligation to be truthful. Such a claim flips a switch, as it were, and converts a context of agnosticism into a context of truthfulness. Kant says that, “Only when a person voices his opinions can another tell what he thinks, and if he declares that he wishes to express his thoughts, he must also do it, for otherwise there can be no sociality among men.” Essentially, this approach amounts to taking the analogy between declaring and promising to its logical extreme. The idea is that, just as saying “I promise” (in normal contexts) commits one to the sincerity of one’s utterance, so too does announcing an intention to express one’s thoughts. However, taking the analogy to its extreme actually shows that the analogy does not work as Kant wants it to. This is, first, because the institution of promising is itself predicated on credulity rather than agnosticism. Inherent in the act of promising is an assumption that the promisee will take the utterance to be sincere—that is, to be truthful. Kant acknowledges as much with his critique of the false promise in the *Groundwork*. According to Kant, a false promise is based on a maxim that, if universalized, makes this assumption while simultaneously undermining the very possibility of holding the assumption to be true.

Promising does not create an obligation to be truthful where simple assertion creates none; it creates a more significant obligation than does assertion. Promising thus

---

more claims made by the utterance or ignore the utterance entirely in much the way one may ignore an utterance made in a language one does not understand.

80 LE 27:444, emphasis added
cannot be understood in the way Kant wants declarations to be—and neither, I argue, can declarations. Saying “I mean it,” or some such phrase to convey that one is making a declaration does not create an obligation to be truthful where before there was none, it just raises the stakes. If a student makes an appointment to meet with me in my office at 3:00 on Thursday, and I fail to show up to that meeting, I cannot absolve myself simply by saying, “Well, I didn’t promise I would be there. I just said I would, and that doesn’t obligate me to show up.” If I say, even without promising, that I will be there, my student has a right to expect that I will be there. That is, the student has the right to expect that I mean what I say, even if I do not explicitly say that I mean it. Consider also the other side of the issue. In a context where there is no presumption of truthfulness, even claiming to promise cannot create an obligation. An actor who says “I promise” in the course of a play, for instance—even breaking the so-called fourth wall and saying it to the audience—does not promise and has no obligation to be truthful.

It is also worth noting that whether one in fact promises is at least somewhat independent of the speaker’s intentions, while Kant’s analysis depends on the same not holding for declarations. Whether an apparent promise constitutes an actual promise is not entirely up to the utterer. One cannot simply cross one’s fingers and thereby nullify the consequences of having made an utterance that is taken as a promise. Even if the utterance of the promise is insincere, if the other conditions for the promise to be felicitous obtain then the utterer can—and typically will—be held responsible for having promised.81 But even outside the context of promising, in order for linguistic deception

81 For purposes of the present discussion, exactly what the other conditions are can be left open. In *Speech Acts* John Searle indicates nine conditions that, he says, must obtain in order for a person to “sincerely and non-defectively” promise. Most relevant to this discussion is condition number 6a, modified from 6
to work it needs to be possible to be believed while saying what one thinks is false. In Kant’s terminology, one must be taken to be declaring while not actually doing so. If it were always clear whether a person were making a declaration, then it would always be clear whether the audience had a right to expect truthfulness. But if it were always clear when the audience had no right to expect truthfulness, then permissible (i.e., non-lying) untruthfulness would fail to deceive—because the agnostic audience would not have reason to take the utterance as truthful.

The analogy between declarations and promises can only work for Kant if it is always clear to one’s audience whether one is making a declaration; but if it is always clear whether one is making a declaration, then declarations cannot help Kant to explain the difference between permissible and impermissible untruthfulness. The only way Kant can account for the possibility of permissible linguistic deception is by recourse to the mental state of the utterer. It must be possible, on Kant’s account, to speak as if one is making a declaration while not actually making a declaration, simply because one “doesn’t mean it”—a sort of modern-day version of mental reservation. But saying whatever one wants so long as one does not take oneself to be making a declaration is the equivalent of crossing one’s fingers to get out of a promise. The fundamental difference between promises and declarations is that promises operate on the assumption that the Principle of Credulity both is and has been in play, whereas declarations are meant to indicate its coming in to play. This is why promises can avoid the regress while declarations cannot. If it were always clear when an utterance is a declaration, then this

precisely to account for insincere promises: “S intends that the utterance of T will make him responsible for intending to do A” (see Searle 57-62).
might not be a problem. But it is not, and it cannot be, at least not when starting from an initial position of agnosticism.

As was mentioned in Section 1, the duty not to lie can be understood alternately as a duty to others or as a duty to oneself. But applying the preceding analysis of declarations to what Kant says reveals that, since starting from agnosticism prevents one from establishing a context of credulity, Kant cannot support either his claim that lying violates a perfect duty to others or his claim that lying violates a perfect duty to oneself.

To examine Kant’s claim that lying is a violation of a duty to others, it is helpful to approach the question from the perspective of the right-holder rather than the duty-bearer. That is to say, instead of asking when an utterer has a duty to be truthful to his audience, it is more informative to ask when the audience has the right to expect (or demand) truthfulness from the utterer. The question posed in this way points out a limitation of Kant’s discussion in “On a Supposed Right to Lie.” There Kant goes directly from criticizing Constant’s claim that one can have a right to the truth to asserting that one can only have a right to truthfulness in one’s own person. In doing so, Kant avoids what appears to be the spirit of Constant’s objection: sometimes a person does, and sometimes a person does not, have a right to expect truthfulness from another. Kant even admits as much when discussing the mugger, but then he seems to take a contrary stance when dealing with Constant. The recourse to declarations, it seems, is meant to address this issue: the audience can expect (or demand) truthfulness when the utterer makes a declaration. But, as we have seen, this only pushes the question back a step. How can the audience know when the utterer is making a declaration?

Agnosticism, no less than outright skepticism, poisons the well.
The agnosticism with regard to utterances that Kant advocates undermines any attempt to understand lying as a violation of a perfect duty to others, because it does not appear that there can ever be a duty to others to be truthful. One never has a right to demand truthfulness, following Kant’s account, because it is never clear that a declaration is being made. From the perspective of the audience, the account Kant provides cannot justify ever taking the utterer to be speaking his mind, even when he says he is. We are forced to conclude that the focus on untruthful declarations cannot do the work Kant needs it to do to justify his absolute condemnation of lying as violating a duty to others. It is possible that Kant came to recognize this limitation of his account, and that is why his discussions of lying in the *Metaphysics of Morals* and “On a Supposed Right to Lie,” both of which were written after the *Groundwork*, focus on lying as a violation of a duty to oneself instead.

The issue of the duty to oneself not to lie is different from that of the duty to others because transparency is no longer an issue: if it is always clear to oneself whether one intends to make a declaration, and thus to be held accountable for the truthfulness of her utterances, then whether one’s interlocutors correctly assess one’s intentions carries much less import. Given that my utterances need not always reflect my thoughts, the question relevant to the issue of a duty to oneself is: how can it be brought about that a particular utterance (call it a declaration) ought to be reflective of one’s thoughts? Kant’s answer to this question is revealed in the case of the mugger considered earlier: the act of declaring my intention to “speak my mind” turns my untruthful utterance into an untruthful *declaration*, and thus confers the obligation in question. It does so, according to Kant, because the “communication of one’s thoughts to someone through words that
yet (intentionally) contain the contrary of what the speaker thinks on the subject is an end that is directly opposed to the natural purposiveness of the speaker’s capacity to communicate his thoughts…”82 If the purpose of language is the communication of one’s thoughts, then professing to use language for this purpose while surreptitiously doing something else amounts to a denial of the dignity of the speaker himself. It is to treat one’s self as a “speaking-machine,” a mere means for bringing about one’s own ends.

This approach by Kant meets with two difficulties. The first difficulty is that it rests on an implausible notion of the nature of language. When he speaks of the capacity to communicate one’s thoughts as having a natural purpose he appears to rely on the same principle that he uses to argue that the purpose of reason is to enable rational beings to be moral. He makes this principle explicit in Section I of the *Groundwork*: “In the natural constitution of an organized being, that is, one constituted purposively for life, we assume as a principle that there will be found in it no instrument for some end other than what is also most appropriate to that end and best adapted to it.”83 This passage betrays a teleological element in Kant’s thinking. Since reason gives rise to the will, and thus to the possibility of a good will, Kant concludes that the good will is the *purpose* behind humans’ possession of reason. Similarly, Kant argues, since language enables people to communicate their thoughts, the communication of one’s thoughts must be the *purpose* of language. That Kant would reason on the basis of this principle is understandable, writing, as he was, nearly a century before Darwin.84 But given what we know today,

---

82 MS 6:429
83 Kant, *Groundwork of the Metaphysics of Morals* (hereafter GMS) 4:395
84 With regard to language specifically, though, Kant is not entirely without fault, even if he was writing before Darwin. Bernard Mandeville, writing well before Kant, presents a theory of the purpose of language that is directly at odds with Kant’s: “If by man’s speaking to be understood, you mean that…men speak in
this position is simply implausible. The use to which an object or a faculty can be put provides little if any insight into the reasons for that object or faculty’s existence. The assumption that there is a purpose behind the faculty of reason or of language is itself suspect; that the use to which it can be put indicates purpose is even more so. It is not the case, for instance, that we have opposable thumbs so that we can grasp objects. Rather, the fact that opposable thumbs give rise to the ability to grasp objects, which itself proved, in the context in which primates first developed them, to be adaptively advantageous, explains why we today have opposable thumbs. Similarly, just because we can use language to communicate our thoughts does not mean that this is the purpose behind language. There is no reason to think that there is any purpose behind language. Language can be put to many uses, but utility does not indicate purpose.  

The second difficulty is, I think, even more damning. If, as according to Kant, in making an untruthful declaration I treat myself as a mere means and thereby violate my own right to truthfulness, then untruthfulness must always be problematic—not just when making declarations. I cannot just decide to make an exception for myself (to not exercise my right to my own truthfulness) this time. For one thing, it is the intelligible self rather than the empirical self that makes the decision; but the intelligible self is the duty-bearer, not the right-holder. And a duty-bearer cannot make a decision on behalf of the right-holder not to exercise the right. For another, exception-making in general is

---

order that their thoughts may be known, and their sentiments laid open and seen through by others…I answer in the negative. The first sign or sound that ever man made, born of a woman, was made in behalf, and intended for the use, of him who made it; and I am of opinion, that the first design of speech was to persuade others, either to give credit to what the speaking person would have them believe; or else to act or suffer such things, as he would compel them to act or suffer, if they were actually in his power” (185).  

Ron Bontekoe has leveled the same critique of Kant’s teleological assumptions, but much more thoroughly, in his book The Nature of Dignity, pages 11-16.
precisely what moral issues, insofar as morality is lawlike for Kant, do not allow. If, as Kant says, lying is a violation of a duty to oneself because “communication of one’s thoughts to someone through words that yet (intentionally) contain the contrary of what the speaker thinks on the subject is an end that is directly opposed to the natural purposiveness of the speaker’s capacity to communicate his thoughts,” then all deception—not just untruthful declarations, not even just linguistic deception, but all deception—is a violation of the duty to oneself to be truthful. Deception of any sort is a matter of conveying to another thoughts that are contrary to one’s own, and is thus directly opposed to one’s ability to communicate one’s thoughts. This is clearly in tension with Kant’s willingness to accept some forms of deception, such as a pretended trip to uncover and thwart a crime.

When beginning from a principle of agnosticism, there is no way to establish a context wherein truthfulness is obligatory; agnosticism precludes the operation of the Principle of Credulity, and thus the possibility of testimony being epistemically fruitful. Kant is thus faced with two apparent choices: he can either abandon his agnosticism and accept the Principle of Credulity, or he can abandon the idea that language can be knowledge-enhancing, and with it his absolute condemnation of lying. I advocate the former alternative.

**Conclusion: The Importance of Context**

All this is not to say that Kant’s focus on untruthful declarations must be abandoned. Quite the contrary. His focus on untruthfulness rather than untruth is an important insight to hold on to, and one that is not recognized often enough in discussions of lying. Similarly, the idea that it is not always problematic for utterances to be
untruthful is crucial for avoiding the objections of extremism that often accompany absolute condemnations of lying. Kant’s mistake in trying to identify what is wrong with lying, as I see it, lies in focusing explicitly on the properties of the utterance, and not on the context in which the utterance is made. If we look at context, it becomes much less problematic to explain how the audience can know when an utterer is making a declaration. For instance, it is apparent not from what the utterer says or how she says it, but rather from the fact that the utterer is on stage in a theatre during a performance, that what is said should not be taken as a declaration. To use a less well-worn example, consider the small talk that frequently occurs at social gatherings such as cocktail parties. If a new acquaintance says to me, “You should come to the gallery opening on Thursday,” and I respond with, “That sounds good,” my acquaintance may expect me to come, and may be disappointed if I do not show, but it is probably just as likely that there will be no such expectation or disappointment. Contrast this with the student who asks, “Can I meet you at 3:00 on Thursday?” to which I respond, “That sounds good.” Here it is clear both that the student expects me to come and that he will be disappointed if I do not show. Further, if my two interlocutors confront me about my absence, a response of, “Oh, something else came up – sorry,” is far more likely to pacify the party acquaintance than the student, and this because of the context in which I made the statement that the proposed meeting “Sounds good.”

The idea of using context to determine when an audience has a right to either expect or demand truthfulness from an utterer is, I think, promising. But it is not clear to me that it is an option Kant ever considers. In fact, it appears to stand in rather stark contrast with what Kant wants to say. If I tell a mugger that my wallet is on the
nightstand in my bedroom when I know that it is actually in my backpack, I have not lied to the mugger. And I have not lied to him precisely because he has no right to demand truthfulness from me. On this much I agree with Kant. But the reason he has no right to expect truthfulness from me is not because I have not declared my intention to inform him of my thoughts; it is because he is mugging me! Substitute my wife for the mugger in the same story, and it does seem plausible to say that I have lied to her. When she asks why I misinformed her regarding the whereabouts of my wallet, I cannot simply respond, as Kant seems to imply, that it is her fault for believing me since I never made it clear that I was making a declaration. Unlike the mugger, my wife does have a right to expect truthfulness from me precisely because she is my wife. In Chapter 3 I will present a theory of what is wrong with lying that focuses more on the importance of the context in which an utterance is made than does Kant’s account. Specifically, my account will examine the role of trust in the blameworthiness of lying.
Chapter 3: On Trust and Lying in a Moral Sense

Introduction

In Chapter 2, I demonstrated that Kant’s absolute condemnation of lying is not as absurdly extreme as it is often taken to be, because the scope of what qualifies as a lie is narrower on Kant’s account than his interpreters often realize. Even so, his characterization of lying as an untruthful declaration, and thereby as contrary to duty, is untenable because it relies on a notion of duties that follow from a particular (flawed) understanding of language-use, considered in the abstract. I then suggested that a more fruitful approach is to examine the duties that follow from the particular relationships in which lying (and utterances in general) occur. Such an approach is precisely what the present chapter will provide. The analysis will reveal that the moral blameworthiness of lying results from a betrayal of trust. A subsequent analysis of trust itself will provide the tools necessary to provide a definition of ‘lie’.

3.1 The Personal Side of Lying

It is anything but novel to criticize Kantian morality on the grounds that it fails to pay sufficient heed to the exigencies of particular situations. We must be careful, though, not to go too far to the other extreme, where we meet with irreconcilable claims, based on conflicting personal intuitions, about what is right in a particular circumstance. To do so would lead into precisely the difficulty we dealt with in Chapter 1. One useful way to navigate this difficulty is to follow Harry Frankfurt and Robert Solomon, who have both (independently) criticized the Kantian objection to lying on the grounds that it is not sufficiently responsive to lived experience. They both begin their critiques from
reflections on the first-person experience of learning that one has been lied to. Frankfurt says that a focus on universal rules misses the point here, because objecting to lying is much more personal than such an account allows for. The problem with lying is not in what would happen if everyone did it, or in the fact that it would be incoherent to think of everyone doing it, but rather in the immediate experience of being lied to, now, by this particular person. “What is most immediately aroused in our response to the liar,” Frankfurt says, “is not public spirit. It’s something more personal…. What stirs us against them, whether or not they have somehow managed to betray all of humankind, is that they have certainly injured us.”¹ Robert Solomon makes much the same point when he says, “Lying is wrong because it constitutes a breach of trust, which is not a principle but a very particular and personal relationship between people.”² The personal reaction that Frankfurt and Solomon mention, the feeling of being wronged in being lied to, is aptly described by Sissela Bok:

> Those [who] learn that they have been lied to in an important matter—say, the identity of their parents, the affection of their spouse, or the integrity of their government—are resentful, disappointed, and suspicious. They feel wronged; they are wary of new overtures. And they look back on their past beliefs and actions in the new light of the discovered lies. They see that they were manipulated, that the deceit made them unable to make choices for themselves according to the most adequate information available, unable to act as they would have wanted to act had they known all along.³

Bok’s description provides an illustration of the two things that Frankfurt says we object to when we object to being lied to. The first is the arrogance, even hubris, that the

² Solomon 100, emphasis in original
³ Bok 21-22
liar demonstrates. Insofar as a liar attempts to cause his dupe to have false beliefs, he attempts to create an alternate reality; he imposes on his dupe a false world of his own creation. Such an imposition of one’s will on another is understandably offensive, particularly when considering paternalistic lies. When I learn that my doctor has misled me about a treatment option because she thinks the actual truth is too complicated for me to understand, or about a prognosis because she thinks the reality of a situation is too difficult for me to deal with, I may genuinely believe that she meant well and took herself to be acting with my best interests at heart, and nonetheless object to what I perceive as disrespect in her decision to mislead me. But paternalism is not the only aspect of lying that comes off as arrogant. In addition to the arrogance of assuming one knows better than I what is best for me, the liar also typically demonstrates the arrogance of assuming that he can lie without being caught, or without incurring significant repercussions for being caught in the lie. This sort of arrogance is taken to the extreme by Carson’s example of the cheating student who lies to the Dean.

Arrogance can most definitely offend our sensibilities, especially when it is perceived to hinder one’s ability to make one’s own choices, but even so it would be a stretch to say that this alone can ground the intensity of the moral objection to lies that is typically found both in philosophical accounts and in folk conceptions of lying. Arrogance that does not involve deception—say, in a friend who has a particularly high regard for his skill at bowling—can be an endearing quality as much as it can be a ground

---

4 Of course there are exceptions to this. Someone living in Nazi Germany who was hiding a Jewish family could lie to the Gestapo about it while both doubtful that the lie would succeed and terrified of the repercussions of the lie’s failure. This does not change the fact that lies typically involve a belief that one will be able to either avoid getting caught or avoid punishment. (I thank Andy Lambert for pressing me on this issue.)

102
of moral condemnation. The second aspect of lying that Frankfurt takes to be objectionable—namely, the sense of self-doubt that being lied to engenders—points us in a more fruitful direction, albeit indirectly. Frankfurt, following Adrienne Rich, says that being lied to makes one feel crazy—the more so the closer the relationship between the liar and oneself. He says that learning that one has been lied to by a friend exposes to us something about ourselves—something far more disturbing than merely that we have miscalculated, or that we have made an error of judgment. It reveals that our own nature (i.e., our second nature) is unreliable, having led us to count on someone whom we should not have trusted. It shows us that we cannot realistically be confident of our own ability to distinguish truth from falsity—our ability, in other words, to recognize the difference between what is real and what is not. Successfully deceiving a friend implies, needless to say, a fault in the one who tells the lie. However, it also shows that the victim of the deception is defective too. The liar betrays him, but he is betrayed by his own feelings as well.5

There is surely something correct in what Frankfurt says here. It is one thing to believe what you tell me because what you say seems plausible and well-researched. It is quite another to believe what you tell me because I know and trust you, and thus accept what you say as uncritically as I usually accept the testimony of my senses. If I find out that you in fact told me a lie, it is particularly galling because you have coaxed me to bring you within my epistemic defenses and then attacked me from within. The lie from a friend is thus a sort of linguistic Trojan Horse.

Nevertheless, I think Frankfurt is mistaken to place the offense in the feeling of self-betrayal rather than in the feeling of betrayal more generally. Naturally, learning that I have trusted someone whom I ought not to have trusted can lead to self-doubt as much as to doubt about the other, but self-doubt is more likely to lead to fear or insecurity than

5 On Truth 83-84, emphasis in original
to anger or moral offense. It is conveniently symmetrical that Frankfurt’s two bases of objection indicate, respectively, a fault in the deceiver and a fault in the deceived, but it seems implausible to think a fault in the deceived can ground the sort of moral condemnation that is typically assigned to lying. While I may become frustrated, upset, or even outright angry at myself in response to my own shortcomings, I am only likely to feel moral outrage at what I perceive to be a moral failure on my part. And a lack of acuity in assessing trustworthiness does not seem to qualify as a moral failure. On the other hand, another’s willingness to exploit my shortcomings—especially when that other is someone I am supposed to be able to trust—does seem plausible as a ground of moral outrage. It thus seems more reasonable to suggest that the feeling of being betrayed itself gives rise to moral offense and to self-doubt than that the feeling of self-doubt gives rise to a feeling of moral offense.

This shift of focus is consistent with the analysis of lying provided by Robert Solomon, which focuses almost exclusively on the notion of betrayal. “Lying is first of all a matter of interpersonal trust,” he says, “and to say that lying is wrong is to point out that a lie breaches the very trust it necessarily presupposes, not in the abstract sense argued by Kant, but in the very personal and concrete sense that usually goes by the name of betrayal.”6 There seems to be something correct in this analysis. Even someone who acknowledges that all lies are wrong can recognize that some lies offend more than others do, and rightfully so. The relevant difference lies not only in the subject matter of the lie, but also in who does the lying, and to whom. I will likely be more upset to find that my spouse or my priest has lied to me than if I find that a used car salesman has lied to me.

---

6 Solomon 101
Especially if I have been raised in an environment that embraces the dictum *caveat emptor*, I may feel more annoyance than betrayal at the used car salesman. A cheating spouse, on the other hand, is to many the epitome of betrayal.

I think Solomon and, to a lesser extent, Frankfurt, are correct to locate the blameworthiness of lying in the very personal feeling of betrayal, and I think their insight is what Nietzsche had in mind when he wrote one of his most vivid aphorisms: “Not that you lied to me, but that I no longer believe you, has shaken me.”7 A problem with lying is a problem with having one’s trust violated. But it would be hasty to conclude from this that violating someone’s trust is always blameworthy, always a *betrayal*. It typically is blameworthy when the trust violated is that of a genuine friend, but the matter becomes much murkier in what might be called immoral trust relationships. If I am a con artist who has taken you in and I take advantage of your trust to drain your bank account, betrayal is an appropriate sentiment for you to feel in response. This fits perfectly under Frankfurt’s account. But if I instead am a drug kingpin who employs your family in slave labor conditions in my poppy fields, and you are my cook, I trust you not to take revenge on my treatment of your family by poisoning me. Perhaps I trust you not to act against my interests because I take you to be too naïve, incompetent, or fearful to do so, or perhaps it simply does not occur to me that you would be anything other than grateful that you and your family have jobs at all. But if you do decide to act against my interests, to take revenge on me for my treatment of you and your family, it is difficult to say that in such a scenario you have betrayed me, since my trust in you had such a nasty basis.

7 Nietzsche, *Beyond Good and Evil* (hereafter JGB) §183
Trust is a complicated issue: more complicated than Frankfurt or Solomon seem to give it credit for. To more fully understand how the moral blameworthiness of lying can be located in its status as a type of betrayal, then, we need to examine the notion of trust in some detail. Specifically, we need to consider the directionality of trust: the fact that two people can trust each other, but in different ways and for different reasons. The ways we trust and the reasons we have for trusting have great relevance for the moral issues surrounding the maintenance and violation of trust.

3.2 The Importance of Trust

Trust is a huge topic in its own right, worthy of an extended treatment far beyond anything that can be provided here. In her seminal essay, “Trust and Antitrust,” Annette Baier points out that the issue of trust has been largely neglected in modern moral discourse. And to the extent that it has been discussed, she points out that trust has been understood largely on the contractual model, as occurring between non-intimates who have roughly equal power and are largely autonomous:

Modern moral philosophy has concentrated on the morality of fairly cool relationships between those who are deemed to be roughly equal in power to determine the rules and to instigate sanctions against rule breakers. It is not surprising, then, that the main form of trust that any attention has been given to is trust in governments, and in parties to voluntary agreements to do what they have agreed to do. As much as possible is absorbed into the latter category, so that we suppose that paying for what one takes from a shop, doing what one is employed to do, returning what one has borrowed, supporting one’s spouse, are all cases of being faithful to binding voluntary agreements, to contracts of some sort.8

Baier wants to move away from this limited perspective and instead analyze trust as a form of entrusting. This makes trust a three-place predicate, including two (typically

---

8 Baier, “Trust and Antitrust” 249-250
distinct) people A and B, and an (abstract or concrete) object C, such that “A trusts B” amounts to an elliptical form of “A entrusts the care of C to B.” There will be instances, she admits, wherein it will be difficult to state explicitly what C is, or even to identify a plausible candidate—and to that extent it is a sort of Procrustean analysis. Nonetheless, I agree with her that the overall account is helpful and informative despite this limitation. Of particular value in Baier’s approach is that it accounts for the directionality of trust, and the moral relevance thereof, that we identified in the previous section.

One of the most salient features of Baier’s analysis of trust is that it involves a significant amount of vulnerability on the part of the truster. This vulnerability arises as a result of two more basic features. The first of these is that the object whose care is entrusted to the other is something important to the truster. Trusting involves letting someone else care for what one cares about. If I ask you to take care of something that is meaningless to me (say, a disposable coffee cup that I was planning on throwing away), I have not really demonstrated trust in you: I stand to lose nothing. One can imagine doing such a thing to test, or to help develop, someone’s trustworthiness (imagine parents who get a toy dog for their child, with the following promise: “If you show that you can care for this toy dog, then we’ll get you a real puppy”), but a test of trustworthiness already

---

9 For a clear articulation of the potential limitations of the entrusting approach, see Section III of Karen Jones’s article, “Trust as an Affective Attitude.” Her primary point is that such an account fails to account sufficiently for the affective component of trust (by which she means to say that trust is not a purely cognitive attitude, that it has an emotional element, and thus it is not wholly a matter of choice whether one trusts). Her point is well-taken, but her critique reveals, I think, that she takes Baier’s use of entrusting too literally, and not as the heuristic that Baier (I think) intends it to be. Jones says the entrusting model struggles to explain the intuition, which she and Baier share, that trust is not a matter of will. For if entrusting is an act of the will, Jones’s argument goes, then trusting should be as well. But one could plausibly argue that when I retrospectively find myself to have trusted, I also find myself to have entrusted something to the care of the trusted person—even if implicitly or unconsciously. We could draw the conclusion that the entrusting model of trust can expand our understanding of entrusting instead of only constraining our understanding of trusting.
reveals a lack of trust.\textsuperscript{10} A romantic partner who professes trust while continually testing his lover’s commitment is either confused or disingenuous.

The second feature of trust that gives rise to the vulnerability of the truster is that it involves affording discretion to the trusted. Discretion is what enables us to differentiate trust from mere reliance. As Baier points out, trust is a form of reliance but is not coextensive with it, for you can rely where you do not trust. A blackmailer, for instance, may rely on his victim’s fear, pride, or sense of shame to pay rather than go public, but can hardly be said to trust the victim not to go to the police: a blackmail victim who chooses to report the crime to the police can not really be said to have violated the blackmailer’s trust. Further, we can rely on machines and other inanimate objects or processes in situations that can be described as trusting only metaphorically. I certainly rely on my car to get me where I am going, on the deadbolt on my door to deter potential burglars, and on my phone to let me know if someone is trying to reach me. Calling such objects “trusty” amounts to calling them reliable—we do not typically call them trustworthy in the way we would a person whom we trusted. And this difference can be attributed to the fact that it makes little sense to speak of affording discretion to such objects. Similarly, when I rely on someone’s psychology or habits I am counting on her to act in a specific, predictable way. But when I trust someone, I do not count on or expect any particular behavior; instead, I count on her to make her own decisions and in so doing to demonstrate a certain sort of attitude—one of benevolence, say, or at least of non-malevolence—toward myself or whatever I have trusted her with. I rely on her to

\textsuperscript{10}But note that a lack of trust does not necessarily imply the presence of distrust. This important distinction will be discussed in Section 3.3.
decide for herself what the best course of action is, and to act on that decision without an over-reliance on direct input from me.

Trust, and the discretion it entails, is a response to human finitude. Individuals have a limited amount of time and energy, and cannot be everywhere at once. Sometimes I simply need to let others take care of something for me. “Since the things we typically do value include such things as we cannot singlehandedly either create or sustain,” Baier notes, “we must allow many other people to get into positions where they can, if they choose, injure what we care about, since those are the same positions that they must be in in order to help us take care of what we care about.” And I cannot dictate precisely how they should express their caring in response to every possible contingency. Even seemingly trivial cases of trusting another person involve discretion, precisely because some contingencies will remain unforeseen. Baier responds to the objection that there is no discretion involved, for instance, in trusting one’s child to put an important letter in the mailbox by pointing out that

life is full of surprises—in Washington on Inauguration day mailboxes were sealed closed as a security precaution, and in some parts of Manhattan mailboxes are regularly sealed after dark. One trusts the child to do the sensible thing if such an unforeseen problem should arise—to bring the letter back, not leave it on the ledge of the sealed mailbox or go too far afield to find another.

In circumstances such as this we do trust the child, and are vulnerable in that trust, but we still have a good idea of what the child should do. We can take further steps to educate the child regarding what to do in different possible circumstances, thereby reducing our vulnerability and/or making our trust more justified. However, there are

---

11 “Trust and Antitrust” 236
12 “Trust and Antitrust” 237, n.4
many more circumstances wherein we trust others when we do not know what to do—in fact, trust is very often predicated on the fact that one does not know what to do, and that the trusted has epistemic privileges that the truster does not share—as when I trust my mechanic to fix my car without performing unnecessary repairs or sabotaging the car, or when I trust my doctor to diagnose and prescribe appropriate treatment for my condition. Kenneth Kipnis uses the example the adversarial system of law to highlight the risks that the fact of epistemic privilege can give rise to. In a system complex enough that average citizens cannot be expected to navigate that system on their own, expert representation is a necessary condition of fair treatment. “But,” he says, “if the laity’s lack of expertise makes it impossible for them to negotiate the court system on their own, it is equally impossible for the uninitiated to distinguish the masters from the pretenders. Under these circumstances, the system of justice itself can be a major source of injustice.” Part of Kipnis’s point is that the system itself must account and correct for this, in part by adopting—and creating a system to reinforce—professional standards of conduct. The importance of systemic trust to the possibility of trust in individuals will be discussed in the next section. For now, the important point is the central role that vulnerability plays in trust.

Trusting people in positions of epistemic authority makes us vulnerable because we cannot properly assess either their behavior or their intentions:

\[J\]ust what they should do to put right what is wrong is something we must leave to them. Should they act incompetently, negligently, or deliberately against our interests, they may conceal these features of their activities from us by pretense that whatever happened occurred as a result of an honest and well-meaning exercise of the discretion given to them.

---

13 Kipnis, “Ethics and the Professional Responsibility of Lawyers” 573
This way they may retain our trust and so have opportunity to harm us yet further. In trusting them, we trust them to use their discretionary powers competently and nonmaliciously, and the latter includes not misleading us about how they have used them.\(^\text{14}\)

For Baier, the species of reliance that qualifies as trust is “reliance on others’ competence and willingness to look after, rather than harm, things one cares about which are entrusted to their care.”\(^\text{15}\) This reliance on another’s good will and competence is what Karen Jones calls the optimistic attitude of trust. We afford others discretion to the extent that we are optimistic about their good will and their competence in executing that will.

In summary, then, trusting is a matter of giving someone else discretion over the care of something one cares about. It follows that betraying someone’s trust would amount to taking unfair advantage of the vulnerability that results from their trusting. The morally weighted notion of unfairness is appropriate here because, as we have already noted, violating trust (as opposed to betraying it) is not always immoral. This is true for two reasons. First, trust itself is not an intrinsic good. Precisely because trust is necessary for so much of social interaction, it is also a breeding ground for much that is immoral: “Exploitation and conspiracy, as much as justice and fellowship, thrive better in an atmosphere of trust.”\(^\text{16}\) Not all relationships that involve trust are morally praiseworthy, nor should all trusting relationships be maintained. And if one finds oneself in a morally rotten trust relationship, the right thing to do might very well involve violating that trust. As Baier cleverly points out, “trust-busting can be a morally proper goal.”\(^\text{17}\) Of course ending a trusting relationship need not always involve a violation of

\(^{14}\) “Trust and Antitrust” 239-240
\(^{15}\) “Trust and Antitrust” 259
\(^{16}\) “Trust and Antitrust” 231-232
\(^{17}\) “Trust and Antitrust” 232
trust; one can honestly and openly opt out of such a relationship instead. That said, it is not always as simple as this statement makes it out to be: someone may persist in trusting me even after I disavow trustworthiness, and in so doing attempt to obligate me in ways that I do not accept. This point helps to illustrate why, in assessing any violation of trust, it is essential to assess the trust relationship in which the violation occurs. We will return to this point shortly.

To see the second reason that violating trust is not always immoral, it will help to recall that Baier points out three ways that one can violate trust: through incompetence, negligence, or ill-will. Violations of trust that stem from ill-will or malevolence seem pretty clearly to qualify as immoral, while it may be less clear how we should assess violations that result from incompetence or negligence. Incompetence and negligence both may be morally blameworthy, but they equally well may not be. Negligence, for instance, could be a result of absent-mindedness, of having taken on too many commitments, or instead of being non-culpably precluded from delivering on a promise. I may be unable to take you to the airport as I promised because I have to take my child to the emergency room or because I lent my car to a neighbor who wanted to use it to impress a date. The former is fully excusable; the latter is less so. And the same considerations apply to incompetence. If I know someone who is well-meaning but unable to deliver, I am as likely to take this to be unfortunate as immoral. But what if I make a promise that I have good reason to believe I will be unable to fulfill? For both incompetence and negligence, the question seems to come down to the will behind the shortcoming. If I am willfully incompetent or negligent then my violation of your trust is more likely immoral than if I do everything in my power to live up to expectations. The
lesson to take from this is twofold: first, the three ways in which one may violate trust are not mutually exclusive; and second, the presence of ill-will is a strong indicator (though perhaps not a necessary condition) of the blameworthiness of a violation of trust. This makes sense, given our understanding of trust as reliance on another’s good will. If the other shows me the opposite trait of what I am relying on, it is understandable that such a behavior will meet with some amount of disapprobation. In the interest of clarity, let us call violations of trust that are immoral betrayals, and non-immoral violations disappointments.

While the distinction between ill-will on the one hand, and negligence and incompetence on the other, helps to differentiate disappointments from betrayals, this is not the end of the story. Life is not that simple: we search in vain if we search for a simple algorithm that will clearly distinguish disappointments of trust from betrayals of trust: there are simply too many potentialities to account for in trust and the violation thereof to know the answer in advance. One issue is that the betrayal-disappointment distinction risks obscuring the fact that this need not be an either-or type of situation. Instead of simply asking whether a particular violation of trust is a betrayal or a disappointment, it will frequently be more useful to ask to what degree a particular violation qualifies as a betrayal, and to what degree it is merely a disappointment. Still, answering the question of the extent to which ill-will motivated a violation of trust will go a long way toward answering the question of the extent to which that violation qualifies as a betrayal. But there are other factors that are important to consider when reasoning about such situations as well.
To decide the extent to which a violation of trust within a personal relationship qualifies as a betrayal, it is necessary to ask both about the trust relationship in which the violation occurs and about the violation itself. As trust can be used to good or to bad ends, there are moral and immoral trust relationships. Baier makes the important point that, “When the trust relationship itself is corrupt and perpetuates brutality, tyranny, or injustice, trusting may be silly self-exposure, and disappointing and betraying trust, including encouraged trust, may be not merely morally permissible but morally praiseworthy.”18 She then provides a sort of expressibility test for deciding whether any particular trust relationship is morally decent, and thus warrants both trustworthiness and trustfulness, or not. Baier argues that morally sound trust relationships could survive the articulation of the basis on which each member trusts the other, while immoral trust relationships would tend to be destroyed by full disclosure. So, for instance, if I trust you because we share a concern for some common good, or because I think you take pride in doing high-quality work, then your awareness that this is why I trust you would be more likely to strengthen than undermine your desire to live up to my trust. If, on the other hand, I trust you because I think you are too stupid or gullible to realize that I am taking advantage of you, or because I think you are too afraid to do anything about it, then knowledge of this fact would tend to undermine your commitment to not letting me down.

This expressibility test provides a helpful guideline, but also leaves room for less generalizable influences, which will differ from one relationship to the next. As Baier

18 “Trust and Antitrust” 253
acknowledges, the test will be more easily applied in some cases than in others, and much will depend on the particularities of the people who comprise the relationship in question:

What will be offensive forms of reliance on one’s psychological state will vary from context to context, depending on the nature of the goods entrusted and on other relationships between the trusting and the trusted. Variations in individual psychology will also make a difference. Some are much more tolerant than others of having their good nature or preoccupation taken advantage of—not merely in that they take longer to recognize that they are victims of this, but they are less stirred to anger or resentment by the awareness that they are being deceived, blackmailed, or exploited in a given trust relation.19

Even in a circumstance where it would be reasonable to take offense at another’s lack of good faith or candor, for instance, one may not take the violation of trust to qualify as a betrayal. And how the truster takes the violation is a relevant factor in the objective moral assessment of the violation itself.

The difficulty inherent in assessing, either in advance or after all the facts are in, the moral status of any particular violation of trust stems at least in part from the role that caring plays in trust relationships. Nel Noddings points out how difficult it can be in looking at any particular situation to decide whether someone’s action (or inaction) qualifies as caring, or perhaps only as attempting to care, or even attempting to appear as if one cares. The value of thinking about which criteria to consider, she says, lies not in generating a comprehensive account of what qualifies as proper caring, but rather in better understanding that such an account is not what we should be looking for to begin with. What we need instead is the ability to be responsive to particular relationships by considering them in their own lights rather than through the lens of a theory that may or may not account fully for present circumstances: “If we can understand how complex and

19 “Trust and Antitrust” 256-257
intricate, indeed how subjective, caring is, we shall perhaps be better equipped to meet
the conflicts and pains it sometimes induces.”

Given the central role of care in trust, there is no way to provide an exhaustive list
of all the considerations that could factor into the moral evaluation of any trust
relationship, and thus into any particular violation of trust, in advance of the judgment’s
need to be made. That said, we can do a better job than has so far been done in the
philosophical literature of setting the parameters for such evaluation. Some factors that
will frequently warrant consideration include: Was the trust solicited? How easily could
the trusted have extricated herself from the trust relationship? Was the trusted aware of
the trust? Was trust a necessary condition of the relationship’s proper functioning? It is
worth examining this final question in some detail because it reveals that an element of
Kantian moral reasoning is more relevant to these considerations than is often
acknowledged by those engaged in such thinking.

Some relationships do not require trust at all, while others rely on trust for the
mere possibility of proper functioning. I need not trust a car salesman in order to do
business with him—a legally binding and enforceable contract can form a sufficient, if
not superior, substitute. My doctor, on the other hand, cannot do her job properly if I do
not trust her enough to share (potentially) relevant personal details with her. To the
extent that trust is necessary for the proper functioning of a relationship, there seems to
be something presumptively morally wrong with violating that trust. This, again,
explains why I am more likely to feel annoyance at an untrustworthy car salesman, but

20 Noddings 11
21 The fact that I may not know which details are relevant makes the need for trust even more acute.
betrayal when faced with an untrustworthy doctor or spouse. This issue is particularly important because it seems to be related to what Kant was trying to get at in condemning lies as violating a principle of language use. While violations of trust do not render sociability and profitable dialogue with others impossible, they can severely limit the possibility of both in certain more narrow contexts. We can see shades of the categorical imperative in the claim that there is something morally wrong in violating trust that is a precondition of the relationship in which the violation occurs. Consequently, we can also see that Frankfurt and Solomon are at least a bit overly hasty in thinking that identifying the wrongness of lying in betrayal is fundamentally at odds with a Kantian analysis.

Given that we are dealing with a morally decent trust relationship, the moral evaluation of a violation of trust turns to the violation itself. To begin with, we need to know whether, and how easily, the violation could have been avoided while still bringing about the same or substantially similar results. Generally speaking, the more easily-avoidable the violation, the more likely it is a betrayal. Consider, for instance, if I hire a teenager in my neighborhood to take care of my landscaping. I trust him to care for my yard, while he trusts me to pay him our agreed amount in exchange and not to make unreasonable demands on him regarding what qualifies as taking appropriate care of my yard. If he mangles my lychee tree by pruning it poorly (and not, let us assume, out of spite), such a violation is a result of incompetence. If he did it because he mistakenly thought that lychee trees respond well to the same sort of pruning technique as do roses, which he is skilled at pruning, such an example of incompetence seems less blameworthy

---

22 This may seem more like reliance than trust. But if neither of us demands a written contract, it can be understood as trust.
than if he had no pruning skill whatsoever, and just thought he should give it a shot on my prize lychee tree.\footnote{Incidentally, if he knows that it’s my prize lychee tree, rather than just a lychee tree, his pruning mishap is more likely to qualify as a betrayal as well. The more the truster cares about something, the more careful must the trusted be in caring for that thing.} If I specifically requested that he prune the tree, on the other hand—especially if he told me that he was unsure how to do it—then his mangling of the tree was less avoidable and consequently less blameworthy.\footnote{At this point, though, it also becomes less clear to what extent he is even violating my trust. If he tells me that he doesn’t know how, and I insist that he do it anyway, that I trust him to do it, then I pretty much need to be okay with any result that clearly qualifies as pruning. His claim that he does not know how to do it raises the issue of whether my trust in him is appropriate.}

Or consider another possibility. In this scenario he is fully competent at all the aspects of landscaping that I ask of him, but he does nothing for two weeks even though we agreed that he would work on the yard at least once a week. Whether such neglect qualifies more as a betrayal or a disappointment of trust depends at least somewhat on whether he was out playing with friends (probably a betrayal) or was busy with end-of-semester school work (probably a disappointment). Even if it was for schoolwork, though, one might contend that his failure to let me know in advance constitutes a betrayal of trust. But we can modify the story such that he did tell me in advance, but, absent-minded professor that I am, I simply forgot that he told me. Given the plausible interpretation that school is more important than a side job, and that he is busy because it is a busy time of the semester and not because he was neglectful of his schoolwork earlier in the semester, then it would have been rather difficult for him to avoid violating my trust, at least without committing the graver offense of neglecting his school work. As a result, this seems more of a disappointment than a betrayal.
These considerations should make it apparent that we cannot expect a clear criterion for assessing the moral status of particular trust violations. Nevertheless, we can use the foregoing discussion to come up with a useful heuristic. One way to think of it is that a violation of trust is likely to constitute a betrayal only if it is a reasonably avoidable violation that occurs in what the person whose trust is violated has good reason to believe, prior to the violation, to be a morally appropriate trust relationship.25

3.3 Reciprocity in Care: Limitations of the Entrusting Model

We turned to Baier’s entrusting model because it accounts for the directionality of trust. But the model has limitations, which Baier acknowledges, when it comes to dealing with systemic trust and with brief trusting encounters. This should not be surprising, as these are the cases that problematize the directionality of trust. Central to the entrusting model of trust is care, but an important aspect of care is reciprocity. As Noddings says, “[M]y caring must be somehow completed in the other if the relation is to

25 The preceding discussion of trusting as a form of entrusting bears strong affinities with discussions of fiduciary relationships within the law. *Black’s Law Dictionary* defines a fiduciary relationship as a relationship in which one person is under a duty to act for the benefit of the other on matters within the scope of the relationship. Fiduciary relationships — such as trustee-beneficiary, guardian-ward, agent-principal, and attorney-client — require an unusually high degree of care. Fiduciary relationships usually arise in one of four situations: (1) when one person places trust in the faithful integrity of another, who as a result gains superiority or influence over the first, (2) when one person assumes control and responsibility over another, (3) when one person has a duty to act for or give advice to another on matters falling within the scope of the relationship, or (4) when there is a specific relationship that has traditionally been recognized as involving fiduciary duties, as with a lawyer and a client or a stockbroker and a customer. A full discussion of the ethics of fiduciaries would be outside the scope of this essay, but it should be clear from this definition that discussions of trust relationships in the everyday sense and of fiduciary relationships in the legal sense can usefully inform one another. Thus, while I think Baier is correct to move away from the legalistic contract model of trust that has been prevalent in much of modern moral philosophy and toward something more responsive to the realities of relationships between non-equals, it is also important to recognize that an understanding of the legal issues can still contribute much to the conversation about trust in a non-legal framework.
be described as caring.‖

I cannot effectively care for you, or for something that is important to you, unless you engage with me in a way that helps me to understand what you take appropriate care to consist in. This is not to say that you need to care for me in the same way I care for you: reciprocity is not the same as symmetry. But if you do not reciprocate at all, if you are unreceptive to my attempts at caring, then I cannot respond to you as a unique individual. If you do not help me to know what you want or need, then I cannot help but be presumptuous in my attempts to care for what is important to you: any attempt on my part to care for you can as easily end up hindering you as helping. One consequence of this is that a lack of trust tends to be self-fulfilling. If you do not trust me, then you will be disinclined to share with me in a way that enables me to care for you, which will lead to my being unable to care for you and thus being in some sense unworthy of your trust.

The reciprocity of care is important at this point in our discussion because it reveals why systemic trust and brief trusting encounters present difficulties for the entrusting model of trust. Both cases are characterized by a lack of reciprocity, and thus make caring a much more difficult prospect. Let us consider each in turn.

When we understand trusting as entrusting we take for granted the network of trust that is presupposed in any particular case of trusting. Baier says, “[T]here are some people whom one would not trust with anything, and that is not because one has considered each good one might entrust to that one and rejected that possibility.”

I might not trust you with anything, simply because I do not trust you. I may think your

---

26 Noddings 4
27 “Trust and Antitrust” 258
interests so fundamentally at odds with mine that I find it safer not to trust or, if possible, even to rely, at all.\textsuperscript{28} This may be because of something specific to you as an individual, or in your relationship to me, but equally likely is the possibility that I identify you as part of a group whose interests I take to be fundamentally opposed to my own, or to the interests of those like me (however I may conceive of “like me”). Trusting relationships become more the exception than the rule if there is no general environment of trust wherein trusting relationships can develop. The absence of a trusting relationship, as we saw with the example of Maggie and Molly in the previous chapter, leads to the skeptic’s regress and thus to epistemic paralysis.

Naomi Scheman says that to overlook, or to take for granted, the network of trust that underlies the possibility of trusting relationships is to ignore the specifically epistemological nature of the issues, in particular, the centrality of the notion of justification and the importance of attending to the background context that, for particular believers, either grounds or undermines the trust that is needed for any reasonable standard of justification to be met.\textsuperscript{29}

While it makes sense to say that an environment of trust is a condition of the development of trusting relationships, it should be clear that this involves a shift in the notion of trust. Systemic trust involves trusting things—cultural norms, social institutions, one’s not-yet-born descendents—to which one frequently cannot provide the necessary feedback in order for appropriate caring to be possible. The reciprocity that is necessary for a trusting relationship is thus absent in (many) instances of systemic trust.

\textsuperscript{28} Note, though, that I may distrust you not because I think you bear me ill will, but rather because I doubt your competence to successfully execute your good will toward me. If I have been disappointed by you too many times—if you are a close friend or a family member who is also an addict, for example—I will be disinclined to make myself vulnerable to further disappointment despite your sincerest protestations that this time you’ll really quit.

\textsuperscript{29} Scheman 36
Scheman argues that systemic trust can be understood in terms of objectivity rather than caring, as the purpose of objectivity is to rationally ground trust in an institution or an epistemic process. A lack of systemic trust indicates a perception that the system in question is not in fact objective. Objectivity is important, Scheman says, because many claims are intended to be both applicable to and engaged with by people who were neither involved in nor privy to the establishment and justification of those claims. If the institutions whose role it is to certify such claims are perceived to be untrustworthy, then the claims they make cannot have their intended force. Even if an institution is a reliable tracker of truth, if it is not trusted, it cannot perform the role it is intended to perform, and thus cannot serve the people it is meant to serve.

While Scheman focuses mostly on science, she illustrates this point effectively with a reference to the American criminal justice system:

The task of the prosecution is to convince a jury that the prosecution’s case demonstrates the defendant’s guilt beyond a reasonable doubt. Meeting that epistemic standard is in general impossible without the jury’s having a great deal of trust in the competence and integrity of those who are presenting the state’s case. To a certain extent what the state (in the person of police, forensic experts, and the prosecuting attorneys themselves) presents can be backed up relatively independently of what those people say, but to a great extent the credibility of testimony rests on the credibility of the testifiers, and, crucially, of the institutions within which they work and that enforce the norms to which they are accountable. When, as arguably in the O. J. Simpson case, the jury is convinced that those institutions are corrupt—that they are, specifically, racist—there does not have to be a story to tell about just how that corruption could have tainted the specific evidence on which the case rests. Rather, all that is necessary for the standard of persuasion beyond a reasonable doubt to be unreachable is that the jury not find the institutions of the state trustworthy.30

30 Scheman 36
To call an institution or an epistemic process objective is to say that one can justifiably let it take the place of doing the epistemic work oneself. That is, an objective institution is one that someone can safely allow to care for what one is not in a position to care for oneself. When systemic trust is lacking, when an institution is perceived to be self-interested or otherwise biased, it is exceedingly difficult to trust an individual who is part of the institution in question. I have to trust the institution you are a part of in order to trust you as a representative of that institution. Even if I think you are sincere in your beliefs and in your efforts to help, if I do not trust the institution that you represent then I will not trust you to care for what I care about. But in many cases there is little I can do to provide the institution with sufficient feedback for it to enable you, in your capacity as a representative of the institution, to care for me.

Baier’s account of trusting as entrusting effectively takes systemic trust as unproblematic, analyzing trust relationships within a framework wherein the legitimacy of the surrounding institutions is either taken for granted or already established. To a certain extent, the discussion of lying in this essay has to do the same thing. This should not be taken to imply that a member of an institution that is taken to be un-objective cannot lie, or that such lying is unimportant, but rather that other considerations take precedence over the question of lying. If a person does not trust the institution I represent, there are bigger issues at play than whether I am being sincere—especially if the mistrust in my institution is well-founded.

The entrusting model’s difficulty with brief trusting encounters has more direct implications for the account of lying that I am developing than does the difficulty with systemic trust. Brief trusting encounters are circumstances wherein we momentarily put
our trust in strangers, people whose trustworthiness we know little or nothing about. Such encounters, as Baier indicates, are quite common: “We trust those we encounter in lonely library stacks to be searching for books, not victims. We sometimes let ourselves fall asleep on trains or planes, trusting neighboring strangers not to take advantage of our defenselessness.”31 Baier says that the expressibility test cannot be applied to brief trusting encounters. It would be better to say that the test can be applied to them, but there is little point to doing so: at stake in the expressibility test is the question of whether to continue the trusting relationship as it stands, but in brief trusting encounters this question of the future generally does not arise. A key difference between trust relationships and trusting encounters is that the former are open-ended and oriented toward the future as much as to the present, while the latter are concerned only with the present moment. This is precisely why it is appropriate to call them brief trusting encounters, rather than trust relationships. Granted, any encounter with another person is a sort of relationship, but brief encounters with strangers are not typically the sort of relationships that are assessed in terms of trust. And to the extent that this is true, we need to acknowledge that the “trust” in “brief trusting encounter” is a bit of a misnomer.

This is not to say that it is incorrect to call such brief encounters trusting encounters any more than it is incorrect to call systemic trust by that name, only that the term is being used in a sense different from how Baier uses it in discussing more traditional, enduring, trust relationships. To see this, consider the primary example that Baier uses to discuss brief trusting encounters: the stranger in the library stacks. As she says, in circumstances such as this, what we expect is generally to be left alone. But my

31 “Trust and Antitrust” 234
desire to be left alone, communicated as it so typically is by leaving the other person alone, means precisely that I am not engaging with this person. Such non-engagement means that I am not enabling this person to care for me in any real sense. So if my orientation toward the stranger in the library stacks is one of trust, it is so only very cursorily.

It could be objected that if a person desires to be left alone, demonstrating the same behavior is the most effective means of engagement. Certainly it would be odd for me, upon first encountering a stranger in the library, to say, “I’d appreciate being left alone, thank you.” This objection has some merit. People can interact with and communicate their interest to one another in a wide variety of ways. But if my desire is to be left alone, not to be interfered with, then there is a very real sense in which what I want is for my interloper not to even attempt to care for me. Especially since I am giving the person so little to work with, any such attempt will involve a substantial amount of discretion, and there is no evidence that I have intended to afford such discretion to anyone when what I really want is just to be left alone. Remember that discretion is a key aspect of the vulnerability that characterizes trust. If this person decides to try to care for my well-being by lecturing me about the dangers of trusting strangers in library stacks, or by preaching the gospel to me, I am more likely to take this to be impertinent and presumptuous than trustworthy behavior. To the extent that I expect a person to leave me alone I do not rely on good will, the exercise of which typically requires discretion, but rather on a lack of ill will. Any sort of good will that the person possesses will either be supererogatory (the person silently wishes me well, for instance) or presumptuous (the person attempts to take liberties with the care of my well-being that I
did not grant). Relying on a lack of ill will does not typically involve affording the other discretion, and so does not involve trust.

As frequently as we rely on others for no more than non-interference, however, we also often rely on strangers for more than this, even in brief trusting encounters. When asking for directions in an unfamiliar city, or asking a stranger what time it is, non-interference is not what I hope for—at least on one understanding of non-interference. It would qualify as non-interference if the person I ask walks away from me without answering my question, for instance, but this would not fulfill my hopes or my expectations. On the other hand, there is another sense of non-interference in which this is precisely what I am hoping for: the sense in which I hope that my interlocutor will function as an extension of the non-sentient instrument that could provide the information I am seeking. When inquiring into the time with a person who has a watch, I want him to give an accurate report of what the watch says, not an interpretation: I want a scribe, not a commentator. Similarly, in asking for directions I want an accurate report of what a map would tell me. In both cases, the accuracy I am counting on is not qualitatively different from the accuracy I would count on from a watch or a map; and to the extent that this is true, it is legitimate to say that I count on my interlocutor for non-interference. To a certain extent, it would be appropriate to say that what I want in these cases is for the person to be trusty rather than trustworthy.

Nevertheless, these accounts differ from those in which I want complete non-interference in that I rely on an amount of competence from my interlocutor in a way I do not from my fellow library patron. Reporting the time requires the ability to tell time, and giving (responsible) directions requires knowing how to get from here to there. This
sort of brief trusting encounter thus comes closer to the sort of trust involved in enduring trust relationships. But such encounters still lack the reciprocity necessary for full-fledged caring to be involved, as can be seen by the limited discretion that is afforded.

Granted, there is more room for discretion here than in the first cases considered, but still not as much as in enduring trust relationships. If there are two ways to get to my destination and one of them takes me through a very dangerous part of town, hopefully my guide will provide me with the other. If the person of whom I ask the time habitually keeps his watch ten minutes fast, I would hope that he decides to subtract those ten minutes before reporting the time to me. But this differs (quantitatively, if not qualitatively) from the discretion my spouse would be afforded in a similar circumstance. My spouse, perhaps knowing that my question about the time reflects a concern that we not be late for dinner reservations, can respond in a number of possible ways that would qualify as trustworthy, many of which do not even involve telling me the time. She could simply say, “We’re fine, don’t worry about it,” or “We still have twenty minutes,” or “Yeah, let’s get going.” My spouse knows more about me, and about my needs and desires, and so is able to care for me in a more thorough way than is the stranger who I ask to tell me the time. To make use of a distinction that I will draw in Chapter 4, my spouse is able to respond to my inquiry rather than merely answering my question, because she knows the intent behind my question.

While systemic trust and brief trusting encounters do not very well fit the entrusting model of trust that Baier provides, this need not be taken as a critique of the model. Rather, the issues considered here reveal that such situations have less to do with trust, at least in the more robust sense that Baier’s account focuses on, than might
otherwise be thought. As with systemic trust, brief trusting encounters frequently lack the
c reciprocity that is necessary for appropriate caring. It follows that such encounters do
not involve entrusting the care of something to the discretion of another—even if they do
involve vulnerability on the part of the truster.

In the beginning of this chapter I considered the possibility that the wrongness of
lying lies in its status as a betrayal of trust. Armed now with a sophisticated notion of
trust, we can turn to the question of what exactly goes wrong in a trust relationship when
one person lies to another.

3.4 Trust and True Beliefs

I said in Chapter 2 that it is possible to believe where one does not trust. Baier’s
distinction between trust and reliance provides a theoretical underpinning to this claim.
When Maggie believes Molly’s cries that she wants the pain to stop, her belief is based
on her knowledge of human physiology and psychology, not on her assessment of
Molly’s sincerity. This is a belief based on reliance rather than on trust. Nevertheless,
our beliefs very often are predicated on trust. If I do not trust you, the range of things
regarding which I will believe you is rather narrow, especially if what you tell me
conflicts with something I already believe. If I trust you, though, I will at least
sometimes believe you even in the face of countervailing evidence. Much of our learning

---

32 Granted, Maggie’s knowledge of physiology and psychology are themselves based on trust—the specific
trust in her teachers and/or in the authors of the books that she learned from, as well as the systemic trust in
the scientific fields of physiology and psychology, among others. But while such considerations serve to
reinforce both the importance and the ubiquity of trust in everyday life, they also do little to help us
understand the reasoning that Maggie engages in when engaged in the act of torturing Molly. To the extent
that she takes her beliefs regarding physiology and psychology to constitute knowledge, she is relying on
them directly and not considering (even implicitly) the reliability or trustworthiness of their sources. And
this will tend to remain true unless and until her expectations are brought up short, at which point she will
be forced to re-consider her beliefs.
occurs when friends, family members, teachers, or others who we take to be authoritative inform us of something that is in tension with or even flatly contradicts what we previously believed, either on the basis of what someone else has told us, or due to our own inferences from our experiences. When we learn that the earth revolves around the sun rather than vice-versa, or that the things we see on movies and television are not real, or that the painful jab of the doctor’s needle is important to make us better, we believe these things because we trust the people who tell us. Much of what we know about ourselves and the world, we know because we trust whoever told us. This trust is precisely what opens the door to the possibility of deception and manipulation. If Othello had not trusted Iago, he never would have come to doubt Desdemona. The consequences of being misled are not typically as dire as they were for Othello and Desdemona, but finding out that someone I trust was wrong in something he told me—or worse, that he intentionally misled me—does lead to my trust in that person (and, as Frankfurt points out, in myself) being undermined.

Determining whether a particular violation of trust amounts to a betrayal does not differ significantly when the violation is (purely or primarily) linguistic than when it is not. Following Baier’s analysis, violations can occur as a result of negligence (failing to tell someone something that is relevant and/or important), incompetence (reporting false or insufficiently supported beliefs), or ill will (glibly revealing a secret, slandering or insulting someone, or other attempts to undermine someone’s well-being). Violating the trust of someone who has good reason to believe the trust relationship to be morally appropriate, when such a violation is reasonably avoidable, will typically qualify as a
betrayal of trust—especially when the violation involves ill-will, including willful incompetence and willful negligence.

With trust in general, the variety of goods that one can entrust to another is limitless. But with the specific sort of trust that is at issue in considerations of the ethics of speaking, including lying and deception, the range is much narrower. The question to be addressed is this: When I believe what you tell me because I trust you, what specifically do I trust you to care for? Or, put another way: When I feel betrayed that you have lied to me, what is the valued thing that is worse off—or at least that has not been cared for as I wanted it to be cared for—as a result of the lie? The answer is straightforward: my beliefs. The debate surrounding the epistemic status of testimony that we visited in Chapter 2 reveals the central issue here. The (reliable) testimony of others is a source of knowledge. Reliance on testimonial knowledge is a response to human finitude in precisely the way that trust itself is a response to human finitude. When I trust the testimony of someone who is in a better state than I am to assess the truth or falsity of the matter under consideration, I trust the care of my beliefs, of my epistemic state, to that person. At least with assertions, to trust what someone tells one is to believe it, and to believe something is to take it to be true. So if in trusting you I trust you to tell me the truth, it follows that in telling me something other than the truth—and even, possibly, in withholding the truth from me—you violate my trust.

This line of reasoning is intuitively plausible, and goes a long way toward explaining why people tend to think that lying and other forms of deception are prima facie wrong. But things are not as simple as this. To trust the care of one’s beliefs to another is not necessarily coextensive with trusting that other to tell one the truth. If
trusting involves affording discretion in the care of some valued thing, then trusting someone to be truthful involves affording that person discretion in how to convey the truth. In fact, there are times when straightforwardly telling someone the truth could actually amount to a betrayal of that person’s trust: precisely those times when doing so undermines the values that having access to the truth is supposed to contribute to.

Similarly, not only neglecting to tell the truth, but even actively misleading someone can, in some circumstances, contribute to that person’s well-being and thus be a trustworthy thing to do. An example will help to illustrate the point.

Beth was leading a group of Girl Scouts on a wilderness adventure. Even though she was properly qualified to lead the group, things got out of hand: an unexpected storm washed out all the major trails and the group wound up lost. Beth went from feeling nervous to being genuinely worried about the survival of the group. Nevertheless, she continued to reassure the Girl Scouts she was leading, never admitting to them the extent of her concern or just how lost they really were. As Lerner points out, in this situation Beth followed exactly the right approach: “Beth’s primary task was to ensure the safety of the group, not share her true feelings.” Not only would telling the girls the truth in this context have had no positive consequences, it would have undermined Beth’s and the girls’ ability to cope effectively with the situation. So in this context, misleading the girls about the severity of their predicament was the trustworthy thing to do. Compare this to what Beth said to the girls during a meeting after the adventure:

33 This example is taken from Lerner.
34 Lerner 119
Beth calmly explained that she had, indeed, felt scared, particularly because she was responsible for their safety. ... The group’s primary task now was emotional rather than functional: the job at hand involved processing the frightening experience they had been through together. Beth felt that anything short of the facts would disqualify the girls’ perceptions, invalidate their sense of reality, and impair their trust in authority.  

Once the difficult situation had passed, it was important for Beth to explain the reality of what happened so that the girls could understand both that she misled them and why she did so, in a way that did not hinder their future ability to trust either themselves, Beth, or other authorities.

An insistence on truth-telling would have ill-served Beth and the Girl Scouts she was responsible for. It is only one case, but it is sufficient to show that truth is not always the most important consideration in deciding what to say. To determine how important of a consideration it is, we will need to examine the value of truth itself. Why do we, in general, value the state of possessing true beliefs over the state of possessing false beliefs, and to what extent are we correct to do so? As Nietzsche so eloquently put it, “Suppose we want truth: why not rather untruth? and uncertainty? even ignorance?”

Harry Frankfurt recognizes the importance of this inquiry, lamenting that “few people are prepared to offer much real illumination of just what it is that makes truth so important,” and writing the short book *On Truth* to at least partially fill the perceived gap. He begins by addressing the obvious reason we value truth: its practical utility. He says,

Individuals require truths in order to negotiate their way effectively through the thicket of hazards and opportunities that all people invariably
confront in going about their lives. They need to know the truth about what to eat and what not to eat, about how to dress (given the facts concerning climatic conditions), about where to live (in view of information about such things as tectonic fault lines, the prevalence of avalanches, and the proximity of shops, jobs, and schools), as well as about how to raise their children, what to think of the people they meet, what they are capable of achieving, what they would like to achieve, and an endless variety of other mundane yet vital matters.\textsuperscript{38}

Despite professions by some post-modernists to mistrust the very notion of truth,

Frankfurt insists, the reality is that at some level we all recognize the importance of truth:

“‘No one in his right mind would rely on a builder, or submit to the care of a physician, who does not care about the truth.’\textsuperscript{39} It is possible to have a too-cavalier attitude toward the truth, and not just when engaged in academic philosophy.\textsuperscript{40}

The general pragmatic success of actions that are based on true beliefs provides an answer to the question of why we tend to value true beliefs. To the extent that my ability to engage in actions that are likely to lead to the accomplishment of my goals—indeed, even my ability to set worthwhile goals, goals that I will be happy to have achieved—depends on my having true beliefs, in trusting you with the care of my beliefs I am trusting you with the care of my own well-being, of my happiness in general. And if you betray this trust you undermine my well-being and happiness. But here a new issue arises: if what was first identified as trusting you to tell me the truth amounts to trusting

\begin{flushleft}
\textsuperscript{38} On Truth 34-35
\textsuperscript{39} On Truth 24
\textsuperscript{40} One important example of this occurred during the lead-up to the 2003 United States’ invasion of Iraq. The proliferation of highly dubious claims—such as a connection between Iraq and al Qaeda, or that Saddam Hussein was stockpiling weapons of mass destruction—can be seen as a reflection of the belief on the part of the Bush Administration that it is not truth, but rather the successful implementation of one’s own narrative, that is of importance. Such an attitude toward truth is a privilege only those in a particular position of power can afford. It is this cavalier attitude, as Scheman points out, that causes some people to sincerely and legitimately doubt the objectivity of hegemonic institutions, and ultimately to see a focus on “truth” as itself an imposition of those untrustworthy institutions. They do not see these institutions as interested in truth, but rather in profit or some other form of personal or material gain.
\end{flushleft}
you with my well-being, then what happens when there is good reason to believe that a particular truth is not or would not be conducive to my well-being? This is the concern that was raised in Chapter 1: the truth can be used as a weapon; it is no more intrinsically valuable than is trust.

This last claim needs more support. The fact that truth has an instrumental value and sometimes fails to bring about that for which it is valued does not by itself mean that truth is not also intrinsically valuable. Bernard Williams has argued that we cannot make sense of the instrumental value of truth unless we also recognize it as having an intrinsic value. He identifies Accuracy and Sincerity as the two “virtues of truth”—those qualities or dispositions that are demonstrated by people who want to know the truth, exert effort to find it out, and tell it to others—and he claims that if the virtues of truth were valuable merely instrumentally, then the general recognition of this fact would actually undermine their instrumental value:

Some people…will cheerfully leave the values of truthfulness as instrumental and accept that there is no intrinsic value to these qualities at all. They will be happy to do without it. But there is a serious question whether they can do without it. They want everyone to agree that the virtues of truth have a purely instrumental value. But what makes them think that if everyone agreed with this, the virtues of truth would still have the same instrumental value—indeed, that they would exist at all?\(^{41}\)

The tendencies to acquire correct beliefs (Accuracy) and to come out with what one believes (Sincerity) are hugely important traits for members of any society to have: they enable community members to benefit from each other’s positional advantage by engaging in a division of epistemic labor. But people are not simply recording-reporting machines; they have their own beliefs and motivations, which affect both how they

---

\(^{41}\) Williams 59
interpret their own experiences and what they decide to say. It does not take much reflection for a person to realize that, if having true beliefs is valuable, then keeping the truth from others can be valuable as well. So while it is beneficial for a society to have members who are truthful, it is not always beneficial for the individual who demonstrates said truthfulness. But once the question of whether to be truthful arises, according to Williams, the instrumental value of truthfulness is insufficient to answer the question, “Why be truthful here and now?” One who believes truthfulness to be only instrumentally valuable will be like the rule-utilitarian who finds himself in a circumstance wherein utility will be maximized by violating a generally utility-maximizing rule. Such a person has nothing to motivate him to act on the disposition except for the knowledge that the disposition is generally valuable, “and this leaves no content to the disposition: he has no thoughts with which to counter the consideration that some alternative action in this situation is the one that has the best Utilitarian consequences.”42 According to Williams, people who abide by truthfulness on rule-utilitarian grounds have no reason to be truthful when not doing so might be more valuable, and thus are not actually motivated by the virtue of truthfulness. People will only tend to tell the truth—that is, demonstrate the disposition toward truthfulness that is necessary for it to be properly considered a virtue—in such circumstances if they see truth as somehow valuable for its own sake.

Williams’s argument is thus that the epistemic division of labor on which the successful functioning of society depends is itself dependent on people’s taking truthfulness to be intrinsically and not merely instrumentally valuable, and that this will

42 Williams 91
only happen if they take truth—that at which truthfulness aims—itself to be intrinsically valuable. He then says that these considerations are sufficient to establish the intrinsic value of truth: “I suggest that it is in fact a sufficient condition for something…to have an intrinsic value that, first, it is necessary (or nearly necessary) for basic human purposes and needs that human beings should treat it as in intrinsic good; and, second, they can coherently treat it as an intrinsic good.”\textsuperscript{43}

I see two problems with Williams’s argument. The first is that a person need not take truth to be intrinsically valuable in order to answer in the affirmative the question whether to be truthful here and now, even when such an answer would be of disvalue to the speaker. I may tell the truth to my own detriment because I think the truth should just be told and that is the end of the story, but I may also do so out of compassion for another, or out of self-loathing, or because I believe the truth is instrumentally valuable and I am a value-maximizing consequentialist, or because I think that I will benefit in the long run from people believing what I say, and that telling the truth here and now will contribute to that. To think that somehow the only possible reasons one can have to tell the truth are self-interest or a belief that truth is intrinsically valuable betrays a rather impoverished view of human motivation.

The second, and more important, concern I have with Williams’s argument for the intrinsic value of truth is that, if his argument establishes anything, it is merely that truthfulness is intrinsically valuable, and not that truth itself is. The one may seem to entail the other—it would seem odd, for instance, to say that the property a compass has of pointing north is intrinsically valuable while there is nothing intrinsically valuable

\textsuperscript{43} Williams 92
about north itself. But the analogy does not work because, as Williams notes, truthfulness does not always point to the truth in the way a compass always points north. The reason Williams capitalizes “Accuracy” and “Sincerity” is to draw attention to the fact that he is using them as terms of art, and not in their everyday sense. With regard to Sincerity in particular, he makes the important point that it cannot simply be a matter of rule-following, i.e., always tell the truth, or never deceive, because there are times when it would be inappropriate, even immoral, to tell the truth. He even makes the point that when deception is warranted, it makes little if any difference whether the deception involves an outright falsehood or a misleading utterance: “[I]f deceit is justified at all…something is wrong if one thinks that it is more honourable to find some weasel words than to tell a lie.” Williams correctly sees that Sincerity, and thus truthfulness, is a matter more of trust than of truth, and that judgment plays a hugely important role in deciding what to say in order to be truthful. In this, it should be clear, I agree with him.

But, as I argued in Chapter 1, a belief that the truth is intrinsically valuable can be pernicious as easily as it can be helpful. And, as I will argue in the next two chapters, a more helpful understanding of truthfulness will take the truthful person to hold a much more sophisticated and nuanced orientation toward truth than simply taking it as intrinsically valuable.

So we are left with the idea that true beliefs are valuable because and insofar as they are conducive to well-being. And this itself means that it is not always valuable to hold only true beliefs with regard to some particular circumstance. Frankfurt says that

---

44 Of course a compass does not always point north; the presence of a stronger magnetic field will disrupt its normal functioning, but this does not undermine the disanalogy between the two.
45 Williams 107
“our chances of dealing successfully with the hazards that [the world] presents will surely be greater if we can see things straight,” but he later acknowledges that this is not always the case:

For instance, a lie may protect us in one way or another from becoming aware of certain states of affairs, when no one (including ourselves) has anything in particular to gain from our being aware of them and when our awareness of them would cause us or others serious distress. Or, a lie may divert us from embarking upon a course of action that we find tempting but that would in fact lead to our doing ourselves more harm than good.

While knowledge generally enables us to go forth confidently and competently in the world, it is possible to know too much. Being presented with too much information can sometimes be paralyzing, and false or incomplete beliefs can sometimes be more helpful than true and complete ones. Concerns about paternalism notwithstanding, sometimes those close to us do in fact know what is best for us, and that can include knowing that we will be better off if kept ignorant of some pernicious, misleading, or merely irrelevant fact. No value would have been served in Beth’s telling the Girl Scouts just how dire their predicament seemed to her to be. Or consider the example of an elderly woman with advanced Alzheimer’s Disease, who does not remember that her husband died several years ago. Whenever she asks where her husband is, if she is told that he died it is as if she is hearing the news for the first time. Rather than put her through that anguish every time she asks, those who care for her can simply say that her husband is out, and will be back in a little while. By the time the “little while” has passed, so too has her memory of asking about her husband’s whereabouts. As with the Girl Scouts, so too in

---

46 On Truth 58
47 On Truth 75-76
48 This example is taken from Kidder.
this circumstance: the best thing to do is to withhold the truth, even if that means saying something that one knows to be false.

Further, the constraints on the value of truth need not be limited to the occasional isolated fact. Nietzsche makes this point repeatedly in his writings. While Frankfurt acknowledges that the truth is not always valuable, Nietzsche argues that it may not be valuable nearly so often as we tend to think. He does not deny that the truth can be valuable, that it is worth pursuing, but he takes this value to be contingent on the truth’s ability to be life-affirming: “For all the value that the true, the truthful, the selfless may deserve, it would still be possible that a higher and more fundamental value for life might have to be ascribed to deception, selfishness, and lust.”

In *The Gay Science* Nietzsche provides an account of how a pragmatically successful worldview could be based on a series of deep metaphysical errors:

Through immense periods of time, the intellect produced nothing but errors; some of them turned out to be useful and species-preserving; those who hit upon or inherited them fought their fight for themselves and their progeny with greater luck. Such erroneous articles of faith, which were passed on by inheritance further and further, and finally almost became part of the basic endowment of the species, are for example: that there are enduring things; that there are identical things; that there are things, kinds of material, bodies; that a thing is what it appears to be; that our will is free; that what is good for me is also good in and for itself.

He makes the point that those who were more cautious in making the inferences that led to such survival-enhancing conclusions were less likely to survive, with the ultimate consequence that the false but pragmatically useful judgments came to have the status of established truth.

---

49 JGB §2
50 Nietzsche, *The Gay Science* §110
51 *The Gay Science* §111
We do not need to agree with Nietzsche’s metaphysical claims, or his understanding of Darwinism, in order to recognize the plausibility of the idea that false beliefs—even systematically false beliefs—could in some circumstances be more conducive to survival than true beliefs—and of the idea that it is often more valuable to draw a hasty conclusion from incomplete information than to remain agnostic until all the data are in. And if false beliefs can be more conducive than true beliefs to survival, it would be sheer prejudice to insist that they cannot also be more conducive than true beliefs to an individual’s or a community’s well-being even when survival is not in question.

Such considerations show, rather convincingly, I think, that it is naïve to believe that truth is always preferable to falsity, and that we are mistaken if we think that the trusted is obligated always to tell the truth. As Nyberg points out, “Trusting a person to look after your best interests is different from trusting a person never to deceive you.”

Within the context of speaking, then, we could say that an utterance is a betrayal of trust when, generally, the occasion of the utterance inspires or is based on trust (that is, it occurs within what is perceived to be a morally decent trust relationship), and the audience’s believing the utterance—or acting on the beliefs generated by the utterance—does not hinder the audience’s well-being. If lying is a betrayal of trust, then it turns out that lying has much less to do with truth (either of utterances or of beliefs) than with truthfulness, understood as a form of trustworthiness.

When I trust you, I trust you to be, either in general or at this particular moment, worthy of my trust. This does not mean that I trust you to tell me the truth, but rather that

52 Nyberg 146
I trust you to tell me something that I can safely act on (with assertions, this involves belief; with requests or commands, it involves fulfillment or obedience). What is important is that believing you will not lead me into misery or otherwise away from my ability to articulate and pursue my own goals. The specific sort of trustworthiness that I am counting on you to demonstrate can be called truthfulness. The analysis given so far reveals that there is much to be said about what truthfulness consists in, especially since it does not amount either to simply telling the truth or to saying what one believes.

In the next two chapters I will develop a notion of truthfulness as a complex social-epistemic virtue; what has been said here about it is necessarily cursory and tentative. The analysis of lying has been instrumentally valuable in getting us to a point where we can see the need for a sophisticated understanding of truthfulness. Before we move on to the virtue of truthfulness, though, it is important to consider the results of our extended exploration of lying.

**Conclusion: A Recommended Definition of “Lie”**

To summarize the key points made so far in this chapter:

- *When lying is wrong, it is so because it constitutes a betrayal of trust.*

- *Trust is betrayed when one member of a trust relationship willfully fails to care appropriately for the good or goods entrusted to that member by another.*

- *Trusting what somebody says is a matter of entrusting one’s beliefs, and thus an important factor in one’s well-being, to the care of the utterer.*

From these three points it follows that betraying another’s trust by means of an utterance is a matter of willfully putting the other’s beliefs into a condition wherein they are less likely to be conducive to that person’s overall well-being. While uttering a falsehood in a manner that is likely to lead to the falsehood’s being believed by the
audience will frequently have such a deleterious effect on another’s set of beliefs, there is no necessary connection between false belief and a detrimental epistemic state. Betrayal of trust in the context of utterances has more to do with the conduciveness of beliefs to the success or failure of subsequent actions than with the (actual or perceived) truth or falsity of an utterance.

If it is also true, as I argued in Chapter 1, that we are best served by conceiving of lying as always wrong, then we have good reason to hold that what was just said about linguistic betrayal is true, *mutatis mutandis*, of lying. A lie, then, is an utterance that betrays someone’s trust by seeking directly to have a detrimental effect on that person’s beliefs.

It is not possible to betray a trust that does not exist. So if lying is always wrong, and wrong because it is a betrayal of trust, then where there is no trust there can be no lie. This is most easily seen in relationships of active distrust, such as our familiar torture scenario. Maggie does not simply lack trust in Molly, she actively distrusts her; and Molly knows this. So there is nothing that Molly could say that could be rationally intended to, or could succeed at, betraying Maggie’s trust. This means, though, that there is nothing Molly could say that would amount to lying to Maggie. In a fundamentally distrustful relationship, lying is not possible.53 As interesting as this consequence may be, however, it is tempered by the rarity of relationships that are as openly distrustful as that of torturer and victim. It is not the case, at least for most people, that we actually

---

53 More specifically, one who is distrusted cannot lie. Maggie could still lie to Molly, though it would not be as easy as one might expect. According to my analysis, Maggie would first have to succeed at getting Molly to trust her—which, presumably, would be difficult to do if Molly knows that Maggie is her torturer. But it is certainly possible. In George Orwell’s *1984*, O’Brien manages to get Winston to trust him even when it is clear to all that he is Winston’s torturer.
distrust a significant number of the people we see or meet. Rather, as we have seen, with strangers and larger institutions, when the issue of trust arises at all it does so only in an extended sense of “trust”. Thus even with a clear and noncontroversial definition of ‘lie’ and a fully articulated understanding of what makes lying wrong, it will still often be unclear whether a person’s utterance does in fact qualify as a lie or not. This is not to say that searching for a definition of ‘lie’ or an understanding of why lying is wrong is not worthwhile, only that achieving either goal cannot do away with the need for deliberation and good judgment when evaluating the ethical status of misleading or deceptive utterances. This is why it is so important for us to shift our focus to the issue of truthfulness.
Chapter 4: The Importance of Truthfulness

Introduction

Pre-reflectively, there seems to be no difference between truthfulness and truth-telling. After all, as any dictionary will tell you, to be truthful just means to tell the truth, right? But it does not take much reflection to realize that there is some tension here, that this dictionary definition is not quite right. The initial plausibility of Siegler’s distinction between lying and telling a lie\(^1\) stems from the fact that something does seem amiss in calling truthful someone who deceives through clever word choice or intonation—was Tartuffe truthful when he confessed that he had attempted to seduce Elmire, but uttered his confession in a manner that all but guaranteed he would not be believed?

Equating truthfulness with truth-telling is a mistake because the two notions apply to different sorts of things. Whether one tells the truth with a particular utterance is a question of the straightforward meaning of the utterance.\(^2\) Whether one is truthful in making that utterance is a question of the contribution to others’ understanding the utterance makes within the context in which it occurs. The difference between truth-telling and truthfulness, then, hinges on the difference between one’s utterance and one’s contribution.

The difference between an utterance and a contribution underlies another distinction that is important to a proper understanding of truthfulness: the difference between answering a question and responding to an inquiry. To answer a question, on

---

\(^1\) See above, Chapter 1

\(^2\) Note that by ‘straightforward’ I do not mean ‘literal’. Metaphorical and other figurative meanings can be straightforward meanings as well. To say that the fence runs from the house to the barn involves a figurative use of ‘run’, but the meaning is straightforward: it is clear under what conditions such a statement would be true or false. By ‘straightforward meaning’ I mean something more along the lines of what Grice calls ‘timeless meaning’. 
my schema, is to focus primarily or even entirely on what is explicitly stated, allowing the lexical meaning to guide the form and content of one’s reply. To respond to an inquiry, on the other hand, is to place primary focus on the attitudes and intentions that motivate the posing of the question, allowing this extra information to influence how one forms and frames one’s response. The utterances that precede my uttering S, as well as other aspects of the context in which I utter it, will influence the way my utterance of S will be taken, and thus the contribution I make in uttering it. This is most easily seen in circumstances wherein I have been asked an explicit question, as the question I am answering will go a long way toward determining how my answer will be understood. But the importance of the distinction transcends straightforward question-and-answer sessions. As Gadamer points out, “[A] person who wants to understand must question what lies behind what is said. He must understand it as an answer to a question.”

Because utterances always occur within a larger context, to narrowly insist on always and only answering the question is frequently to miss the point. In order to make a truthful contribution it is often important to respond not just to the straightforward meaning of the other’s utterance, but to the contribution that utterance makes or seeks to make—to respond to the inquiry. The narrator of *The Great Typo Hunt* tells the story of himself and a friend in a conversation with an African-American shopkeeper in Atlanta about their mutual support of Barack Obama during the 2008 election season. The shopkeeper asks them why they support Obama, but it is only after the conversation has continued for a while that the narrator realizes the intent behind the shopkeeper’s question: “She wasn’t asking ‘fellow supporters’ to tell their favorite thing about Obama;
she wanted to know how we two white kids had come to vote for a black candidate.”

Misunderstandings of this sort, if they go unrecognized, can prevent all parties to a conversation from making truthful contributions, because they misunderstand each other and thus talk past each other, answering each other’s questions all the while without ever realizing that they are not responding to each other’s inquiries.

Just as a factually accurate (i.e., true) utterance may contribute to a factually inaccurate (i.e., false) understanding, so too can a factually inaccurate utterance contribute to a factually accurate understanding. Consider, for example, two roommates, Evan and Cesar, who are on their way home after a movie and decide to order a pizza. Evan calls the pizza restaurant and places the order. Soon after, they decide to change the order. Evan is driving, and since it is illegal in their state (and, Evan and Cesar both recognize, unsafe) to talk on a cell phone while driving, Evan asks Cesar to call back and change the order. Cesar calls, saying “This is Evan. I just placed an order, and would like to change it,” and proceeds to change the order according to their preference. Clearly, in claiming to be Evan, Cesar has said something that is untrue, while intending the person he is speaking with to believe what he says. But it seems just as clear that he has not done anything wrong. In fact, under the circumstances, I would argue that he has done exactly what he ought to do. Had Cesar insisted on telling the truth, he would have told the pizza worker much more information than was necessary, to the point of being unhelpful. In fact, had Cesar said, “My name is Cesar. My friend Evan just placed an order, and we want to change it,” the pizza worker would have had to call Evan back to confirm the change, which would have led either to Evan not answering his phone (and

---

4 Deck and Herson 54
the order thus not being changed), to him talking on his cell phone while driving, or to him having Cesar answer on his behalf—which would return us to the difficulty that precipitated this whole thing. So Cesar did the right thing by asserting a falsehood with the intention of bringing about a false belief in his interlocutor. And this is because doing so was the most efficacious way to bring the person he was speaking with to a correct understanding of the situation: that Evan wanted his order changed. What Cesar has said—“This is Evan”—is false, but what he has conveyed—that Evan wants his order changed—is true. Thus Cesar has made a false statement, but a truthful contribution.

Understood in this way, truthfulness is more important than truth-telling. It is the truthful utterance—and thus, ultimately, the truthful person—that is trustworthy. Thus we proceed from the pre-reflective idea that truthfulness just is truth-telling to the more sophisticated understanding of truthfulness as trustworthy telling, as trustworthiness in speaking.

The intimate connection between truthfulness and trustworthiness reveals, based on the analysis of trust provided in the previous chapter, that the gap between truthfulness and truth-telling is greater even than the discussion up to this point has indicated. Remember that one of the things that differentiates a trust relationship from a brief trusting encounter is the open-endedness, the future-orientation, of the former. In an established trust relationship, one can sometimes take a longer view of what it means to make a truthful contribution—even to the extent that one makes not just untrue utterances but even what seem to be untruthful contributions along the way.

The more trustworthy a person has demonstrated herself to be, the more latitude she is likely to be granted in using untrue utterances and instrumentally useful untruthful
contributions in the service of ultimately making a truthful contribution. In the end, truthfulness properly comes to be seen as a quality of the person making the utterance rather than of the contribution the utterance makes. A truthful—that is, trustworthy, conversationally responsible—person is one who makes contributions that can be safely trusted. And the discretion regarding the care of some valued thing that is afforded to the trusted in general is afforded to the trusted utterer in the care of my web of beliefs. This discretion can involve simply being tactful in telling the truth, or it can involve using falsity to bring about a more beneficial understanding of the truth. A primary guideline for ethical speaking, then, is this: when speaking, respond to the (present) inquiry in such a way as to maximize the chance that you will make a truthful contribution to the conversation at hand.

All of this I will argue for in greater detail in this chapter. The argument will begin with the relationship between truth and truthfulness and a limitation in Bernard Williams’s account thereof. From there we will shift to the classical Indian philosophical tradition to gain a different perspective on the ethics of truth and on the invaluable role of falsity in truthfulness. I will then elaborate more on the difference between a true utterance and a truthful contribution, and between answering a question and responding to an inquiry. I will conclude this chapter with an articulation of truthfulness as a social-epistemic virtue. In Chapter 5 I will address concerns about whether this understanding of truthfulness advocates an objectionable form of paternalism.

4.1 Truth and Truthfulness

Bernard Williams has conducted an extensive and informative study of the relationship between truth and truthfulness. I have already mentioned that he identifies
Accuracy and Sincerity as the two virtues of truth, but there is more to his account than just this. In calling Accuracy and Sincerity virtues of truth, for instance, he implicitly acknowledges that truth is more than mere correspondence, disquotation, or whatever other contemporary analytic account one may prefer. Accuracy and Sincerity are virtues of people, not of some abstract property of propositions or statements. They are, specifically, virtues demonstrated by people who value the having of true beliefs. In calling them virtues of truth, instead of, say, virtues of truthfulness, Williams is acknowledging that truth has an ethical dimension—that it is not merely an epistemic notion.

Accuracy is Williams’s blanket concept for the epistemic tendencies demonstrated by those who are properly concerned with truth. It is a necessary component of being a truthful person that one take care to have true beliefs. This involves both trusting appropriate sources of information and being sufficiently skeptical in the face of dubious claims. But it also means, and Williams does a commendable job of pointing this out, recognizing that some knowledge is not worth the effort it would take to acquire. A responsible epistemic agent must recognize when the point of sufficiently diminished marginal returns has been reached and not invest an inordinate amount of energy into an investigation. Importantly, sometimes that point is at the outset of the investigation. That is to say, some knowledge is simply not worth having, either at all or given what it would take to get it. So for a truthful speaker, the concern is not to

5 The Tuskegee syphilis experiment, conducted from 1932-1972 in Tuskegee, Alabama, is an example of what can happen when epistemic agents fail to recognize that the information to be gained is not worth the cost. In the experiment, hundreds of poor African-American men with syphilis were neither informed that they had the disease nor treated for it, in order to research the natural progression the disease would take if left untreated.
be *perfectly* accurate, but rather to be accurate *enough*, knowledgeable enough to be responsible in conveying one’s beliefs to others.

Perhaps Williams’s most significant contribution to the understanding of truthfulness, however, is the substantial amount of attention he pays to Sincerity. While Accuracy is the aspect of truthfulness that accounts for doing one’s best to have true beliefs, Sincerity is the aspect that deals with making sure what one says is trustworthy. Accuracy is about being epistemically responsible; Sincerity is about being *conversationally* responsible. Christine Swanton makes the point that people’s interactions with one another aim at something more basic than finding out the truth:

> [T]he central practical task of ethics is not *simply* the search for truth. That search is constrained by an even more fundamental problem: of our needing to live together, solving our problems in ways consistent with this end. Dialogue does not *just* serve an epistemic truth-seeking goal, it serves also the social goal of solving problems.\(^6\)

It is in his discussion of Sincerity that Williams does the most to de-couple truthfulness from mere truth-telling. First, he points out that telling the truth is not sufficient for truthfulness. Equivocation, mental reservation, and exploitation of conversational implicatures are all, he says, at best cheap and ill-conceived attempts at avoiding the telling of a truth that one wants to keep hidden. And this is true regardless of whether one’s motivation for hiding the truth is noble or malicious: “If lying is inherently an abuse of assertion, then so is deliberately exploiting the way in which one’s hearer can be expected to understand one’s choice of assertion,” and “if deceit is justified at all, as in defending the innocent fugitive, something is wrong if one thinks that it is more

---

\(^6\) Swanton 250-251, emphasis in original
honourable to find some weasel words than to tell a lie.”\(^7\) Just as there are situations wherein the effort expended to learn the truth would not be worth it, so too are there situations wherein telling the truth would not be worth it, and it is better to remain silent or even deceive one’s interlocutor—even by means of an outright falsehood. Williams concludes,

> Trustworthiness is more than the avoidance of lying, and if we want Sincerity to be the virtue of trustworthiness in speech, there must be more to be said about it. We have to ask what beliefs, and how much of one’s beliefs, one may be expected to express in a given situation.\(^8\)

What Williams says of Sincerity is insightful, but he does not go far enough in distinguishing truthfulness from truth-telling. This can be seen in the quotation above as well as in the way he poses the question that guides his inquiry into the subject: “The question is, what dispositions does a speaker need to have if he is to be trusted to say what he believes about some matter…?”\(^9\) This phrasing may seem innocuous enough, but the focus on the speaker’s saying “what he believes” is mistaken. If what is important in Sincerity is that the speaker say what she believes, then equivocation and exploitation of conversational implicature cannot be rejected as forms of truthfulness.

Williams provides a brilliant example of such exploitation:

> “Someone has been opening your mail,” she helpfully says, and you, trusting her, take it that it was not the speaker herself. If you discover that it was the speaker, you will have to agree (if through clenched teeth) that what she said was true.\(^10\)

You will also have to agree that she said what she believed and thus, on Williams’s account, that she was truthful. But absent other, extenuating, details of the story, she

---

\(^7\) Williams 107  
\(^8\) Williams 97  
\(^9\) Williams 94  
\(^10\) Williams 96
does not qualify as a trustworthy speaker in this instance. What is important is not that
the utterer can be trusted to say what she believes, but rather that the utterer can simply
be trusted—more specifically, that the audience can trust the beliefs that are generated as
a result of hearing the utterance. What is important is that the utterer make a truthful
contribution. And the example Williams gives is clearly a case wherein this
consideration is violated. A better way to put the initial question is thus, “What
dispositions does a speaker need to have if he is to be trusted in what he says about some
matter?” Removing concern with the speaker’s beliefs from the question also has the
additional benefit of allowing the question to apply to those speech acts that do not
involve directly expressing one’s beliefs (questions, commands, and promises, for
instance).

“Truthfulness,” Williams says, “implies a respect for the truth.”11 But respect
does not imply blind obedience. I can respect my parents or my teachers and still not
always follow their advice. I may think they are mistaken on this particular occasion, or I
may have good reason to believe that there is a better way to bring about the outcome
they want me to achieve. Sincerity means that what you say reveals what you take to be
ture, but it may be revealed only gradually. And the gradual process of revelation can
involve falsity along the way. Williams recognizes this to a limited extent, even making
explicit use of it in adopting Nietzsche’s strategy of using a fictional genealogy to make
sense of a real phenomenon. The State of Nature story12 he uses to explain the

---

11 Williams 11
12 The State of Nature story that Williams provides is an “imaginary genealogy” that is intended to
illuminate the virtues of truth by showing how they could have come to play the important role for us that
they do in fact play. He acknowledges that his approach is inspired by Nietzsche’s in On the Genealogy
of Morals.
importance of truthfulness to humans is, he acknowledges, a fiction. But, he insists, its
status as fiction is no objection:

As a whole, it hopes to make sense of our most basic commitments to
truth and truthfulness. Whether, if it were to succeed in doing that, it
could as a whole properly be called “true” is doubtful, but unimportant. It
is certainly less important than that the story as a whole should be
truthful.13

The importance of a story, parable, or myth lies not in its stating something true, but
rather in its revealing or conveying something true—that is, in its being truthful, making a
truthful contribution. Fictional stories can be useful educational tools, and must not
simply be dismissed as lies or propaganda. Of course, one may insist, part of a story’s
being truthful also involves its being presented as a story, and not as truth. Williams
appears to agree with this insistence. While he accepts that one should not be expected to
say everything one believes all at once regardless of circumstances or consequences, he
does not appear to accept the surreptitious use of falsehoods for bringing about an
understanding of the truth.14

The opposing view—the idea that even deception can be used in the service of
truthfulness—is explored extensively in the Indian philosophical tradition. Jonardon
Ganeri’s The Concealed Art of the Soul examines the way this approach is used
particularly in the Upaniṣads and Buddhism when it comes to teaching about the nature
of the self (ātman).15 But as a pedagogical tool it is applied much more broadly than just

---

13 Williams 19
14 Williams does accept the surreptitious use of falsehoods, though. He is perfectly okay with lying to the
would-be murderer from Kant’s famous example, for instance. But this is because he does not see bringing the
culprit around to an understanding of the truth (of the victim’s whereabouts) as a laudable goal.
15 Ganeri, The Concealed Art of the Soul: Theories of Self and Practices of Truth in Indian Ethics and
Epistemology (hereafter Concealed). Ganeri has also explored the idea in a different vein, discussing the
viability of a “procedural epistemology” that he sees as advanced by some exponents of Advaita Vedānta.
See his “Can Error Lead to Truth? The Procedural Epistemology of Mandanamśra”.

153
to such deep metaphysical issues. The tradition seems to recognize that truth, understood as literal, factual accuracy, is not intrinsically valuable, that the value of true beliefs is a product of the consequences of holding true beliefs—and that when the consequences would be bad, so is truth-telling.

We thus find in the Indian philosophical tradition an understanding of truth as at once ethical and epistemological. This complex notion of truth is conveyed by the Sanskrit word ‘satya’. Derived from the verbal root ‘as’, which means ‘to be’, ‘satya’ connotes not only a sense of truth but also of reality. Epistemological debates in the Sanskrit literature tend to focus on issues of prāmāṇya (from the verbal root pra + mā, to measure properly, and thus to know), in part because classical Indian epistemological debates focused on what makes a cognition (jñāṇa) veridical (pramāṇa) rather than on what makes a belief true (yathārtha). Satya tends to convey a more metaphysical sense, as in the four noble truths of Buddhism and Gandhi’s guiding principle of satyagraha, of holding fast to the truth.

As the references to Buddhism and to Gandhi indicate, the metaphysical connotations of satya are often also understood to carry with them an ethical aspect as well. Sitansu Chakravarti says: “Satya is truth that corresponds to facts, brute or institutional. It is as things are, out there, even irrespective of human involvement. Such

---

16 A failure to appreciate this difference leads to frequent confusions regarding classical Indian epistemological debates. The word ‘jñāṇa’ is cognate both with ‘knowledge’ and with ‘cognition’. The question of whether a particular jñāṇa is pramāṇa or apramāṇa is often translated into English as the question of whether a particular knowledge-episode is true or false. But the parties to the debates did not hold the counterintuitive notion that there can be both true knowledge and false knowledge; rather, they held the quite commonsensical notion that some of our cognitions are accurate and others are mistaken. The debates centered on how to tell whether a cognition was accurate or mistaken, and what made it so.
truths relate to ethics insofar as one is supposed to live up to them as they literally are.”  

Chaturvedi Badrinath, in his extensive commentary on the *Mahābhārata*, takes this point even further:

All schools of Indian philosophy, excepting the Materialists taken as a general group, were united in regarding truth as a great deal more than correspondence with facts. That did not imply disregard to correspondence with facts as an essential aspect of truth. But to say that it is truth that sustains and enhances human worth is not to say that correspondence with facts is all that there is to truth. Jainism and Buddhism, like Yoga and the other philosophical schools, subjected ‘correspondence with facts’ to a much deeper view of truth, *satya*.  

In the next section I will explore this ‘deeper view of truth’ and its relationship to the ethics of speaking. I will examine how we are to make sense of this robust, at times counterintuitive notion of truth, and what the implications are for the notions of lying, truth-telling, and trustworthiness. The ethical-cum-epistemological notion of truth that we find in the epic literature and elsewhere throughout the Indian tradition will provide us with some useful resources for articulating a notion of truthfulness as something far more than—and indeed sometimes in direct conflict with—simply telling the truth.

### 4.2 Truth in the *Mahābhārata*

To see what the Indian tradition has to say about the ethics of truth, we need to begin not with philosophical treatises but with the epic literature—specifically, in our case, the *Mahābhārata*. As Matilal points out, classical Indian philosophical treatises are often rather reticent about ethical matters: “The philosophical texts of classical India (*darśanas*) had...surprisingly precious little to say about moral philosophy, although they

---

17 Chakravarti 59, emphasis in original  
18 Badrinath 182, emphasis in original
had plenty to say on epistemology, logical theories and metaphysics.”\(^{19}\) It would be a mistake to infer from this, though, that ethics was unimportant to the thinkers and writers of the period:

The tradition in India was extremely self-conscious about moral conflicts and disputes about ethical principles, commitments and priorities, and moral emotions, love, passion, and self-control. This becomes crystal clear once we look at the great epics and other narrative literature. … Indeed the epics were not just heroic tales. If anything, they were also practical lessons in morals and dharma deliberations. The dharmaśāstras supply only a skeletal account of dharma. The epic stories and narrative literature add flesh and blood to this skeleton. The richness and ambiguity of the concept of dharma is interwoven with the narrative at every step.\(^{20}\)

Since our inquiry into the Indian tradition will begin with an examination of the Mahābhārata, it will be helpful to begin with some background information regarding the epic itself. One of the two great epics of Indian literature (the other is the Rāmāyaṇa), the Mahābhārata is a vast tome—comprising as many as 100,000 two-line verses in some editions. (By comparison, the Iliad and the Odyssey together comprise fewer than 30,000 lines.) The well-known Bhagavad Gītā, in which the hero Arjuna despairs at the prospect of fighting, even in a war he takes to be just, because those lined up against him are “Teachers, fathers, sons, and even grandfathers; maternal uncles, fathers-in-law, grandsons, brothers-in-law and other relatives,”\(^{21}\) is a scene from the Mahābhārata. The battle at Kurukṣetra that immediately follows the Gītā is the central event in the epic, and it pits the five sons of Pāṇdu (the Pāṇḍavas) against the one

\(^{19}\) Matilal, “Elusiveness and Ambiguity in Dharma-Ethics” (hereafter “Elusiveness”) 38.

\(^{20}\) “Elusiveness” 38-39. This approach is not unique to the Indian tradition. As Alasdair MacIntyre points out in After Virtue (especially Chapter 15), the Western tradition has a similar heritage, “according to which the telling of stories has a key part in educating us into the virtues” (216).


156
hundred sons of Dhṛtarāṣtra (the Kauravas). After eighteen days of brutal fighting the Pāṇḍavas emerge victorious, but each side has seen so many casualties that it is in many ways a pyrrhic victory. In order to win, the Pāṇḍavas have had to kill not just their cousins who were trying to usurp their right to the kingdom, but also their most beloved teachers and friends: Bhīṣma, Droṇa, Karṇa, and Śalya, each of whom took command of the Kaurava armies at different times, as well as many others. Bhīṣma, who is the great uncle of the Pāṇḍavas and the Kauravas (though in terms of their actual relations he is more like a grandfather; and in terms of age more like a great-grandfather), is recognized by all as the greatest human authority on dharma. Even though he was defeated as the initial commander of the Kaurava army, he is still alive at the end of the war, lying on a bed of arrows and awaiting the appropriate time to die. In this condition he agrees to let Yudhiṣṭhira, eldest of the Pāṇḍavas and king after the war, question him about the nature of dharma.

22 Technically, since Pāṇdu and Dhṛtarāṣtra are both descendents of Kuru, all of the cousins can properly be called Kauravas (the sons of Dhṛtarāṣtra can be more precisely called the Dhartarāṣtras), but standard practice is to use ‘Kaurava’ to refer only to the Dhartarāṣtras.

23 ‘Dharma’ is a notoriously difficult word to translate. B. K. Matilal explains, “In various contexts, the word dharma may mean: law, justice, custom, morality, ethics, religion, duty, nature, or virtue. … This almost breathtaking complexity (and, I venture to add, richness) of the dharma-concept seems to underline links between ways of living, ways of seeing and ways of relating to life’s ultimate issues” (“Elusiveness” 37). Dharma is at once an ethical, social, political, and legal notion, as these are frequently not considered separate fields in the classical Indian tradition. Chaturvedi Badrinath accordingly calls dharma “the universal foundation of life and relationships” (78).

24 Earlier in the epic Bhīṣma’s father granted him a boon that allowed him to choose the time of his death. It is worth noting that Bhīṣma is initially hesitant to give a discourse on dharma, for two reasons. First, he is in excruciating pain, and he fears the pain will serve as a distraction, causing him to speak falsely. Second, he feels that it would be inappropriate for him to discourse on dharma in the presence of Kṛṣṇa, whom he recognizes to be an incarnation of the God Viṣṇu. To speak as an authority in front of someone who is a greater authority would be, he fears, to speak out of turn. To allay these two fears, Kṛṣṇa grants Bhīṣma freedom from the pain of the arrows so that he will be able to remain clear-headed, and the divine eye with which to see all so that he will be equal in knowledge and authority to Kṛṣṇa himself. With this boon, Bhīṣma consents to answer Yudhiṣṭhira’s questions.
Yudhiṣṭhira questions Bhīṣma extensively regarding the details about the proper conduct of kings and members of other castes, about the conduct that is suitable to the various stages of life, about the proper running of a state in times of peace and times of war, and about the proper treatment of friends and enemies. Eventually, Yudhiṣṭhira’s inquiry turns to the nature of truth. He says, “Brahmins, ancestors, and Gods alike praise the dharma of truth. I want to learn about truth; tell me about this, grandfather! What is the defining characteristic of truth, oh King, and how is it reached? What results from its attainment, and why is it cherished so?”

Bhīṣma replies that truth is an eternal duty (satyaṃ dharmaḥ sanātanaḥ) and that everything relies upon truth. He then says,

In all the world, son of Bharata, truth takes thirteen forms. Undoubtedly, it is truth itself, as well as fairness and self control, unselfishness, tolerance, modesty, and benevolent patience. It is, moreover, renunciation, contemplation, nobility, constancy, unwavering steadfastness, and non-violence. These, great king, are the thirteen aspects of truth.

Bhīṣma goes on to describe the features in some detail, explaining what each consists in and how each can be achieved. He concludes by reiterating that truth establishes dharma, and that there is no greater crime than lying (nānṛtat pātakam param).

This list was clearly not compiled by someone working with a disquotational, minimalist, or even a correspondence conception of truth. With the possible exception of the first item, truth itself, this is a list of virtues. Bhīṣma, like Bernard Williams, understands truth as more than epistemological. The only way to understand Bhīṣma’s list as a list of aspects of truth is to also understand truth as an ethical consideration.

---

26 Mahābhārata (hereafter Mbh) XII.156.1-2.
27 Mbh XII.156.4
28 Mbh XII.156.7-9
29 Mbh XII.156.24
Note, in this connection, that satya is contrasted not with asatya (falsity, illusoriness) but with anṛta (that which is unlawful; a lie). Truth, as it is here described by Bhīṣma, is a property not merely of facts or propositions, and not even simply of utterances, but of the person making the utterance as well. This should not be particularly surprising, given that the discussion occurs within the context of a discourse on dharma, which is irreducibly about interpersonal relations. But Bhīṣma is not simply equating ‘truth’ with something like pleasant speech or a good disposition. What one says and does must be grounded in reality in order to qualify as true. The point is that being grounded in reality is not sufficient for an utterance to qualify as true in a robust sense; one’s attitude must also be appropriately attuned, both toward reality and toward whomever one is speaking with. In discussing the understanding of truth under which the Mahābhārata operates, Ganeri points out that, “The underlying theme is that of the mind as unsettled by emotions like fear, anger, greed or even attachment and attraction. A mind thus wracked by strong emotion will place the satisfaction of desire above the determination of truth.”

The character traits that Bhīṣma lists are valuable both for coming to know the truth and for communicating it effectively to others—they contribute to Accuracy and Sincerity, precisely the two traits that Williams calls the virtues of truth.

We get a glimpse of how this understanding of truth plays out in the Yoga Sūtras. In Chapter II, verse 30, Patañjali lists satya as one of the restraints of yoga. In the commentary Vyāsa explains what is meant by satya:

[Satya] consists in word and thought being in accord with facts. Speech and mind corresponds to what has been seen, heard and inferred as such.

---

30 Consider that ‘sat’, the adjectival form of ‘satya’, is translated most often not as ‘true’ but as ‘real’.
31 Concealed 235
Speech is uttered for the purpose of transferring one’s knowledge to another. It can only be said to have been employed for the good of others and not for their injury if it is not deceptive, confused, or barren in knowledge. If, however, it proves to be injurious to living beings even though uttered as such, it is not truth; it is a sin only. By this outward appearance, this is a facsimile of virtue, and one gets into painful darkness. Therefore, let everyone examine well and then utter truth for the benefit of all living beings.32

Here, Vyāsa agrees with the Mahābhārata account in seeing satya as opposed not to asatya but to sin (pāpa). Even if what one says is accurate, Vācaspati Miśra clarifies, if it is not spoken for the sake of benefitting others, “it is only the semblance of truth, not truth.”33 Vācaspati explains that deception, confusion, and futility are three conditions that preclude an utterance’s qualifying as true. A deceptive (vañcita) utterance, he says, is one that does not communicate the idea that is in the mind of the utterer to the mind of the audience. This would apply to strategic uses of equivocation and to mental reservation as well as to straightforward misinformation. A confused (bhrānta) utterance is one wherein the utterer is sincere but does not actually know the truth. An utterance can be futile, or ‘barren of knowledge’ (pratipattivandhya), for two reasons. The audience may not understand it, perhaps due to a language barrier or because the utterer and audience are not in close enough proximity. (I cannot claim to have told you something if I whispered it under my breath while you were across the room—or across the country.) Alternatively, the utterance may be nonsensical, as in Lewis Carroll’s famous poem “Jabberwocky”; or it may make sense but have no object, as would be the case in discussing whether the queen of the United States parts her hair on the left or right, or attempting to describe a location north of the North Pole.

32 Patañjali II.30 (p. 156)
33 Patañjali II.30 (p. 157)
This analysis is reminiscent of what Paul Grice says about what he calls the ‘Cooperative Principle’, which constrains what can qualify as an appropriate contribution to a conversation at a given point.\(^{34}\) Vācaspati Miśra’s insistence that an utterance that is deceptive, pointless, or otherwise harmful is thereby untrue amounts to a claim that such utterances violate one or more maxims of manner, according to which one should avoid obscurity, ambiguity, and verbosity. The basic idea is of striving to be helpful (Grice uses the word ‘perspicuous’) in what one says. Grice would disagree with Vyāsa and Vācaspati when they say that such unhelpful utterances are *untrue*, however. He makes the point elsewhere that an utterance can be “linguistically correct and true, even if pointless or misleading…”\(^{35}\) It is apparent that Yoga focuses more on the effect on the audience in assessing truth than does Grice. The fact that an utterance is misleading, for instance, overrides the fact that it is factually accurate: “But what I said is true!” is no defense, in the understanding of truth that we are discussing from the Indian tradition, if it was clear that what you would communicate with your utterance was false. Pataṅjali and his commentators, like Bhīṣma, are more interested in people making truthful contributions than true utterances.

One may be inclined to argue, as Grice probably would, that what is being described, both in the *Mahābhārata* and by the commentators on the *Yoga Sūtras*, is not truth but something more like sincerity, or maybe friendliness. Perhaps ‘truth’ is a poor translation of ‘*satya*’, or perhaps the word is being used in a non-standard sense that

---

\(^{34}\) The Cooperative Principle is “a rough general principle which participants will be expected (ceteris paribus) to observe, namely: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged” (Grice 26).

\(^{35}\) Grice 241 (cf. also 228).
should be acknowledged. What is true, after all, is supposed to be objective, not dependent on a person’s attitude or intentions. Virtue is not a condition of truth. The fact that what I say is hurtful to you, for instance, has no bearing whatsoever on the truth-value of what I say. Or so the argument goes. I must admit that I am a bit sympathetic toward this objection. This is why I have drawn a distinction between true utterances and *truthful* contributions, rather than *true* contributions. But to insist on this perspective is to overlook the fact that there is a notion of truth available to us that is richer than those that have been favored by analytic philosophers over the past century. To challenge the translation of ‘*satya*’ as ‘truth’ is to beg the question, assuming a very specific, impoverished notion of truth that misses the very point of what is important about truth.

In English as well as in Sanskrit, ‘true’ has many connotations. While facts and propositions can be true, so can carpenter’s squares, tuning forks, and wheels. What ‘true’ connotes in this usage is the notion of reliability, of not deviating from a norm. And this, in turn, helps us to understand what is meant when we speak of a true friend. A true friend is most definitely not someone who always and only tells the truth; a true friend is a reliable, or *trustworthy*, friend. In a sense this is unremarkable, given the etymological connection between the words ‘truth’ and ‘trust’. Both come from the same Old English root word, *tréowþ*, which comes into modern English as ‘truth,’ ‘troth,’ and ‘trust’. 36 Bromhead points out that ‘truth’ and ‘troth’ were originally useable interchangeably. By the 16th and 17th Centuries, ‘truth’ had taken on the abstract and

36 Bromhead 133. The *American Heritage Dictionary of Indo-European Roots* identifies *tréowþ* as deriving from the Proto-Indo-European root word ‘*deru*’, meaning “To be firm, solid, steadfast; hence specialized senses ‘wood,’ ‘tree,’ and derivatives referring to objects made of wood” (12). ‘*Deru*’ is also closely connected with the Proto-Indo-European word ‘*dher*’ ‘to hold firmly, support,” whence the Sanskrit word ‘*dharma*’ (14).
externalist connotation that is common today, while ‘troth’ continued to be used more to indicate “a personal act of constancy.” Whatever differences may be present in the two notions, both agree in conveying the sense of reliability, fidelity, or fitness to be trusted that was present in the original *tréowþ*, and that is present in *satya* as well. In a sense, perhaps the best English-language translation of *satya* would be ‘*tréowþ*’, if *tréowþ* were still a part of the English language.

The insight of the *Mahābhārata*, as of the *Yoga Sūtras*, is that genuine truth consists not just in the external and subject-independent notion that is captured by the modern English ‘truth’, but also in the internal and interpersonal notion that is captured by the modern English ‘troth’. In this way they seem to anticipate (by several hundred years) J. L. Austin’s insistence that

> It is essential to realize that ‘true’ and ‘false’, like ‘free’ and ‘unfree’, do not stand for anything simple at all; but only for a general dimension of being a right or proper thing to say as opposed to a wrong thing, in these circumstances, to this audience, for these purposes and with these intentions.

Austin uses the intriguing example, ‘France is hexagonal’ to make his point that both audience and intention are central to evaluating whether an utterance is true or false. Such a sentence, he says, “is good enough for a top-ranking general, perhaps, but not for a geographer.”

---

37 Bromhead 123. She quotes Geoffrey Hughes, saying, “The development of these two cognate terms has been highly significant. Whereas *troth* has remained personal, binding, and largely archaic, *truth* has become public, demonstrable by evidence and, in a sense, greater than the individual” (39).

38 We should also note here that the English ‘true’ is cognate with the German ‘*treu*’ but not ‘*wahr*’. ‘*Satya*’ seems to capture both notions at the same time.

39 Austin 145

40 Austin 143
A person can be guided by the truth not simply in believing that what one says should be true, but also in believing that one should remain true (i.e., trustworthy) to oneself and to others. Sitansu Chakravarti says that, according to the Mahābhārata, an insistence on truth-telling betrays the wrong attitude on the part of the utterer: “Following the attitude of truth one is expected to behave truthfully, and not necessarily report what has happened exactly, or carry out a promise to the letter.” An insistence on telling the truth, he says, “may be associated with egotism, self-attachment, harm to others, or a lack of sympathy and forgiveness.” This is the point that Bhīṣma’s list is meant to remind us of, and it is this that we must keep in mind if we are to properly understand what Bhīṣma—and the Yoga Sūtras—say about truth.

While a more robust notion of truth that looks to behaviors, attitudes, and other aspects of context in its determination can be appealing, it is also notoriously complex—so much so that it is easy to understand why theorists have turned to simpler, external conceptions of truth. The virtuosity of truth requires the consideration of so many variables and participants that it is often taught not as rules or principles, but rather in the context of stories and narratives. Indeed, as has already been noted, such teaching is one of the primary functions of the Mahābhārata. Accordingly, one must dig into the narrative and analysis of the great epic in order to fully unpack the complexity of the ethics of truth.

A particularly effective place in the narrative to begin this digging occurs during the fifteenth day of the war at Kurukṣetra. After Bhīṣma’s defeat, Droṇa, who taught

---

41 Chakravarti 15
42 Chakravarti 15
military strategy to the Pāṇḍavas and Kauravas, has taken over as leader of the Kaurava army. Droṇa’s fighting skill is so great that nobody can overcome him in combat, and the Pāṇḍavas know this. The Pāṇḍavas experience repeated defeats when facing the Droṇa-led Kauravas and begin to despair. In the midst of this carnage, Kṛṣṇa explains that the Pāṇḍavas will surely lose if they do not find a way to defeat Droṇa. The only way Droṇa can be defeated, Kṛṣṇa insists, is if he is somehow convinced to lay down his weapons; and this will only happen if he believes his son, Aśvatthāma, to have been killed. The problem is that Aśvatthāma, like his father, cannot be killed in battle. The solution, which Kṛṣṇa proposes, is that the Pāṇḍavas kill an elephant that happens to also be named Aśvatthāma. Bhīma then announces in the midst of battle, such that Droṇa is sure to hear it, that Aśvatthāma has been slain. It does not have the desired effect: Droṇa suspects that it is a ruse and unleashes great fury, killing thousands of the Pāṇḍava troops.

Droṇa’s wrath is so great that the gods themselves come down and criticize him for conduct unbecoming a brahmin. He begins to wonder whether what Bhīma said was true, and turns to Yudhiṣṭhīra to resolve his doubt. From anyone else he would suspect deceit; but Yudhiṣṭhīra is well-known to be perfectly pious, unflagging in his adherence to dharma (in fact he is the son of the God Dharma). Yudhiṣṭhīra does not want to mislead his beloved teacher, but he succumbs to pressure from his brother Bhīma and from Kṛṣṇa: in order to avert catastrophe and win the war, he says that Aśvatthāma is slain. In order to avoid uttering a falsehood, though, he then mutters under his breath, “the elephant.” Convinced by this act of deception, Droṇa lays aside his weapons and begins to meditate, which enables him to be killed.\footnote{The story occurs at Mbh VII.164.65-165.46.}

43
Yudhiṣṭhira’s deception of Droṇa appears to be a simultaneous act both of equivocation and of mental reservation—equivocation because ‘Aśvatthāma’ has two referents; and mental reservation because he says to himself, rather than aloud, that he is referring to the elephant. But there is more going on here. Unlike Bhīma, Yudhiṣṭhira does not use Aśvatthāma’s name. Instead, he simply says, “Lord, he is killed” (rājan hataḥ).

Ganeri suggests that the word that Yudhiṣṭhira uses for ‘elephant’, ‘kuñjara’, can also be understood as an affectionate name for Aśvatthāma and was uttered aloud rather than under his breath. Whether this is an accurate interpretation or not, the fact that Droṇa clearly intended to ask about his son’s fate means that he would infer from Yudhiṣṭhira’s utterance that his son had been killed. But this means that Yudhiṣṭhira’s deception of Droṇa trades not on equivocation, but instead on conversational implicature. (And if ‘kuñjara’ was uttered aloud, then there was no mental reservation, either.) One can understand how Yudhiṣṭhira might have preferred such an approach to equivocation, as it would have been easier to convince himself that the responsibility for the deception lay on his teacher for drawing the inference rather than on himself for making the utterance. But he cannot escape responsibility so easily, for this approach will bind him in the same difficulties that beset Kant when he attempted to make a similar move.

Further, as Williams points out, there is nothing noble in deceiving by hiding behind clever words. Arjuna appears to agree with Williams on this point, calling Yudhiṣṭhira’s deception “a lie wearing the truth as armor.” Ganeri’s assessment of the incident is, I think, the correct one: “Yudhiṣṭhira is responsible because Droṇa trusted him, and by

---

44 Mbh VII.164.106
45 Concealed 82-84
46 See above, Chapter 2.
trusting him, as it were put himself in Yudhiṣṭhira’s care. Yudhiṣṭhira failed in the duty of care his very trustworthiness created.”

Given this analysis, it appears that Yudhiṣṭhira was wrong in his decision to mislead Droṇa as he did. But things are not quite that straightforward. It is not simply a question of whether Droṇa had a right to be able to trust what Yudhiṣṭhira said, but whether he had a right to demand the information from Yudhiṣṭhira as he did. Ganeri explains:

One reason to think that he does not, after all, have a right to the truth about his son from Yudhiṣṭhira is that he is ‘trading upon’ Yudhiṣṭhira’s virtue, exploiting his reputation for truthfulness for his own ends and surely virtue is not so easily abused. Droṇa, we might say, is trying to be a ‘free-rider’ on Yudhiṣṭhira’s goodness.

If this is the case, if Droṇa is akin to a mugger asking where my wallet is, then he does not have a right to the information he asks for and Yudhiṣṭhira may be justified in deceiving him. But it is not immediately clear that the comparison is appropriate. There is another episode in the Mahābhārata that deals with one party exploiting another’s truthfulness, one wherein the comparison is more obviously appropriate. An analysis of the story of Kauśika and the bandits will help us to disambiguate Yudhiṣṭhira’s deception of Droṇa.

As Kṛṣṇa recounts the story to Arjuna, Kauśika was a brahmin who took a vow to tell only the truth, and who consequently gained renown as a truth-teller. One day some travelers who were being pursued by bandits passed by the part of the forest where Kauśika lived. When the bandits came by a little later, they appealed to Kauśika’s vow

---

47 Concealed 71
48 Concealed 84
49 See Mbh VIII.49.
and asked him to tell them where the travelers went. Kauśika told them the truth, which
enabled the bandits to find and kill their victims. Kṛṣṇa tells Arjuna that Kauśika was
reborn into a particularly heinous hell as a result of this act of cruelty.

Kṛṣṇa tells the story to help Arjuna resolve a dilemma. Arjuna had taken a vow to
kill anyone who would dare to insult his bow Gāṇḍīva, which he received as a gift from
the God Agni. But Yudhiṣṭhira, frustrated at the slow pace of the war and at what he
perceives to be Arjuna’s reluctance to fight to his fullest ability, says that the bow is
useless in his arms and he should give it to someone else. Arjuna thus sees himself as
committed, in order not to break his vow, to killing his brother. Kṛṣṇa explains that
believing this to be his only option betrays a shallow understanding of dharma, and he
tells the story of Kauśika to make his point. While there is nothing greater than truth,
Kṛṣṇa explains, truth is exceedingly difficult to understand and to practice.\textsuperscript{50} Kauśika’s
goal in taking the vow of truth was to attain heaven, and he opted to stay true to his vow
even at the expense of others’ lives in order to achieve this end. But by following the
letter of the vow so closely he undermined his own purposes. Even though he spoke the
factual truth, Kauśika transgressed the ethics of truth in at least two ways. First, the
deaths that resulted directly from his utterance meant that he failed to uphold nonviolence
(\textit{ahimsā}); and second, in his prioritization of the desire for heaven over the lives of the
travelers he failed to demonstrate unselfishness (\textit{amātsaryam}). Kauśika’s goal in taking
the vow of truth was to become virtuous enough to attain heaven, but his act of truth-
telling demonstrated a lack of the very virtue he sought. The important lesson to be
learned from this story is that Kauśika did not have the correct attitude toward the truth or

\textsuperscript{50} Mbh VIII.49.27. Bhīṣma reiterates the point when discussing \textit{satya} with Yudhiṣṭhira.
toward those with whom he spoke. For him, truth had become a fetish, which prevented him from realizing the true value of truth. If, as in this case, telling the truth will lead to violence, and if falsity will lead to nonviolence (while not violating the other conditions of truth), then falsity is tantamount to truth and truth to falsity.\textsuperscript{51}

In killing his brother, Kṛṣṇa argues, Arjuna would be acting in a manner similar to Kauśika. Kṛṣṇa provides Arjuna with an alternative that allows him to stay true to the spirit of his vow without following it to the letter: he suggests that Arjuna should disrespect Yudhiṣṭhira, as disrespect for an elder brother is tantamount to death. Even assuming this is a sufficient resolution to the problem,\textsuperscript{52} however, it is not clear that such a convenient solution had been available to Kauśika. It is tempting to say that Kauśika should have remained silent if he wanted to both maintain his vow and avoid being an accessory to murder. But, as Matilal points out, this may not be as viable of an option as it at first seems.\textsuperscript{53} For starters, the bandits may be particularly determined to find their targets and thus willing to resort to violence and torture to extract the information from Kauśika. It would then become a question of how much Kauśika is, or should be expected to be, willing to endure to avoid breaking his vow. But the problem is avoided if he is simply willing to utter a falsehood at the beginning. More important, however, is the fact that a refusal to speak leaves the bandits with the knowledge that they are ignorant of the travelers’ whereabouts and may lead them to redouble their efforts to find them. A successful deception would thus be more likely to succeed at saving the

\textsuperscript{51} See Mbh VIII.49.28
\textsuperscript{52} It certainly seems flimsy. However, if the issue is not the keeping of the vow but rather Arjuna’s ability to feel as if he has kept his vow, this may be a brilliant solution. As the story itself indicates, Arjuna is quite the moral absolutist; this could be Kṛṣṇa’s way of placating him so that he can get on with his real responsibility of helping his brothers win the war.
innocent lives than would principled silence, and would have enabled Kauśika to stay true to the value of truth.

So what does the story of Kauśika tell us about Yudhiṣṭhira’s deception of Droṇa? It is clear that the bandits intended to exploit Kauśika’s truthfulness for their benefit, and that Kauśika’s fetishization of truth enabled them to succeed in their nefarious plan. And there is a sense in which Droṇa intends to make use of Yudhiṣṭhira’s truthfulness for his benefit as well. But precisely because his plan is not to harm anyone with the information he is asking for, it is difficult to argue that his intention is blameworthy.

Whether one’s son is dead or alive seems to be information that a person (at least absent some pretty extreme extenuating circumstances) has a right to, and a trusted friend and student seems an appropriate person to turn to when one is in doubt about the issue. The bandits are at fault both for demanding the information and for what they did with it, and Kauśika is at fault for providing them with that information. Droṇa, it seems, is not at fault for demanding the information. He is simply acting out of self-defense, trying to avoid being manipulated. This then puts the responsibility for the deception squarely on Yudhiṣṭhira’s shoulders. But there is still the question of whether that responsibility can appropriately be called blame—whether it was wrong for Yudhiṣṭhira to deceive Droṇa. One’s answer to this question will depend on how one understands Yudhiṣṭhira’s motivation.

One of the main themes of the Bhagavad Gītā is to not be overly concerned with the consequences of one’s actions. Kṛṣṇa advises Arjuna at length not to let his concern about the expected outcome of fighting against his friends and family prevent him from recognizing and acting on his duty to fight. He tells Arjuna to act for the sake of dharma
rather than for the sake of consequences. It is a bit ironic, then, that Kṛṣṇa here advocates that the Pāṇḍavas deceive Droṇa (violating the dharma of truth) for the sake of the very specific consequence of winning the war.\(^54\) If Yudhiṣṭhira is understood as letting his desire for victory overpower his desire to follow dharma, then he is no less blameworthy than are his cousins who acted against dharma in pursuing the war. If, however, he is understood as acknowledging and accepting the consequences of his deception (loss of reputation, perhaps a lower rebirth) and doing it anyway because he recognizes that the only just outcome is a Pāṇḍava victory, and thus that the right thing to do is whatever it takes to win the war—just as Kṛṣṇa told Arjuna that the right thing to do was to fight in the war, irrespective of the fact that doing so would involve killing his loved ones—then his action seems to be more in keeping with dharma and with the ethics of the Mahābhārata in general.

Before we leave the story of Yudhiṣṭhira and Droṇa, there is one more point that needs making. After Droṇa asks Yudhiṣṭhira about the fate of Aśvatthāma, Kṛṣṇa assures Yudhiṣṭhira that one who utters a falsehood in order to protect life does not incur blame, and that this is just such a case because Droṇa will kill all the Pāṇḍava forces within half a day if he continues fighting. However, as soon as Yudhiṣṭhira utters the fateful sentence, his chariot that had always floated a few inches above the ground because of his piety drops to the earth. Further, at the end of the Mahābhārata Yudhiṣṭhira is made to

\(^54\) This is only one of many times that Kṛṣṇa counsels the Pāṇḍavas to apparently transgress dharma for the sake of winning the war. His reasoning is always the same: there is no way the Pāṇḍavas can overcome the forces against them straightforwardly, so for the sake of ensuring the proper outcome they should resort to some stratagem that he has devised. This is important to keep in mind when interpreting the Bhagavad Gītā, as a counterpoint to the apparently absolutist approach that Kṛṣṇa seems to advocate there. In many ways, the Kṛṣṇa of the Mahābhārata appears much more like a strict consequentialist than like the deontologist he seems to be in the Gītā.
visit hell before he can go on to heaven. When he asks why, he is told that it is necessary in order to atone for deceiving Droṇa.\textsuperscript{55} Now, since Kṛṣṇa is the incarnation of Viṣṇu, it is implausible to suppose that he was unaware that these would be the consequences of Yudhiṣṭhira’s action. It is interesting to consider whether Yudhiṣṭhira would have acted the same way if he had known what the consequences would be, as well as whether Kṛṣṇa was justified in deceiving (lying to?) Yudhiṣṭhira as he did. The answer to the final question will probably not affect one’s assessment of Yudhiṣṭhira’s action, as he clearly intended to deceive Droṇa. But if one thinks that Yudhiṣṭhira would not have consented to deceive Droṇa if he had full knowledge of the consequences for himself, then Kṛṣṇa’s deception of Yudhiṣṭhira may have been just as necessary to bring about the desired outcome as was Yudhiṣṭhira’s deception of Droṇa.

It is obvious that the desire for a particular outcome can cloud one’s judgment, leading to rationalization about what is justified in a particular circumstance. When you have a particular goal in mind, it can be difficult to tell the difference between cunning strategizing and outright cheating. This is one reason the idea that one should always tell the truth is so appealing: rationalization and self-interested deception is always a risk, and affording people discretion regarding whether to tell the truth exacerbates this risk. As the stories of Yudhiṣṭhira and Kauśika demonstrate, our own interests can play a significant role in how we address moral issues, and our assessment of others’ motivation plays an important role in how we assess their actions. The difficulty of knowing with confidence someone’s—even one’s own—motivation, and the important role that self-interest and self-deception play in motivation, will be addressed in the Conclusion of this

\textsuperscript{55} See Mbh XVIII.3.
essay. For now, let it suffice to say: when in doubt, it is generally a good policy to
default to truth-telling. But, as we will see in the next section, it can sometimes be the
case that asserting something false is the right way to bring it about that a person
understands the truth.

4.3 Truthfulness in The Upaniṣads and Buddhism

A common theme throughout the Upaniṣads is that knowledge of the nature of the
self is not to be transmitted baldly, incautiously, or indiscriminately. The standard trope
is of the sage being reluctant to speak, of doing so only when given no alternative. This
reluctance is born from an awareness that the information being requested is powerful,
even dangerous. The sages’ reticence, Ganeri says,

[is] rooted not so much in a self-serving secretiveness or a disinclination to
share the knowledge that gives them power, but...exists rather as a
response to a deep respect for the power of that knowledge, and a
recognition of the need not to be frivolous either with the knowledge itself
or with its potential recipients.\footnote{Concealed 13}

One prominent example of sagely reticence is the story of Indra and Prajāpati from the
Chāndogya Upaniṣad.\footnote{Reprinted in Radhakrishnan and Moore, 72-76.} In the story the God Indra and the demon Virocana seek out
Prajāpati to ask him about the nature of the self. Prajāpati has them look at their
reflection in a pan of water and asks them what they see. That, he says, is the true self,
and he sends them on their way. Indra soon realizes that this cannot be correct, for one’s
appearance changes radically over time but the self remains the same. He returns to
Prajāpati and demands a more satisfactory answer to his question. Prajāpati again gives
him a false answer, this time that the self that is aware during dreams is the real self.
Indra is again convinced only temporarily before returning for further explanation. The process is repeated again, but ultimately Prajāpati teaches Indra the truth, the familiar account from the Upaniṣads of ātman as the unperceived perceiver, the agent behind consciousness.

The story of Indra and Prajāpati raises a number of questions. Why is Prajāpati so hesitant to tell Indra the truth? Why does Indra trust him after being misled so many times? And what about Virocana, still under the delusion that Prajāpati’s first answer was correct? We can understand the appropriateness of Prajāpati’s treatment of Indra best by examining this last question first. At issue is what ill-fortune awaits Virocana as a result of his false view, which Prajāpati gave him, and how that could possibly be better than just telling him the truth and letting things take care of themselves. First, it is important to remember that according to the story Virocana is a demon. As such, it does not strain credulity to hypothesize that he had bad intentions in seeking out the nature of the self. Telling him the truth could well be akin to Kauśika’s mistake in answering the bandits. To insist that Prajāpati should have told the truth to Virocana would be to agree with Kant’s unsatisfactory analysis in “On a Supposed Right to Lie”. Second, and more importantly, there is a very real sense in which Virocana’s false belief is his own fault, and not Prajāpati’s. Virocana did not really engage with the answer that he was given; he accepted it and then went about his life unchanged. It is apparent that he did not let his new belief affect him, for if he had, he would have realized that it was incorrect and would have continued inquiring as Indra did. To borrow a notion from Jonardon Ganeri, Virocana was not appropriately receptive to the truth.
One of the ideas behind the Upaniṣadic sage’s reticence about the self is that someone who is not receptive to the truth thereby does not deserve the truth. Deserving the truth is a notion we explored in Chapter 2, in light of Benjamin Constant’s critique of Kant, but it now needs elaboration. Defenses of deception that trade on the idea of someone’s not deserving the truth focus almost exclusively on the example of people demanding information that they have no right to, such as the would-be murderer asking about his prospective victim’s whereabouts. And this is an important case to keep in mind. When the bandits asked Kauśika where the travelers went, they asked for information that they did not deserve. But to say that someone does not deserve something can have two very different meanings. If a friend tells me about how his wife treats him and I respond by saying, “Man, you don’t deserve that,” I can either mean that his wife treats him much better than he deserves or that she treats him much worse than he deserves. Someone can be undeserving of something pleasant or something painful, and both circumstances can provide legitimate reasons for withholding what is requested.

The argument can be made that, when Indra asked Prajāpati about the nature of the self, he asked for information that he did not deserve in the sense that he had a right not to experience the negative consequences that such knowledge would bring. It is less clear with Virocana whether he was undeserving of the truth in the positive or negative way. But perhaps it was both: he may have had no right to the truth because of the use to which he intended to put it, but also had a right not to experience the negative consequences that learning the truth would have generated.

The idea of being receptive to the truth, of engaging with what one is told rather than simply believing or disbelieving, also helps to explain why Indra would believe the
final account of the self that Prajāpati gives. He believes it not because Prajāpati says, “Okay, this time I’ll tell you the truth,” but because he considers the idea on its own merits, just as he has with the previous stories. But this one fits, at least as far as Indra can tell. He might not have been able to come up with it on his own, but with Prajāpati’s help he is able to get there.

In this way we can understand Prajāpati to be following a rather unexceptional pedagogical technique. If I want to teach students about Kant’s epistemology, a good way to do so is to begin by introducing the Rationalists. But rather than simply saying, “Here’s what Descartes and others thought and here’s why they’re wrong,” a more effective approach is to teach the Rationalists as if I agree with them wholeheartedly and let my students come on their own to recognize the limitations of the views being expressed. Even if I can articulate the objections more lucidly than my students can, they will learn more by realizing on their own that the objections are there. I can then repeat the process with the Empiricists, thereby helping my students come to a point where they will be able and ready to fully appreciate the view that Kant espoused.

In discussing the story of Indra and Prajāpati, Ganeri introduces the notion of a ‘preparatory condition’ to help explicate one aspect of what it means to engage with the truth. He says:

Among two doctrines or propositions, let us say that the first is a ‘preparatory condition’ for the second if understanding that the first is false is required of someone who is to be in a position even to speculate upon the truth or falsity of the second. Indra could not begin even to appreciate the virtues of the less obvious doctrine that the true self is the self one encounters in a dream had he not already understood as wrong the more obvious idea about the true self being the body’s reflected image or one’s appearance to others. And now Prajāpati’s reluctance might be seen as following from a further idea, that for one doctrine about the self to be a
preparatory condition of another, it must not merely be understood that it is false, but this understanding must be the result of one’s own personal investigation and discovery. If that is so, then Prajāpati could not have told Indra the final doctrine straight away, nor could he have told him that any of the preceding doctrines was false. He could only ‘feed’ Indra a preparatory doctrine and wait to see if Indra discovers it to be false. Prajāpati’s reluctance…is a reluctance born of didactic necessity when what is being taught is the nature of the self.58

Ganeri’s notion of a preparatory condition is, I think, a valuable contribution to the discussion, both about how to understand this particular story and about how to understand truthfulness in general. It gives concrete shape to the idea that one should not always strive simply to tell the truth, and that deception can be born not just out of self-interest or malevolence but also out of compassion. Indra and Prajāpati thus provide us with a vivid illustration of the reality that making an untrue utterance can be necessary in order to make a truthful contribution. The preliminary answers Prajāpati gives to Indra are designed to help him come to the point where he is ready to contemplate the nature of the reality that he seeks to understand.

This may seem like paternalism. After all, who is Prajāpati to decide whether Indra is “ready” for the truth? I will address the paternalism objection more fully in Chapter 5, but it is worth briefly addressing here. It is important to keep in mind that Prajāpati’s goal is ultimately to bring Indra around to a correct understanding. Indra wants to understand the truth, and Prajāpati answers in such a way as to maximize the chances of that happening. Of course, given the nature of Indra’s inquiry, whether he ultimately comes to that understanding is at least partly up to him—it is contingent on his doing the work to realize the limitations of the views he is initially provided. Neither the

---

58 Concealed 19
intention nor the consequence of Prajāpati’s machinations is to keep in the dark, however; the purpose is rather to enlighten efficaciously. And it is Indra’s ability to respond appropriately to the incorrect answers that demonstrates his readiness for the full and nuanced truth. So this is not just a case of one gatekeeper deciding who receives the knowledge and who is to remain ignorant; rather, it is a case of whoever is willing to work for something of value (Indra) receiving it and whoever is unwilling to do so (Virocana) not receiving it. Prajāpati did not paternalistically decide that Indra was ready; Indra demonstrated that he was, and Prajāpati responded accordingly.

Probably the most famous locus within the Indian tradition for the oblique approach to the truth that we have been discussing is not the Upaniṣads but rather Buddhism. The notion of ‘skillful means’ (upāya kauśalya, or just upāya), especially as it is used in the later (Mahāyāna) tradition, has been studied extensively.59 In short, the idea of skillful means is of the Buddha catering his teachings to the propensities of different audiences. Donald Lopez notes that “The Buddha is said to have taught different things to different people based on their interests, dispositions, capacities, and levels of intelligence.”60 As a hermeneutical device, upāya is useful because it makes possible a ready resolution to the presence in the canonical texts of teachings of the Buddha that contradict one another. Lopez makes it clear just how pressing this issue is in Buddhism:

59 See, for instance, Michael Pye’s *Skillful Means: A Concept in Mahayana Buddhism* and John W. Schroeder’s *Skillful Means: The Heart of Buddhist Compassion*. Probably the most famous canonical discussion of upāya occurs in the *Lotus Sūtra*, in particular Chapters 2 and 3, where the Buddha is compared to a father who misleads his children in order to get them to escape from a burning house. Upāya is also a common topic throughout the Prajñāparamitā literature.
60 Lopez 3
[H]ow is one to harmonize the statement “the self is one’s protector” with the statement that “there is no self”? How can the advice that suffering is to be identified, its origin abandoned, its cessation actualized, and the path to that cessation cultivated be seen as compatible with the declaration that “there is no suffering, no origin, no cessation, no path”?\(^{61}\)

Ganeri uses Candrakīrti as an example of how one is able to use upāya to explain away the Buddha’s apparently inconsistent teachings with regard to the self, in a passage that is worth quoting at length:

The Buddha’s definitive teaching about the self is that there is neither self nor non-self, or so claim the Mādhyamikas… In spite of this being his true position, the Buddha was not unwilling to assert both that there is a self and that there is no self. The assertion that there is a self is made for an audience of materialist hedonists who take it that all there is are the material elements, and who can make no sense within that metaphysics of a notion of personal moral responsibility. The Buddha supplements their beliefs with another, that there is such a thing as personal identity and so personal responsibility, and in doing so provides them with the beginnings of a world-view in which their moral scepticism ceases to make sense. Another audience does indeed have a well-entrenched conception of personal identity and moral responsibility, the problem now being that it is too well-entrenched. Their prudential self-concern is exaggerated to the point that they are incapable of attaching any value to altruistic moral motivation… To help an audience such as this, the Buddha again seeks to effect a revision in their beliefs, this time a revision by deletion rather than by supplementation. Without the belief that there is a self, but with the remaining structure of their prudential world-view in place, the implication that self-concern is the only form of moral concern ceases to be warranted. Teaching ‘no-self’ to this audience does not turn them into materialist hedonists for the simple reason that the supporting structure of their beliefs—the residue after the belief in self has been deleted—is quite different from the belief-scheme of a hedonist. Someone who responds to this teaching of no-self with a reversion to materialistic hedonism would have grasped the Buddha’s teaching in quite the wrong way.\(^{62}\)

---

\(^{61}\) Lopez 3

\(^{62}\) Concealed 113-114. The processual approach to teaching people about selflessness that is described in this passage also provides a powerful response to the common objection that selflessness would imply not an ethical ideal but an amoral nightmare. (See, for instance, Paul Williams’s *Altruism and Reality* and Tom Tillemans’s “What Would it be Like to be Selfless?”) According to this approach, the problem lies not with selflessness but rather with a premature exposure to the doctrine.
A similar approach could be used to explain mutually contradictory statements about *duḥkha*, the Eightfold Path, or any other topic. This understanding of the Buddha’s multiple and contradictory teachings about the nature of the self bears obvious resemblances to the analysis of Prajāpati given above. We can understand the Buddha as Prajāpati, the doctrine of the self taught to materialist hedonists as the teaching that Prajāpati gave Indra and Virocana, and the no-self doctrine as the teaching that Prajāpati gave to Indra after he had demonstrated his readiness. We can thus see the Buddha as, like Prajāpati, prioritizing truthful contributions over true utterances.

Of course, *upāya* as a hermeneutical tool also has the noteworthy drawback of making it exceedingly difficult to know with confidence what, if anything, the Buddha actually believed: it gives commentators the wherewithal to argue for any interpretation, on the grounds that whatever apparently contradicts that interpretation is just a case of the Buddha’s *upāya*. Fortunately, the stories to which such interpretations are frequently applied provide enough context that such explanations as that found in the above quotation are expected to be pulled from the text rather than made up out of whole cloth to support a particular commentator’s idiosyncratic interpretation. Overall, this simply means that in interpreting the Buddha’s teachings no less than in interpreting any other text, one must consider the particular passage one is dealing with in light of the larger text, and in light of the whole body of work of which it is a part.

As striking as it is that the Buddha seems to contradict himself on issues of basic doctrinal significance, far more pronounced in the Buddhist literature is the Buddha’s

---

63 Of course, the similarities are in pedagogical technique, not in content. The Buddhist and Upaniṣadic understandings of self (*ātman*) are radically different teachings.
tendency to remain silent in the face of direct and persistent questioning. One of the most famous examples of this occurs in “The Shorter Discourse to Mālunkyāputta” (Cūlamālunkya Sutta) wherein the monk Mālunkyāputta finds himself disappointed that the Buddha has not taught him about certain metaphysical issues, such as whether the world is eternal, the nature of the relationship between the soul and the body, and whether there is life after death, among other things.64 Mālunkyāputta approaches the Buddha and requests answers to these questions, claiming that he will leave the monastic life if the Buddha refuses to teach him about these things.

Instead of answering the questions that are put to him, the Buddha chides Mālunkyāputta for his willfulness. He compares one who would make such demands to someone who is shot with a poison arrow but refuses to allow the arrow to be removed until he knows who shot the arrow, where the archer is from, what kind of bow was used, what the bowstring was made of, what kind of arrowhead and feathers the arrow is fitted with, and all sorts of other irrelevant and immaterial details. The Buddha says that such a person would die without having learned any of the information demanded. He then explains that he has no intention of answering any of the questions Mālunkyāputta put to him, not because he does not know, nor because he is keeping something secret, but because there is no point to seeking out the answers to such questions. He tells Mālunkyāputta that it is noteworthy that he has refused to answer these questions: “Why have I left that undeclared? Because it is unbeneficial, it does not belong to the fundamentals of the holy life, it does not lead to disenchantment, to dispassion, to cessation, to peace, to direct knowledge, to enlightenment, to Nibbāna. That is why I

64 Majjhima Nikāya 63 (Ñāṇamoli and Bodhi 533-536)
have left it undeclared.” The Buddha’s silence can thus be understood as motivated by compassion rather than ignorance or self-interest. It can also be understood as intended to convey a greater truth (about the nature and goals of life) than can be explained in words.

In another sutta the Buddha refuses to answer the ascetic Vacchagotta when the latter asks him whether there is a self. In the face of the Buddha’s persisting silence, Vacchagotta eventually leaves. When the monk Ānanda asks the Buddha why he did not answer Vacchagotta’s question, the Buddha replies:

The teaching of no-self that I give to the [monks] is a means to guide you to look deeply in your meditation. It is not an ideology. If you make it into an ideology, you will be caught in it. I believe the ascetic Vacchagotta was looking for an ideology and not for a teaching to help him in the practice. So I remained silent. I did not want him to be caught by the teachings. If I had told him there is a self, that would not have been correct. If I told him there is no self, he would have clung to that dogmatically and made it into a theory, and that would not have been helpful either.

It is interesting to note that, while the Buddha’s motivation for not asserting that there is a self is the same as his reason for not asserting that there is no self, he does not take the two views to be interchangeable. One statement is correct and the other incorrect, but neither would be useful, given Vacchagotta’s predilections. The Buddha refuses to answer Vacchagotta because Vacchagotta has the wrong motivation. He is not appropriately receptive to the truth; he is not ready to let the truth fully affect him, and so the Buddha chooses to say nothing.

65 Majjhima Nikāya 63.8
66 Samyutta Nikāya IV.44.10 (Bodhi 1393-4)
67 Hanh 37-38. While Hanh cites this explanation in his discussion of the Buddha’s ‘thundering silence’ in the face of Vacchagotta’s questioning, it should be noted that the passage does not actually appear in the Pali text of the Samyutta Nikāya. Nevertheless, the sentiment conveyed is in keeping with what appears in the Pali version.
The Buddha’s unwillingness to answer certain questions bears a strong affinity to Vyāsa’s discussion of satya in the *Yoga Sūtras* that was discussed above. The Buddha is embodying Vyāsa’s insistence that a truth that will serve to harm those who hear it does not deserve to be spoken as truth, and that to insist on speaking it is a mere pretense of virtue. As mentioned above, Vācaspati Miśra says that an utterance can be futile because unintelligible or nonsensical. The Buddha seems, quite reasonably, to be adding ‘unbeneficial’ to the list of properties that can render an utterance futile.

The Buddha’s distaste for the idea that one should tell the truth merely because it is the truth, without regard for the effect it will have on those who hear it, is illustrated poignantly by the story of story of Ariṭṭha. The story, which comes from “The Simile of the Snake” (*Alagaddūpama Sutta*), begins with the monk Ariṭṭha expounding a non-standard interpretation of the Buddha’s teaching. Specifically, he says, “As I understand the Dhamma taught by the Blessed One, those things called obstructions by the Blessed One are not able to obstruct one who engages in them.” Unable to convince him to change his mind or renounce this interpretation, Ariṭṭha’s fellow monks bring him before the Buddha for reckoning. Now, the view that Ariṭṭha espouses flatly contradicts what the Buddha says elsewhere. Interestingly, though, the Buddha rebukes Ariṭṭha not for having a false view but for having a pernicious view: “[Y]ou, misguided man, have misrepresented us by your wrong grasp and injured yourself and stored up much demerit;
for this will lead to your harm and suffering for a long time.”\textsuperscript{71} It is not that what Ariṭṭha said was wrong (even if it was wrong), but rather that he was wrong to say it. As Ganeri says,

[Ariṭṭha’s] error was to think that all that mattered was to declare the full and nuanced truth as he understood it, without regard for the effect of that truth on himself or on the spiritual progress of his companions. In other words, he attached to the truth an unconditional value, a value independent of the well-being of those who come to believe it.\textsuperscript{72}

Like Kauśika, Ariṭṭha has fetishized the truth, and this fetishization has blinded him to the possible destructive consequences of asserting to others what he believes. This, it appears, is what the Buddha criticizes him for.

The Buddha uses the analogy of grasping a snake inappropriately to elaborate on the nature of Ariṭṭha’s error:

Suppose a man needing a snake, seeking a snake, wandering in search of a snake, saw a large snake and grasped its coils or its tail. It would turn back on him and bite his hand or his arm or one of his limbs, and because of that he would come to death or deadly suffering. Why is that? Because of his wrong grasp of the snake.\textsuperscript{73}

Just as grasping a snake inappropriately will harm the snake-seeker, so too will an inappropriate grasp of the truth harm the truth-seeker. It also runs the risk, as in Ariṭṭha’s case, of harming those who listen too readily to the professions of the misguided truth-seeker. Just as one must be cautious and skillful in attempting to grasp a snake, so too must one be cautious and skillful in attempting to grasp the truth. One must not be too hasty, or too careless, lest one end up worse off than one began. There is thus a value to be found in strategically concealing (or strategically revealing) the truth when presented

\textsuperscript{71} Majjhima Nikāya 22.6
\textsuperscript{72} Concealed 45
\textsuperscript{73} Majjhima Nikāya 22.10
with an overzealous truth-seeker. Especially within the context of Buddhism, where compassion is the pinnacle of virtue, it is important to allow the other’s receptivity to the truth to constrain the way in which one conveys the truth. My interlocutor’s mental state conditions the form that compassionate sincerity can take. Just as the candy coating on a pill makes it more palatable without hindering its effectiveness, so too can the method of delivering the truth be aimed at making the truth easier to swallow—without, thereby, undermining its status as truth. As Ganeri says, “Just as a table covered over by a cloth is still seen as a table, and a person wrapped up in shawls is still seen as a person, so too a truth ‘covered up’ is still seen as a truth. Not every guise is a disguise.”

He goes on to say, “The effectiveness of such a truth is a sign that there is a real truth underneath the wrapping, even if it is not a truth ‘in the raw’. That it works is the sign that underneath the clothing there is something true.”

4.4 Hume on Veiled Truths

While it is more fully developed in Indian philosophy, the Indian tradition does not have a monopoly on the idea that the truth can be more efficacious when veiled. We find echoes of the sentiment, for instance, within Hume’s Treatise of Human Nature. In fact, Hume’s discussion provides a ready response to one possible worry about the approach to truth that we have been discussing. This worry is that the Buddhist and Upaniṣadic approaches inappropriately prioritize tact over truth and try to paper over this fact by calling tactfulness truthfulness. The concern is that tact is at best a social virtue that has little to do with truth, and at worst a pernicious and unethical form of

---

74 Concealed 57. Notice the difference between the analysis here and Yudhiṣṭhira’s deception of Droṇa, which Arjuna called “a lie wearing the truth as armor.”

75 Concealed 58
deviousness. Hume’s point, which I think those we have been discussing from the Indian tradition would agree with, is that tactlessness is itself both harmful and unethical.

He makes this point in explaining why people prefer criticism to be given indirectly rather than explicitly. The fact that people have to follow more obscure rules of inference to conclude someone’s censure from subtle rather than explicitly condemnatory comments, Hume argues, renders the conclusion itself less forceful. He makes the point that in both cases the conclusion can be drawn with equal certainty; what is at stake is the influence this conclusion has on the person. Because of the reduced force and vivacity that subtle forms of communication lend to their conclusions, he says, we consider such veiled comments to be more respectful: “A secret intimation of anger or contempt shows that we still have some consideration for the person, and avoid the directly abusing him.”

Here one may be inclined to object that Hume is mistaken, that a secret intimation of contempt in fact betrays a lack of respect for the person criticized. Veiled criticisms and innuendo can be seen as an attempt to disguise the truth, to keep the object of such opinions in the dark as to people’s actual opinions of him or her. Such behavior when viewed in this light is understood to be manipulative, belittling, even outright insulting because it betrays the paternalistic assumption that one is either not deserving of or incapable of handling the truth. Respect is most properly demonstrated, according to this argument, not by disguising or softening the truth, but by simply speaking the truth—by being forthright and honest. One who argued in this manner against Hume would in all likelihood put the same critique to Prajāpati and the Buddha as well.

---

76 Hume, *A Treatise of Human Nature* 1.3.13
But Hume can respond to this challenge in much the same way Buddha and, I suspect, Prajāpati would: by reminding the objector to keep in mind our reasons for praising and censuring others. When we express disapproval of another, the intent is to bring about a change in the objectionable behavior. Following the analysis of the passions that Hume provides in Book II of the Treatise, we can understand criticism to be effective insofar as it brings about a sense of humility in the person criticized. Such humility can then lead to a desire to improve. But to the extent that the criticism is public, and thus stands to be embarrassing, the wrong passion is brought about. Instead of yielding a sense of humility toward oneself, such criticism yields contempt for the critic, which in turn leads to pride in oneself: exactly the opposite of the intended effect. Rather than disrespect, a concern for efficacy reveals a more fundamental respect and compassion for the other than does direct criticism. The direct criticism is a mere pretense of virtue. To be subtle, tactful, is to be responsible, allowing the object of your criticism to save face, and ultimately to respond well to the criticism.

It is important also to recognize here that Hume’s analysis does not advocate lying. This is not a question of whether or not to tell someone the truth; it is a question of how to go about conveying that truth. It is not about disguising or softening the truth per se, but about softening the impact of the truth, so that real, valuable change can be brought about. And this understanding sheds further light on the Buddha’s teachings as well. Regardless of surface grammar, the Buddha’s teachings can always be understood at least partly as exhortations to improve one’s behavior, and ultimately one’s way of understanding the world. But there is a very important sense in which much of what anyone says is aimed at helping people improve the way they understand the world. And
an unreflective insistence on truth-telling will too often fail to realize this aim.

Sometimes uttering a falsehood with the intent to deceive is the best, most efficacious, or even the only way to help one’s audience improve their understanding of the world, and with it their ability to function successfully.

4.5 Truthfulness and the Ethics of Truth

So what are we to make of truth and truthfulness in light of the foregoing discussion? I do not follow Kṛṣṇa or Vācaspati Miśra in holding that an utterance that harms another, or that is ill-motivated, is thereby untrue. It is important to be able to convey the sense of an objective standard against which claims can be measured, independently of the interests or desires of those involved. The revelation of Valerie Plame’s status as a CIA agent to the public (via reporter Robert Novak) clearly harmed many people and was probably ill-motivated, but this does not change the fact that it was true at the time of the revelation that Valerie Plame was a CIA agent. Similarly, Kauśika told the truth when he told the bandits where the travelers were, even if his motivation in doing so was selfish. Someone who reveals another’s secret tells the truth. But such a person is not truthful, because he does not deserve the credit that an attribution of truthfulness implies—he is not trustworthy, cannot be counted on. Kauśika’s mistake was to confuse the act of truth-telling for the virtue of truthfulness. Similarly, a person who simply says what other people want to hear, perhaps in order to make them happy, but without regard for the truth of her utterances, is not truthful. Such a flatterer is just as untrustworthy as the one who fetishizes truth.

But a person who says what is helpful (as opposed to what is merely pleasant) can be truthful even in the making of untrue utterances. Importantly, it is not enough that one
merely try or intend to be helpful: one’s utterances must be properly grounded in and responsive to reality—the reality both of the subject matter and of the audience. But there is a difference between being grounded in reality and directly reporting it. A helpful example from popular culture of a person being truthful in uttering an untruth occurs in the first *Star Wars* movie, when Obi-Wan Kenobi tells Luke Skywalker that Darth Vader betrayed and killed Luke’s father, Anakin. When Luke subsequently learns that Vader is his father, he confronts the spirit of Obi-Wan with this realization, asking why Obi-Wan did not tell him the truth. Obi-Wan responds by explaining that what he said was true from a particular perspective, as Anakin ceased to exist when he became Darth Vader. He further defends this explanation by claiming that many of the truths we take for granted depend on one’s point of view.

More important than whether what Obi-Wan said to Luke was *true* is whether he was truthful in saying it. Yoda seems to be in agreement with Obi-Wan, saying that it is “unfortunate” that Vader told Luke the truth when he did, as Luke was “not ready for the burden”. There is a telling contrast here between Yoda and Obi-Wan’s motivation to keep the truth from Luke and Vader’s motivation in telling him. Vader’s motivation was not simply to inform Luke of some fact that he had a right to know; it was to manipulate, to weaken Luke’s resolution to fight and to bring him over to the dark side of the force. This seems to be an example of what Vyāsa has in mind when he refers to non-compassionate truth-telling as a mere pretense of virtue. Had Luke acted as Vader wanted him to, it would have led to Luke’s great suffering and the suffering of many others. What Vader said, though true, was untrustworthy. Obi-Wan, on the other hand,
did not betray Luke’s trust—even if Luke may at first have felt otherwise. Obi-Wan remained true to Luke, and to their relationship, even in uttering an untruth.

It is a commonplace that truth has ethical implications: what is true has an effect on what I or others ought to do, or what it would be best to do. What is less readily accepted, at least in modern and contemporary mainstream Western philosophy, is that truth has an ethical *component* as well. Whatever notion of truth we adopt, it must remain broad enough to account for the fact that what Darth Vader *said* was true while also allowing for the fact that what Obi-Wan Kenobi *did* was true: it has to allow for both truth-telling and truthfulness. The aspects of truth that are concerned with factual accuracy are the ones that apply to truth-telling, while those that are concerned with trustworthiness are the ones that apply to truthfulness. Bhīṣma’s discussion of truth, the Upaniṣadic sages’ reticence about the nature of ātman, and the Buddha’s use of upāya provide models of what an understanding of truth that incorporates both elements (in these cases, under the term ‘*satya*’) would look like in practice. From here on out, our concern will be with the aspects of truth that fall under truthfulness.

Truthfulness is a species of trustworthiness. Specifically, it is trustworthiness in speech (where “speech” is broadly construed, as discussed in Chapter 1, to include such things as conspicuous silence). To be truthful entails telling the truth precisely insofar as telling the truth is a necessary condition for trustworthiness; and this condition, while frequently fulfilled, is not fulfilled nearly as often as it is commonly thought to be. Indeed, as the discussion of ‘truth’ and ‘troth’ above shows, to insist that a truthful person is concerned only with truth-telling, even at the expense of trust or trustworthiness, misses a very important part of what it has meant, historically, to say that a person is
truthful. One of the insights we must take away from the foregoing discussion is that truthfulness is not just a matter of correspondence with the utterer’s beliefs; it is a matter of the audience’s ability to safely trust. And sometimes the audience can only, or can best, trust an utterance that is in tension with the utterer’s (true) beliefs. When deciding what to say in any particular circumstance, it is important to remember that the most trustworthy thing to say may not always be what you believe, and thus that the most truthful thing to do may sometimes be to deliberately deceive your audience—just as Obi-Wan Kenobi, Prajāpati, and even the Buddha himself sometimes did.

One characteristic of trustworthy speakers is that they prioritize the making of truthful contributions over true utterances. Part of making a truthful contribution involves being responsive enough to your audience that you can respond to more than the words that are spoken: you can respond to the inquiry that motivates the question rather than simply answering the question. An example will help to make clear what I am getting at with this distinction.

Tom’s parents are divorced, and his mother re-married, giving him both a father and a step-father. Tom’s step-father, Mark, frequently refers to Tom as his son, rather than his step-son, both because that is how he thinks of Tom and because it is simply more convenient. One day Tom is walking through the mall and sees Anne, whom he knows to be one of his step-father’s acquaintances, and to whom Tom was introduced as Mark’s son. As they exchange pleasantries, Anne asks, “How has your dad been?” Tom knows that Anne means to refer to Mark, his step-father. In this situation Tom could choose to answer about his father, since that is technically who Anne asked about, but that would be misleading and unhelpful. He could explain to Anne that the man she takes
to be his father is actually his step-father, and then answer the question about the person she intended to ask about. And, generally speaking, there would be nothing wrong with him doing this. But one can also imagine, especially if Anne is an acquaintance rather than a close friend of Mark, and if both she and Tom are busy with their holiday shopping and thus not engaged in a long, detailed conversation, that taking the time to provide such a clarification would be more of a nuisance than a help. Anne may well feel put out, or imposed upon, by the more detailed answer than she expected, in much the same way she might if her question to the cashier about how his day is going were met with a long diatribe about his toothache and his girlfriend walking out on him.

Of course, Tom has a third—and, I would argue, more conversationally responsible—option: he can simply respond to Anne’s question how his father has been by telling her how his step-father has been. Not only is this response more conversationally responsible, it is truthful because it responds to the intention behind Anne’s question: it helps her to understand precisely what she sought to understand. The idea that being truthful amounts to telling the truth takes answering the question to be the primary obligation of the utterer—and thereby takes Tom’s first two options to be his only options if he wants to remain truthful. But by now it should be clear that responding to the inquiry is, in general, more important—and more truthful—than simply answering the question.

It is important to note, though, that sometimes there is no appreciable difference between answering a question and responding to an inquiry. Questions can be straightforward, motivated by nothing more than a desire to know the answer. In these circumstances, making a truthful contribution may or may not be the same as answering
the question. The bandits’ question to Kauśika was motivated by a desire to know, so responding to their inquiry would seem to simply amount to answering their question. But their desire to know was in turn motivated by the desire to commit injustice. The correct response to such an inquiry is to try to prevent accomplishment of the sought-after evil, not to just answer the question that has been put to one. Telling the truth to the bandits may have seemed like a trustworthy act to the bandits, but it involved betraying the trust of the travelers, who should have had a greater right than the bandits to trust Kauśika; telling the truth was thus untrustworthy overall, and therefore untruthful.

While responding to an inquiry rather than merely answering a question can be helpful in ensuring that one’s audience generates true rather than false beliefs, making a truthful contribution is not simply a matter of generating true beliefs in one’s audience. If it were, then Kauśika would be truthful in telling the bandits where the travelers went, and Prajāpati’s truthfulness would be dependent on Indra’s recognizing the limitations of the false views Prajāpati gives him and returning for further instruction. Obi-Wan was truthful even though Luke did not keep questioning or figure out the truth for himself; Prajāpati would have been as well even if Indra had not returned. A truthful contribution is one that generates beliefs that the audience can safely and successfully act upon. Truthfulness, once again, is a matter of trustworthiness. This orientation is clear in the Buddha’s behavior: his silence and mutually contradictory answers aim at bringing about a shift in his interlocutors’ perspectives, so that they can come to understand either what they seek to understand, or that their efforts are better spent seeking to understand something else.
Even with a perfect understanding of the context surrounding an utterance, sometimes it will still be unclear what a truthful contribution would entail. When I ask my wife what time it is, I may just be curious what time it is, or I may be concerned about being late for our dinner reservations; I may be attempting to humorously point out that she forgot to put her watch on when we left the house, or passive-aggressively pointing out how long we have spent shopping for shoes. Each of these intentions warrants a different response from her. In the former two cases, truthfulness is relatively unproblematic. If I am curious what time it is, “Don’t worry, we’re fine,” is not helpful in the way it is if I am worried about dinner reservations. In the second instance her contribution is truthful; in the first it is not (though it is not untruthful, either: it is simply irrelevant, essentially a non-response). In the latter two cases, however, it is tricky to identify what, if anything, would qualify as a truthful response on her part. For instance, if my intent is to point out her missing watch, if she responds by pulling out her cell phone and reporting to me the time her answer will be true, but she has probably missed the point of my question.77

Questions are motivated by a desire to learn something, communicate something, or both—and that ‘something’ may not be the subject-matter of the question itself. And this is true not just of questions, but of utterances in general. As I mentioned in the Introduction to this chapter, the distinction between answering a question and responding to an inquiry is of use not just when deciding how to proceed in the face of an explicit question. Utterances are frequently responses to implicit questions, and taking them to be

77 Unless, of course, her intention is not to tell me the time but rather to deny me the satisfaction of making a successful joke at her expense.
so is a necessary part of understanding and interpreting much of what anyone says. The context of the utterance helps to create or call attention to certain assumptions and preconceived notions that are implicit in the utterance itself, and which the audience will (implicitly or explicitly) account for in interpreting the utterance.

The importance of considering context when interpreting an utterance has been discussed commendably by two quite disparate twentieth-century philosophers: Hans-Georg Gadamer and Paul Grice. Grice’s work has contributed perhaps more than any other philosopher’s to an understanding of what goes into understanding both what people say and what they mean in saying what they say, at least in the context of everyday conversations. Gadamer’s discussion centers on how to understand a text, but much of what he says on the subject reveals textual understanding to be not qualitatively different from engaging in conversation. For instance, he explicitly appropriates insights from inquiries into the I-Thou relationship to illuminate what he takes to be responsible engagement with a text. Even when an utterance (including something written in a text, or the entire text itself) is not explicitly framed as the answer to a question, “the meaning of a sentence is relative to the question to which it is a reply, but that implies that its meaning necessarily exceeds what is said in it.”\(^{78}\) Understanding what is said cannot be accomplished independently of understanding the context in which it is said, but doing so necessarily means understanding the utterance as saying more than what is said. The meaning that exceeds what is said in the sentence is part of what makes up the contribution, and understanding and responding to this additional meaning is an important part of making one’s own contribution truthful.

\(^{78}\) Gadamer 370
To understand an utterance you have to, in a sense, know where the utterer is coming from. To know what to say, you have to know where your audience is coming from, so that you can anticipate their understanding of where you are coming from. This is a development of the Gricean notion that meaning involves the speaker uttering something with the intention of the utterance being understood to have a certain meaning, where that understanding itself is based in part on the audience’s recognition of the speaker’s intention. This non-vicious, hermeneutic, circularity of meaning is captured well by the notion of openness, which begins with receptivity. In his discussion of receptivity as a virtue of truth, Jonardon Ganeri says that one who seeks the truth (or truths) with the attitude of a collector, or with the thought that knowing the truth will help to win debates, is not appropriately receptive to the truth. He takes the Buddha’s diagnosis of Mālunkyaḥutta, Ariṭṭha, and Vacchagotta to be that they are not receptive to the truth. Receptivity, according to Ganeri, is a matter of letting new information influence one’s web of beliefs. He accordingly takes the Buddha’s refusal to answer Mālunkyaḥutta’s and Vacchagotta’s questions to be, like the Upaniṣadic sage’s reticence, born of compassion.

Ganeri’s discussion is enlightening, but it needs to be taken further than he takes it. His focus on the receptivity of the truth-seeker either overlooks or takes for granted how important receptivity is for the other party to the conversation. As an utterer, one must be receptive to the audience’s condition, and allow that condition to affect one’s own conception of what it would mean to be truthful, in this instance, to this particular person. This double-receptivity of utterer and audience is what Gadamer discusses under the name ‘openness’:
In human relations the important thing is...to experience the Thou truly as a Thou—i.e., not to overlook his claim but to let him really say something to us. Here is where openness belongs. But ultimately this openness does not exist only for the person who speaks; rather, anyone who listens is fundamentally open.79

Openness is an essential trait for all parties to a conversation. Partly this is because, in conversation more so than when reading a text, one is almost never exclusively utterer or audience: a conversation is a dynamic process, with utterer and audience constantly trading places to the point that the very distinction begins to lose sense. It is also because really understanding what another is saying requires coming to the conversation without the idea that one already knows all the questions as well as the answers. Gadamer almost sounds as if he is commenting on Ganeri’s discussion of receptivity when, in discussing Platonic Dialectic, he says,

To someone who engages in dialogue only to prove himself right and not to gain insight, asking questions will indeed seem easier than answering them. There is no risk that he will be unable to answer a question. In fact, however, the continual failure of the interlocutor shows that people who think they know better cannot even ask the right questions. In order to be able to ask, one must want to know, and that means knowing that one does not know.80

This, again, sheds light on the Buddha’s silence. Mālunkyāputta is not asking the right questions, because he is not aware of just how deep his ignorance and delusion go. Vacchagotta asks better questions, but the attitude he brings to his questioning is not right. (In this way he is more like Virocana than like Indra.) Ariṭṭha does not ask any questions at all, but what he says reveals that he has been bringing the wrong sort of questions to bear on what he has heard—leading him to misunderstand both what the Buddha has taught and why he has taught it in the manner he has.

79 Gadamer 361
80 Gadamer 363
A person who is not receptive to the truth will tend to ask questions that, in Gadamer’s language, are not fully open: questions that retain false, dubious, or uncritically accepted presuppositions. Such a question, Gadamer says, “pretends to an openness and susceptibility to decision that it does not have.”

To insist on truth-telling as a guide to speaking, to focus on answering the question without due consideration for the additional factors that influence and motivate one’s interlocutor, is to accept uncritically many assumptions that frequently turn out to be false. Such a doctrine amounts not to epistemic responsibility, but to conversational laziness.

It is all fine and good to assert that we must look beyond what is explicitly stated in order to understand what is meant and how to respond, but such an assertion is little more than useless unless accompanied by suggestions for how to go about doing so in a responsible manner. Grice’s work on conversation helps with this by providing a heuristic for understanding how we do go about interpreting and contributing to conversations. I have already mentioned the Cooperative Principle, which Grice takes to be operative in virtually all conversations, and which constrains both what it is appropriate to say and how any particular utterance can be appropriately interpreted.

Echoing Kant’s epistemology, Grice suggests four categories of maxims that comprise the Cooperative Principle: Quantity, Quality, Relation, and Manner. Maxims of Quantity prescribe that participants in a conversation should be informative enough without being overly informative. For instance, if I ask an attorney how long he has been practicing, I normally will be looking for an answer rounded to the nearest year or so. If he responds by saying that he has been a practicing attorney for “Less than two hundred years,” his

---

81 Gadamer 364
answer, while true, is not particularly helpful: it violates the maxim to make one’s contribution as informative as is required. But if he were instead to tell me that he has been a practicing attorney for 9,372 days, this answer would violate the maxim not to make one’s contribution more informative than necessary, and would be only slightly more helpful than the first answer.

Maxims that fall under the category of Quality pertain to the truth or falsity of one’s utterance. They amount, Grice says, to not asserting what you do not believe, and not asserting that for which you lack sufficient evidence. We will return to the category of Quality shortly, as it bears greatly on the distinction between a true utterance and a truthful contribution. Maxims of Relation bear on relevance. It would be odd, and generally inappropriate, for me to respond to your question about how many planets there are in the solar system by telling you about my first pet hamster. Maxims of Manner, as discussed above, differ from the other three types by pertaining not to what one says, but rather to how one says it. They amount to an exhortation to be helpful or perspicuous. Grice suggests as a maxim of Manner, “Facilitate in your form of expression the appropriate reply.”\textsuperscript{82} This maxim raises the important point that even if the audience is responsible for interpreting an utterance, the utterer is responsible for ensuring that the utterance is such that the audience has a reasonable chance of interpreting it correctly.

Because Gricean maxims can always be assumed to be in play in everyday conversational contexts, even overt violations of the maxims are typically informative. To blatantly violate, or flout, a maxim raises the issue of why the speaker should, for instance, be intentionally vague or uninformative, or assert a blatant falsehood. In such a

\textsuperscript{82} Grice 273
circumstance, Grice contends, the audience is generally entitled to assume that the Cooperative Principle is still in play. And here is where Grice’s now-canonical account of conversational implicature comes in. When a maxim is flouted, this calls attention to the utterer’s intention and forces a re-appraisal of the utterance in question. During an interview for a job at an office, the Hiring Manager asks me my age; I respond with, “Over 21.” My response, while true, blatantly violates a maxim of Quantity by being under-informative.\(^83\) By flouting this maxim, I have sought to convey to the Hiring Manager the impropriety of such a question. This is a conversational implicature: something that is not logically entailed by what is said (and thus not, strictly speaking, an implication), but that is strongly suggested by the way one chooses to convey what one does say. Of course I could have simply responded by stating my disapproval of the question. Flouting a maxim is by no means the only way to communicate an idea. But my evasion of the question has the benefit of being more decorous, and of being (marginally) less likely to offend the Hiring Manager.

The fact that implicated meanings can be consistently and reliably derived from utterances that do not logically imply them is what enables conversations to go forward. It is simply not possible to say everything that one means—one must invariably leave some things suggested, implied, or otherwise unsaid. It also shows that people can rely on, and can be expected to rely on, subtext in conversational interactions.

Grice raises an important concern about maxims of Quality. To the category of Quality Grice attributes a supermaxim, namely “Try to make your contribution one that is

---

\(^{83}\) Assume for the sake of the example that the job will not involve the handling of alcohol, and thus that the particular age I have chosen, 21, is not in fact relevant to the hiring process.
true,” under which he says fall two maxims, “Do not say what you believe to be false,” and “Do not say that for which you lack adequate evidence.”\textsuperscript{84} But truth is so important a factor in conversation that he wonders whether the category of Quality should be understood as a precondition for the other categories to even be a factor. Perhaps, that is, the maxims of Quality are superordinate to, rather than coordinate with, the maxims of Quantity, Relation, and Manner. He revisits this concern later, when he says that truth is not just one of a number of ways to contribute to an inquiry, but rather a condition of an utterance’s being a contribution at all: “False information is not an inferior kind of information; it just is not information.”\textsuperscript{85}

In a sense, of course, Grice is right. Generally speaking, we expect people to strive to tell the truth, not to intentionally mislead by saying what they do not believe or by overstating their confidence; and to the extent that people fail to even try to achieve this standard we take them to be negligent. This is part of why lying is so universally condemned, and also why Frankfurt takes bullshit to be such an affront to our sensibilities. But Grice takes too short a view when explicating what he takes to be the maxims of Quality. I would argue that there is just one maxim of Quality, and it is what Grice has identified as the supermaxim: “Try to make your contribution one that is true.” His phrasing is, I think, fortuitous: truth is important, but the important truth-bearer is not one’s utterance but rather one’s \textit{contribution}.\textsuperscript{86} It is important to remember, as Grice points out himself, that an utterance is typically not an isolated event—it is a part of an

\begin{footnotes}
\item[84] Grice 27
\item[85] Grice 371
\item[86] This is my own distinction, not Grice’s. His mistake comes, I think, from the failure to fully appreciate the distinction between the contribution one makes and the utterance with which one makes it—a surprising failure, since the difference between speaker’s meaning and sentence meaning is central to so much of Grice’s philosophy of language.
\end{footnotes}
ongoing conversation. And one of the things we have learned in this chapter is that one’s overall contribution to a conversation can be truthful even if one said along the way something one knew to be false. Falsity can be instrumental to making a truthful contribution. So Grice’s first maxim of Quality does not follow, and is not a necessary contributor to, the overall goal of ensuring one’s contribution to be truthful. This line of reasoning is in keeping with, but expands upon, Grice’s own insistence that things such as sarcasm and irony (which, taken literally, are false) can contribute positively—indeed truthfully—to a conversation.

Conclusion: Truthfulness as a Social-Epistemic Virtue

In discussing the difference between an utterance and a contribution I have been careful to always use the phrases ‘true utterance’ and ‘truthful contribution’ to drive home the distinction and the priority of the ethical value of one’s contribution over the truth-value of one’s utterance. But if truthfulness is a species of trustworthiness as I have been arguing, then there is a problem with calling the contribution one makes truthful rather than simply true. Calling a contribution truthful as opposed to true would be akin to calling an inanimate object trustworthy rather than trusty. We can understand what is meant, but it involves extending the meaning of the word—which in both cases is unnecessary because we already have a word that can do the job.

When understood as a species of trustworthiness, with the attendant focus on discretion such an understanding entails, truthfulness must be understood as a quality of the utterer rather than of either the utterance or the contribution. And like trustworthiness, truthfulness is typically attributed to a person not on the basis of one or two instances in which the person demonstrated the trait in question, but rather in
response to a demonstrated tendency or disposition to act in the relevant manner over an extended period of time. So, while a person can demonstrate truthfulness on a particular occasion, only those who consistently demonstrate trustworthiness in what they say can properly be called truthful people. Being a truthful person is thus much like being a liar: one lie does not a liar make, nor does one truthful act make one a truthful person.

This analysis will naturally call to mind Aristotle’s discussion of virtue (aretē). The parallel is not accidental. Truthfulness is, I think, best understood as a virtue in Aristotle’s sense. It is not my intention to get into the many discussions regarding how best to interpret Aristotle, or whether his is the best starting point for formulating a robust virtue theory. My point is simply that his discussion of virtue provides a powerful model for understanding truthfulness. Annette Baier has described virtues as “personal traits that contribute to a good climate of trust between people, when trust is taken to be acceptance of being, to some degree and in some respects, in another’s power.”87 If Baier is correct that properly practiced virtues contribute to a healthy climate of trust, truthfulness may well be the central virtue of human interaction. Certainly it is the central virtue of speaking.

Truthfulness as a virtue is a disposition to speak (or refrain from speaking) in a manner that is appropriate to the context. And this disposition is something that must be cultivated. Christine Swanton points this out when she characterizes Aristotle’s notion of aretē by saying, “virtue is a state of appropriate responsiveness to, or acknowledgement of, what I call ‘the demands of the world’, where the criteria of ‘appropriateness’ of that acknowledgement are shaped, at least in part, by a correct conception of human

87 Baier, “Trust, Suffering, and the Aesculapian Virtues” 135
flourishing." She rightly focuses on “the demands of the world” in her discussion, pointing out that a virtuous disposition involves creativity and responsiveness. Being virtuous requires the exercise of judgment in novel situations.

This is clearly the case with regard to truthfulness. Because truthfulness when understood as distinct from truth-telling requires adaptability to the particularities of the situation, there are no rules or algorithms that one can follow in order to be assured of speaking truthfully. The best we can come up with are guidelines, which are more or less appropriate at different times and for different people depending on circumstances. Further, even if there were some rule, it would not be enough that someone’s utterances conformed to that rule; to be virtuous, actions must be expressive of the right sort of attitude. Just as personal charm can be used to make a nervous person feel comfortable either out of compassion or out of a desire to seduce that person, so too can one tell the truth for either noble or self-interested reasons—only the former should be a candidate for truthfulness. The importance of being properly motivated, of doing things for the right reasons (Aristotle might say, “in accord with right reason”) means that truthfulness in practice involves a healthy measure of judgment, of responsiveness to the particular demands of the situation: truthfulness involves phronēsis.

It is also very important—central, in fact, to my account—that truthfulness as a virtue lies in a mean between an extreme of excess and one of deficiency. It need hardly be argued that people can demonstrate a deficiency with regard to truthfulness: the inveterate liar is clearly a vicious (in the sense of being characterized by vice) person. But, as Frankfurt has pointed out, the liar is not the epitome of the speaker who is

---

88 Swanton 8
deficient in truthfulness. Even more deficient than the liar is the bullshitter. Insofar as the liar wishes to be taken to be telling the truth, any lies she tells must be informed by the truth—what is said must *seem* true in order for the lie to have any significant chance of succeeding as a lie. And an attempt to make one’s speech *seem* true is just as much a truth-responsive enterprise as is an attempt to make one’s speech actually *be* true: “The liar is inescapably concerned with truth-values. In order to invent a lie at all, he must think he knows what is true. And in order to invent an effective lie, he must design his falsehood under the guidance of that truth.” The liar can be understood as a cheater, as one who respects the game enough to want to be evaluated according to the rules of the game—as evidenced by the cheater’s desire *not to be caught* cheating—even if he does not want his own actions to be constrained by those rules. But to the extent that the cheater does not want to be caught, he is constrained by the rules just as much as is the one who follows the rules: he is just constrained differently. The bullshitter, by contrast, ignores the rules entirely. Attempting to carry on a conversation with a bullshitter is akin to attempting to play a game of checkers with someone who is only interested in making interesting patterns on the board. Trying to interact in a rule-governed context with someone who refuses to acknowledge that there are rules, as opposed to someone who seeks surreptitiously to violate the rules, can be not just frustrating but downright offensive. Such a person is effectively saying that the game in question is not even worth playing. It is not the liar, but rather the bullshitter, who speaks without regard for the truth or falsity of what is said, and who is thus maximally deficient in truthfulness.

---

89 *On Bullshit* 51-52
Of equal importance, but somewhat less widely recognized, is the extreme of excess with regard to truthfulness. One well-worn example is the case of the would-be murderer at the door. I take it as uncontroversial that someone who would unhesitatingly reveal the location of the potential victim has missed the mark, and is overly concerned with truth-telling. Just as the rash person is not courageous, so too this truth-teller is not truthful. Rather, this person is guilty of a slavish obedience to the truth. The slavish truth-teller, like Kauśika in the Mahābhārata and Ariṭṭha in the Simile of the Snake, has turned truth into a fetish.

To say that truthfulness is a *social-epistemic* virtue is simply to point out that being truthful involves appropriate responsiveness to the person or people one is engaging with, including oneself (social), as much as to the independent facts that one is discussing (epistemic). Swanton says that virtuous action is responsive to the demands of the world; my point (and Swanton would, I believe, agree) is that these demands come not only from various impersonal states of affairs in the world, but also from one’s interlocutors. So whether one in fact speaks truthfully cannot be wholly dependent on the relationship between the meaning of one’s utterance and the state of the world (truth-value); it depends to a large extent as well on the relationship between the utterer and the audience.

According to Dostoyevski’s Ivan Karamazov, without God, everything is permitted. An important reservation about my account could be expressed in a way that parallels this insight. In advocating for a notion of truthfulness that is not informed by truth-telling, the concern could be that I am opening the floodgates for all sorts of manipulation. The concern, then, is that without truth to serve as ultimate arbiter,
everything is permitted. Even if truth-telling sometimes causes pain, the argument goes, it is better than the alternative; and even if we could somehow justify in theory the idea that truthfulness involves the strategic, therapeutic use of falsehoods, in actuality people cannot be trusted to do so in an ethical way: advocating such an account simply invites people to deceive themselves about their motivations in withholding information or otherwise deceiving people. In the next chapter I will respond to this concern, specifically in light of the objection that my account advocates an unacceptable form of paternalism.
Chapter 5: In Defense of Truthfulness

Introduction

In advocating an ethics of speaking that affords as much discretion to the utterer as I do, I open myself up to at least two objections. The first objection is that my account is practically incoherent. That is to say, any attempt to put it into practice would prove self-defeating because an awareness of one’s intention to deceive renders any attempt at deception futile. The idea is that in order to work, deception needs to be covert. Part of the reason Kant takes the lying promise to fail the universalizability test is that if everyone made a standard practice of making promises that they had no intention of keeping, people would lose any faith in promises and the act of promising would cease to have any meaning. Similarly, if the ethics of speaking were to advocate deception as a standard means of informing others, then the general awareness of this attitude should render impotent any attempt to misinform another. Worse, though, is that it would also render impotent any attempt to inform another. Ultimately, then, my account of truthfulness would lead to the consequence that it would no longer mean anything to speak of being truthful.

The second objection is that my account advocates a morally reprehensible form of paternalism. Following Buchanan’s seminal article, “Medical Paternalism,” I understand paternalism to describe an act of inhibiting another person’s freedom of action or freedom of choice, and intending to do so for the sake of the other’s benefit.¹ This can

¹ There is an important distinction drawn in discussions of paternalism between “soft” and “hard” (or sometimes between “weak” and “strong”) paternalism. Generally speaking, soft paternalism involves interfering with the choices of someone who is acting non-voluntarily (due to external compulsion, ignorance, or cognitive incompetence, for example). Not everyone agrees on how precisely to draw the
include physical restraint—typically to prevent the other from engaging in an action that one judges would be harmful to the one acting—or it can involve withholding relevant information or providing disinformation. The latter two are typically utilized to cause the other to make what one judges to be an optimal decision, and they are the incarnations of paternalism that my account presumably advocates.

In this chapter I will articulate and respond to both of these objections. I will then take up the issue of medical paternalism more fully in order to show both why paternalism is problematic and why my account actually discourages such behavior. If I can show how my account can be applied within the practice of medicine without amounting to paternalism, I think this will count as strong evidence that it can be similarly applied elsewhere as well.

5.1 Consenting to be Deceived

The idea behind what I call the “self-innoculation” objection is that if I know people are willing to engage in deception in their interactions with me, then their attempts at deception will be fruitless precisely because of my awareness of their intention. An overt policy of deception is thus self-innoculating. For instance, if I am listening to the Dalai Lama give a public lecture, but I know that he is perfectly willing to engage in upāya, then I will not take his statements in the same way I would take them if I took him to be engaged in a practice of truth-telling. I will listen to him instead, perhaps, as a story-teller. I may find his stories interesting; I may even repeat them to my friends. But my beliefs and actions will not change in the way they might if I took him to be telling
the truth about the world or about the human condition. Ironically, then, a willingness to be “skillful” in teaching people about the nature of the human condition actually renders the Buddhist teachings unable to teach people about the nature of the human condition.

Similar problems arise if I know that my doctor is willing to engage in such behavior. My doctor may, in classic paternalist fashion, tell me that my condition is not terminal in order to make me feel better and to make sure I do not give up hope. But if I am aware that my doctor is willing to do this, I will not feel any better upon hearing the “news”. Worse, if my condition really is not terminal, my doctor has no way to convey this to me if I know that he is willing to engage in paternalistic deception: the skeptic’s regress rears its head yet again. In short, then, the argument is this: if I don’t take you to be telling the truth, then I will not believe you; and if I don’t believe you, then your utterance cannot have the intended effect of “telling”. And if your utterance cannot amount to telling, then you cannot either inform or misinform me; your utterance is reduced to mere noise, and “truthfulness” ceases to have any sense.

There are three weaknesses in this argument. First, things are not as simple as the objection makes them out to be. As all sorts of game theory and prisoner’s dilemma scenarios demonstrate, knowledge that a person is attempting to mislead you does not serve as inoculation against being misled—especially if that person is aware that you know of the deceptive intent. Bluffing in poker games is a perfect example of this. Bluffing at poker is at least sometimes successful not because people naïvely believe everything that is said at the poker table, but because the bluffer is able to intellectually out-maneuver the target in a context wherein each knows that this is a legitimate strategy—indeed, the fact that people know not to believe everything that is said in this
context is part of what makes it bluffing instead of lying. Granted, it becomes both more difficult and more complicated to deceive if one’s deceptive intent is common knowledge. But by no means does that knowledge make deception impossible.

Second, the objection relies on an illegitimate identification of willingness with intent. Just because you are willing to deceive me does not mean that you intend to deceive me at this moment with this utterance. I have been careful in discussing my own idea of truthfulness to say that truthfulness can sometimes legitimately involve deception, not that deception can or should be practiced as a matter of course. If I know that you are willing to deceive me, perhaps because you have done so before, but you have done so only on a handful of occasions in the more than ten years I have known you, then I am not epistemically irresponsible in continuing to believe much of what you say. I have good reason to believe that any particular utterance from you is sincere, even if I know that some of your utterances may not be. If the boy who cried wolf did so a dozen times over the course of a year, and ten of those times there really was a wolf, many of the townsfolk would have continued coming to his aid even in the face of the two counterexamples.

Finally, the skeptic’s regress does not threaten here because even awareness of an intent to deceive, let alone awareness of a willingness to deceive, does not preclude the possibility of trust. People in ancient Greece knew that the Oracle at Delphi was cryptic and misleading. Such knowledge did not prevent people from going to the Oracle, seeking advice, and acting on that advice. And this despite the fact that acting on the Oracle’s advice often led to others’ downfall. How much easier must it be, then, to posit a continued willingness to trust even in the wake of deception when previous trustings
have led to success rather than failure? An example of this occurs in *The Matrix* movie trilogy. When the protagonist, Neo, consults with the character known as the Oracle to learn about his fate, he receives disappointing news. He believes what the Oracle tells him, and he acts on that belief. When he later learns that what the Oracle told him was false, Morpheus (his mentor) explains to him, “She told you exactly what you needed to hear.” This incident does not prevent Neo from seeking out the Oracle and trusting what she tells him on subsequent occasions. On the contrary: it makes her seem even more wise, and thus more trustworthy, because she was able to anticipate what he would do with the information he was given, and thus to speak so as to maximize the likelihood of a positive outcome for everyone involved.

The plausibility of Neo’s continued trust in the Oracle is confirmed by the historical fact that Buddhism did not become impotent as a result of *upāya*, even after the *Lotus Sūtra* explicitly articulated *upāya* as an educational methodology. To this day, people continue reading, listenting to, and studying Buddhist teachings (the *Lotus Sūtra* included), and taking them not just as interesting stories but as something to be believed as true. And, much as with the Oracle in *The Matrix*, the Buddha’s adeptness at catering his message to the propensities of his audience is taken to be a sign of his greater trustworthiness rather than grounds for suspicion. Similarly, people continued consulting with and trusting doctors from the time of Hippocrates up through the twentieth century, when the movement toward patient autonomy and physician honesty began to take hold. Believing that a person may mislead you does not preclude you trusting that person, regardless of whether the deception is taken to be motivated by self-interest or intended for your benefit.
It is important to remember that, since truthfulness is a species of trustworthiness more generally, deception in the name of truthfulness will be motivated by a desire to be trustworthy. When I trust you to be truthful, I trust you with the care of my beliefs, and thus with my overall well-being. If I reflect on and endorse this trust—that is, if I take ours to be a relationship of healthy trust—then I can continue to trust you even if I believe you are keeping something from me, because I trust that if it is important for me to know, then you will let me know. Of course, I can be mistaken in my choice to trust, in whom I choose to trust, or in what I choose to trust someone with. But this is true in any context, and does not pose a particular objection to my account of truthfulness.

The more you know about me, the more you can cater your actions appropriately. If you know that my spouse and I are recovering alcoholics, for instance, you may choose not to bring a bottle of wine if you come over to our home as a dinner guest. The more I trust you, the more I will let you know about me, and consequently the more you will be able to adapt your behavior in order to care for me. Such a catering of one’s behavior need not be objectionable when it is manifested in what one says any more than it needs to be objectionable when manifested in what one does. A relationship of healthy trust is a key to caring, and it is a key to truthfulness. The more distant the relationship between two people, the more presumptuous it is to attempt to care in any manner other than non-interference, and consequently the more important it is to adhere to a guideline of telling the truth. The more familiar one is with one’s audience, however, the more liberty one has in deciding how to convey the truth.

It is important to note that there are contexts in which it is not appropriate to take liberties with conveying the truth regardless of how well one knows one’s audience. It is
utterly inappropriate during legal proceedings, for instance, where truth-telling is both the letter and the spirit of the law. If I am asked about my father while I am under oath and it is not clear whether the questioner intends my biological or adoptive father, I should ask for clarification rather than simply rely on context to figure out whom the questioner intends. But the world is not a courtroom, and human interaction would be greatly impoverished if we were to act as if it were. As David Nyberg argues, “[C]ommunication absolutely free of deception of all kinds is suitable only for people who like each other very little, or don’t plan to be together very long… To live decently with one another we do not need moral purity, we need discretion—which means tact in regard to truth.”

5.2 Paternalism and Truthfulness

The second, and more important, concern about my account is that it advocates an objectionable form of paternalism. To fully articulate this objection, I am going to consider two actual cases in which behavior that I think can appropriately be labeled paternalistic could apparently find a defense in my account of truthfulness. I accept that the agents in question could use my account in defense of their acts of deception, arguing that deception was justified in their case because it helped to bring about an

---

2 A question that is worthy of further consideration is to what extent a witness should be expected to ask a lawyer to clarify a question that is asked and to what extent it is the lawyer’s job to ask the correct question to begin with. In United States v. DeZarn, for instance, the lawyer apparently intended to ask the defendant DeZarn whether there was any political fundraising at a 1990 party, but instead asked about a 1991 party. Since there was no fundraising at the 1991 party, DeZarn’s denial was accurate. Nevertheless, he was convicted of perjury—a conviction that was upheld by the Court of Appeals—on the grounds that he could reasonably be expected to understand that the lawyer intended to ask about the 1990 party. (For discussion of the case, see Green, pp. 178-180.) The more standard legal position, as expressed by the court in Bronston v. United States, is that literal falsity is necessary for perjury, and, “although a witness’ testimony might be misleading, it is the responsibility of the questioning lawyer to probe until the truth can be uncovered. If the lawyer fails to do so adequately, the witness is not guilty of perjury” (Green 177).

3 Nyberg 80
understanding of a more significant truth. But their defenses rely on a distortion of my actual account, and thereby amount to a misappropriation of my thesis.

The first case is the lead-up to the 2003 United States’ invasion of Iraq, during which George W. Bush and other members of his administration argued repeatedly that an invasion was necessary because Saddam Hussein was stockpiling “weapons of mass destruction” (WMDs) and because he was harboring high-level members of Al Qaeda in Iraq. It is now clear not only that these claims were dubious at best, but that those who advanced them had reason at the time they advanced them to doubt their veracity. One could argue, as a rebuttal to those who interpret the situation in the way suggested above,\(^4\) that perhaps they were convinced that the claims about WMDs and a connection between Iraq and Al Qaeda were true, even if the evidence they had so far obtained was insufficient to establish them.\(^5\) In this case, they could well have believed that they were using falsity in the service of truth: convincing people that X qualifies as conclusive evidence for Y, even when it does not, in order to bring about the greater good that people believe (correctly) that Y. Alternatively, they could have thought it immaterial whether Iraq in fact had WMDs or close ties with Al Qaeda, but instead thought it of vital importance that the American people support a war that the administration took to be necessary in order to stabilize a region and remove a murderous dictator from power. On this interpretation, they could have viewed the claims about WMDs and Al Qaeda as useful fictions, understanding themselves to be engaged in something akin to *upāya*.

\(^4\) Page 133

\(^5\) In *Lying and Deception* Carson suggests this as a plausible interpretation, at least of the claims regarding WMDs. He does a commendable job both of summarizing the literature available on the subject and of assessing the extent to which the various claims made by those in the Bush administration can be legitimately called lies or cases of intentional deception (pages 212-221). While I disagree with the definition of ‘lie’ with which he operates, I largely agree with his overall analysis of the events.
Assuming this interpretation is accurate, those in the Bush Administration (including the President himself) who perpetrated the deception with regard to weapons of mass destruction and the supposed link between Iraq and Al Qaeda were untruthful in their utterances because they failed to ensure that they knew what they were talking about. They conveyed a greater confidence than was warranted in the veracity of their evidentiary claims as well as in the status of those claims as evidence for their conclusions. They may have been guilty of self-deception or of wishful thinking. Williams claims that one of the greatest values of the virtue of Accuracy is precisely as a corrective on the tendency to believe what we want to believe instead of what the evidence indicates we ought to believe. It is also difficult to understand the false claims advanced by members of the Bush Administration as akin to upāya, as the motivations appear to be opposed to such an interpretation. There is no evidence that the information provided was catered to help people come to a sophisticated understanding of the truth. Rather, the stories told appear to be intended to advance the particular agenda of those telling the stories by preventing people from understanding the truth—a truth that the people had a right to (providing the information would not have amounted to a breach of national security, for instance); and a truth that, were it known, would not have had a detrimental effect on anyone other than those who were trying to keep it obscured. This is not virtuous truthfulness; it is vicious, self-interested deception.

The second example is of a doctor who diagnoses his patient with thyroid cancer and recommends a thyroidectomy, but does not inform the patient that vocal cord

---

6 I am not claiming that this *is* an accurate assessment of their motivations. I am using this case merely as a thought experiment.
paralysis is a possible side effect of the surgery. He does not tell her about the possibility because he suspects that she would opt for a less effective treatment option due to the fear of losing her voice. The physician could argue that keeping the truth of the risk from his patient helped her to understand the more significant truth that a thyroidectomy was the best course of action, that he withheld the information to save her from unnecessary doubt and confusion.⁷

The primary mistake this doctor made, it seems to me, is that he failed to be appropriately receptive to his patient’s own point of view. There is no evidence from the case that the physician knew the patient well enough to presume to know either that the patient would have opted for a treatment with a lower chance of success, or that such a decision would have been an irrational one (or unsuccessful treatment a suboptimal outcome) for this patient. The doctor considered what he took to be the patient’s medical good without giving due consideration to the other components of the patient’s overall good. Further, it is possible that the physician exploited the level of authority his recommendation would have just by virtue of his status as a medical doctor in order to cause the patient to make the decision he deemed to be ideal. Either way, this behavior is not virtuous truthfulness; it is paternalistic manipulation.

The doctor who convinces a patient to undergo a procedure without disclosing the possible side effects is, absent extenuating circumstances, acting paternalistically. So is the politician who falsifies evidence to drum up support for a political aim. And I cannot deny that my account risks making it easier to rationalize such behavior. Be that as it

---

⁷ This example is modified from a malpractice case that Buchanan cites. In the case, the physician’s testimony as to why he chose not to inform his patient of the risks reveals a blatantly paternalistic attitude: “I feel that were I to point out all the complications—or even half the complications—many people would refuse to have anything done, and therefore would be much worse off” (Buchanan 373-4).
may, my account does not justify paternalistic behavior. People who want to manipulate others for their own benefit will find a way to rationalize their actions whether they are familiar with my account of truthfulness or not. As Benjamin Franklin pithily says in his autobiography, “So convenient a thing it is to be a reasonable creature, since it enables one to find or make a reason for everything one has a mind to do.” The fact that my account can be misappropriated for such purposes no more counts as an argument against its correctness when understood properly than does the fact that ephedrine can be used to make methamphetamine count as an argument against its usefulness as a decongestant.

Just as the ease of availability of methamphetamine makes education about its dangers that much more imperative, so too does the ease with which my account could be understood to countenance paternalism require both an extra caution against paternalism and a clear articulation of just how my account should be put to use. To that end, in the next section I will dig deeper into the issue of medical paternalism, particularly as it applies to the doctrine of informed consent, to show how my account of truthfulness can actually discourage paternalistic attitudes and decision-making.

5.3 Paternalism and Truthfulness in Medical Ethics

Historically, paternalism of the kind under consideration has been both common and widely accepted in the practice of medicine, at least in the West. In their book Principles of Biomedical Ethics, Beauchamp and Childress point out that the concern for a physician’s truthfulness in patient interactions is a rather recent development:

---

8 Franklin 32
9 My research into the area for this essay has only included the practice of medicine in the tradition spawning from Hippocrates. I make no claims about the historical practice of medicine in other cultural traditions. Such research will have to wait for a subsequent project.
Codes of medical ethics have traditionally ignored obligations and virtues of veracity. The Hippocratic Oath does not recommend veracity, nor does the Declaration of Geneva of the World Medical Association. The Principles of Medical Ethics of the American Medical Association (AMA) from its origins until 1980 made no mention of an obligation or virtue of veracity, giving physicians unrestricted discretion about what to divulge to patients. The 1980 revision recommended, without elaboration, that physicians “deal honestly with patients and colleagues,” and the 2001 revision indicates that physicians shall “be honest in all professional interactions.”

But what does, and what should, such honesty entail? According to Beauchamp and Childress, “Veracity in the health care setting refers to comprehensive, accurate, and objective transmission of information, as well as to the way the professional fosters the patient’s or subject’s understanding.” I will return to this characterization later, as I will argue that the final clause provides a crucial constraint on the rest of the sentence.

Despite the shift in recent decades toward greater patient participation in medical treatment and decision-making, the paternalism model is still prevalent in a number of cultures. For instance, Tolga Guven cites “the only currently available Turkish clinical ethics textbook,” which was published in 2001, as evidence that paternalism is still accepted in Turkey. The text includes a quotation from the author, a professor of surgery, recommending a paternalistic attitude toward cancer patients:

Some want to know the answer and ask: “What do you think it might be?”; whereas others may ask candidly: “I have cancer, don’t I?” However, such patient[s] actually want to hear that they do not have cancer. Even in this situation, the answer: “yes, you have cancer” should not be given.

This passage aptly illustrates the risks that inhere in following my advice to respond to someone’s inquiry rather than simply answering the question asked. It is easy for me to

---

10 Beauchamp and Childress 288
11 Beauchamp and Childress 289
12 Guven 2
convince myself illegitimately not only that I know what is best for you, but even that I know better than you do what you want. As a result, I risk inappropriately giving you the answer to a question that is not the question you intended to ask.\textsuperscript{13}

Allen Buchanan describes what he takes to be the three most common arguments at the heart of defenses of medical paternalism, and he argues that all three fail to support the practice. The first is the Prevention of Harm argument. The idea behind this argument is that the physician’s duty is to prevent or minimize harm to the patient, and thus that the physician is justified in withholding information when providing that information would cause more harm than would withholding it. This explains the historical tendency of physicians to withhold cancer diagnoses: “They assumed that if they told the patient he had cancer they would be depriving him of all hope and that the loss of hope would result in suicidal depression or at least in a serious worsening of the patient’s condition.”\textsuperscript{14} Buchanan points out multiple problems with this argument. For one thing, the paternalistic physicians’ cost-benefit analyses almost uniformly fail to take into consideration the harm that is generated by the patient’s finding out that she has been deceived by her physician. Such a realization can fundamentally undermine the patient’s trust in her physician, making subsequent treatment difficult or even impossible.

But even assuming the physician could somehow guarantee that the deception would go undetected, or that the patient would understand and continue trusting, the Prevention of Harm argument is, Buchanan contends, fatally flawed. There is precious little evidence for the common concern that diagnosing a patient with a terminal illness

\textsuperscript{13} Note, though, that part of the reason I advocate responding to the inquiry instead of merely answering the question is that answering the question can have this consequence as well, because sometimes the question one asks is not the question that one intended to ask or that one takes oneself to have asked.

\textsuperscript{14} Buchanan 372-3
actually leads to depression or suicidal tendencies. Physicians’ arguments in favor of such paternalism tend to rely on vague intuitions or anecdotes rather than on empirical studies. More importantly, the risk of suicide is a psychiatric assessment that physicians in general are unqualified to make. (Even psychiatric specialists are seldom able to predict with confidence whether learning some particular piece of information would make a person suicidal). Further, the idea that suicide is not a rational choice for the (terminally ill) patient is itself uncritically assumed by the paternalistic physician.\(^\text{15}\)

The Prevention of Harm argument fails because the physician cannot know with sufficient confidence that providing the information will be harmful to the patient at all, let alone more harmful than withholding or misleading. Physicians who use the Prevention of Harm argument to support keeping their patients in the dark seem to have their own idea of what the patient’s good consists in, and impose their own views onto the situation. But as a rule, a physician is simply not in a better position than is the patient to decide what is good for the patient from the patient’s perspective. Arguments in favor of paternalism that rely on the Prevention of Harm argument fail to appreciate this.

The second argument for paternalism that Buchanan criticizes is what he calls the Contractual Version of the Prevention of Harm argument. This holds that the nature of the patient-physician relationship is such that the physician is contractually obligated to withhold information when doing so would serve the greater good of the patient, and the nature of the contract implies that the patient consents (at least implicitly) to such a stipulation. This argument, Buchanan points out, is subject to all the same criticisms as

\(^{15}\) For more, see Buchanan, pages 377-383.
the straightforward Prevention of Harm argument. He argues that this argument suffers from an additional flaw, however, in that not all patients take themselves to be consenting to such a contract when they seek medical help, and the law supports them.\textsuperscript{16} At most, then, he argues, a physician can assume it is okay to mislead a patient if that patient explicitly authorizes such behavior. But this cannot work, for “if in each case he inquires as to whether the patient wishes to make such an authorization, he will defeat the purpose of the authorization by undermining the patient’s trust.”\textsuperscript{17} As I argued in the previous section, this consequence does not follow. Nevertheless, the standard arguments against the Prevention of Harm argument are, I think, sufficient to undermine the Contractual Version as well.

The third argument for paternalism that Buchanan seeks to refute is the Inability to Understand argument. The idea here is that there is no duty to provide patients or their families with information that they will be unable to understand or to process, and that it is better to withhold such information than to confuse the patient. This argument is more plausible than the previous two, because it really is the case that many patients either do not or cannot understand much of what a medical practitioner understands about the practice of medicine. And it is reasonable to posit that too much information, in health care decision-making no less than elsewhere in life, can be as problematic as too little information. Indeed, as Beauchamp and Childress indicate, there is some empirical evidence that this is the case:

Some studies have uncovered difficulties in processing information about risks, indicating that risk disclosures commonly lead subjects to distort

\textsuperscript{16} Buchanan 384
\textsuperscript{17} Buchanan 384
information and promote inferential errors and disproportionate fears of some risks. Some ways of framing information are so misleading that both health professionals and patients regularly misconstrue the content.\textsuperscript{18}

Nevertheless, Buchanan argues that these considerations do not justify a physician in withholding information from a patient. It is not the physician’s obligation, he points out, to ensure perfect understanding; the physician’s obligation is rather “to make a reasonable effort to be understood.”\textsuperscript{19} He does not elaborate on what qualifies as a reasonable effort, but presumably part of that effort would include being aware of and accounting for such framing effects as Beauchamp and Childress point out. Precisely what a reasonable effort to be understood involves or should involve is another point to which we shall return.

The loss of trust that many patients experience when they find they have been misled, under-informed, or excluded from the decision-making process by health care professionals illustrates one of the most pernicious aspects of medical paternalism. I made the point above that a relationship of healthy trust is an important part of successful caring, and this is abundantly true in the field of health care. Because of the essential role that trust plays in any successful physician-patient relationship, which we will discuss in more detail below, any practice that tends to undermine trust in this context must be avoided.

Recognition of the problems inherent in a paternalistic approach to patient care over the last (roughly) half-century has led to a marked shift toward a focus on patient autonomy, at least in the United States. One of the most pronounced ways in which this focus on patient autonomy manifests is in the importance of informed consent. It is now

\textsuperscript{18} Beauchamp and Childress 130
\textsuperscript{19} Buchanan 386
recognized that patients have a right to be informed of the nature of their condition (including prognosis), their treatment options, and the risks and benefits associated with those options, as well as a right to authorize or refuse medical procedures. A focus on informed consent and patient autonomy has resulted in a “shift in the locus of decision making from the physician to the patient.”

Because of the importance of informed consent, the demands of respect for patient autonomy are frequently understood to call for truth-telling. The first rule that Beauchamp and Childress identify as justified by a focus on respect for autonomy is “Tell the truth.” Ironically, however, an insistence on truth-telling in the name of patient autonomy can have the effect of undermining patient autonomy—and with it, patient well-being. Consider, for instance, the challenges faced by a physician treating someone who subscribes to the traditional Navajo belief that thought and language do not merely describe reality but actually have the power to shape and to alter it. The example is mentioned by Beauchamp and Childress, and is discussed more thoroughly by Ruth Macklin. Macklin relays the story of an elder who refused to undergo a bypass after the surgeon explained to him the (very small) risk of not waking up from general anaesthesia. To the patient, mentioning the possibility that he might not wake up from surgery was tantamount to guaranteeing that he would not wake up from surgery. This reality raises a dilemma for western-trained doctors working with Navajo populations:

Should they adhere to the ethical and legal standards pertaining to informed consent now in force in the rest of the United States and risk harming their patients by "talking in a negative way"? Or should they adhere to the Navajo belief system with the aim of avoiding harm to the

---

20 Pellegrino and Thomasma 11
21 Beauchamp and Childress 104
22 Beauchamp and Childress 106
patients but at the same time violating the ethical requirement of disclosure to patients of potential risks and future contingencies.\textsuperscript{23} Macklin argues, and I agree, that in such a situation the health care professional must clearly give priority to the patient’s beliefs and withhold the information if there is no way to present it in a way that would qualify as talking in a positive way.\textsuperscript{24} But a straightforward rule of informed consent, insisting that anything less undermines patient autonomy, cannot sufficiently account for this. Even within the field of medicine, an insistence on full disclosure, on telling the truth, can be as paternalistic as can withholding information. If I insist on telling someone the truth regardless of other considerations, then I am guilty of assuming uncritically that I know what is best for my audience, and that the best thing for my audience is to be told the truth. A surgeon who insists on disclosing all risks to a patient who subscribes to the traditional Navajo view can be understood as paternalistically assuming that her own worldview or value system is superior to her patient’s.

This issue is not limited to those who work with Navajo patients. The example is indicative of a larger concern with the relationship between truth-telling and patient autonomy. “To respect autonomous agents,” Beauchamp and Childress say, “is to acknowledge their right to hold views, to make choices, and to take actions based on their personal values and beliefs.”\textsuperscript{25} Thus, to withhold information from people disrespects their autonomy insofar as it undermines their ability to do these things. But telling people

\textsuperscript{23} Macklin 10  
\textsuperscript{24} I do not know enough about the Navajo worldview to know whether, for instance, mentioning the percent chance of success rather than of failure (or of waking up from general anaesthesia rather than of not waking up) would succeed at respecting the injunction against speaking negatively.  
\textsuperscript{25} Beauchamp and Childress 103
something that undermines their ability to do these things is also to fail to respect their autonomy. As Guven points out,

Paternalistic physicians believe that they already know what patients do or do not want... When health care professionals assume (without probing the expectations of individual patients) that the truth must be disclosed to patients under all conditions, they are performing a paternalistic act and disrespecting the autonomy of their patients because they alone determine what patients should know. Forcing the truth on patients who do not wish to receive such information is, in fact, not compatible with the implications of respect for patient autonomy. Such an act is no less problematic than withholding the truth from patients who demand to be given detailed information about their disease.\(^{26}\)

It should come as no surprise that I would advocate for replacing the rule, “Tell the truth” with the guideline, “Be truthful.” What is important is not that physicians always tell the full and unadulterated truth to their patients, but rather that they be trustworthy in what they say. When treating a Navajo patient, that may (depending on the patient) involve leaving out some of the risks of the proposed course of treatment. Below, we will consider additional cases wherein truth-telling and trustworthy telling (that is, truthfulness) are in conflict, and I will show why trustworthiness is more important. First, though, it is worth exploring the importance of trust to the physician-patient relationship.

Beauchamp and Childress state unequivocally the importance of trust in health care: “Trustworthiness has the practical outcome of promoting the image of the professional and making health care effective. Nothing is more important in health care organizations than the maintenance of a culture of trust.”\(^{27}\) As I mentioned in Chapter 3, the patient must trust the physician enough to share any potentially medically relevant

\(^{26}\) Guven 163
\(^{27}\) Beauchamp and Childress 41
information. This is just one of the reasons it is so important to guarantee doctor-patient confidentiality. And what can qualify as medically relevant is much broader than may at first be apparent. Pellegrino and Thomasma point out that the patient must be forthcoming not only about whatever condition or ailment has brought him to seek the physician’s assistance, but also about “his values and obligations that will influence his choices of alternatives. The last is necessary so that the physician may act as the patient’s advocate and also understand where their values might differ.”

If a committed Jehovah’s Witness does not disclose her religious values to her surgeon, for instance, the surgeon may well end up giving her a blood transfusion, unaware that such treatment is contrary to her wishes. The health care provider’s ability to care properly is limited by the patient’s willingness to share fully, which in turn is limited by the patient’s willingness to trust.

So, following the analysis of trust provided in Chapter 3, the obvious question is: what is the good that the patient entrusts to the care of the health care professional? It may seem that the obvious question has an obvious answer: health. They are, after all, health care practitioners. But it is not quite as simple as this. If the only concern were the care of the patient’s health, then paternalism would seem a legitimate approach, as the physician is likely to know much more than the patient about the nature of the condition or ailment and any medically indicated treatment.

A much more promising approach is that taken by Pellegrino in his article “Professing Medicine, Virtue Based Ethics, and the Retrieval of Professionalism”. He

---

28 Pellegrino and Thomasma 107
argues that the end or goal of medicine is the good of the patient, broadly construed.\(^\text{29}\)

He points out that one’s medical good is only one of many considerations that contribute to one’s overall good. It is not terribly difficult to imagine cases in which there is some degree of conflict between a patient’s medical good and that patient’s more comprehensive notion of her own good, such as a dancer who feels that it would be better to be dead than to have a foot amputated—even if it may be difficult for many people, the physician included, to agree with the dancer’s sentiment. Less difficult to understand may be the Jehovah’s witness who would rather die than transgress the will of God by receiving a blood transfusion. Pellegrino argues that four auxiliary goods contribute to the overall good of the patient, and that the physician should be cognizant of each. They are the patient’s medical, personal, human, and spiritual goods. I do not intend here to evaluate Pellegrino’s list of types of goods: to do so would take us too far afield. The approach is more important to me than the particular list, and the approach is, I think, the correct one. The virtues of health care professionals are those traits that will characteristically bring about the proper end of medicine, which is the overall good of the patient. And to bring about this end, one must take into consideration not only the patient’s state of health, but also the patient’s values and priorities. One must take account of what the patient takes to be her own good. The patient’s legitimate trust in the physician obligates the physician to care for what the patient cares about.

Now, to the extent that the patient’s conception of her own good lies beyond what may be considered her medical good, the physician’s obligation to care for it typically

\(^{29}\) This works, he says, for the other professions (but not necessarily for, say, trades) as well: the end of law is the good of the client, of education the good of the student, etc.
extends no further than respect and non-interference. It would be as inappropriate and presumptuous for a physician who is a committed atheist to try to disabuse her patient of his considered religious beliefs as it would be to disregard those beliefs. Just as I trust other library patrons and my fellow train-passengers not to bother trying to care for certain aspects of my well-being, so too do I trust my doctor’s care to involve simply leaving me alone in certain respects.

The focus on the physician’s ability to care appropriately for the patient is central to the beneficence model of health care that Pellegrino and Thomasma advocate as an alternative to paternalism. Unlike paternalism, the beneficence model recognizes the importance of patient autonomy and strives to ensure that autonomy as a value informs the broader ideal of respect and care for persons. While disagreeing with some of the details of Pellegrino and Thomasma’s account, Beauchamp and Childress advocate for something that is broadly similar. Beauchamp and Childress identify caring as the central virtue for health care professionals, and they point out that what caring entails is quite complex: “Considerations of trust, compassion, objective assessment, caring responsiveness, reassurance, and the like must be balanced.” They identify compassion, discernment, trustworthiness, integrity, and conscientiousness as five “focal virtues” that contribute to the overall virtue of caring.

It is noteworthy how similar this list of virtues is to the list of aspects of truth that Bhīṣma gives in the *Mahābhārata*, which we discussed in the previous chapter. Later in

---

30 In fact, Pellegrino and Thomasma see beneficence as a sort of middle ground between the extremes of medical paternalism and an absolute focus on autonomy. They understand a pure focus on autonomy to be mistaken because, “To accord autonomy superiority over beneficence is sometimes to abandon the patient in time of need” (6).
31 Beauchamp and Childress 22
32 Beauchamp and Childress 38
the chapter, we identified truth as having aspects that are concerned with trustworthiness as well as those that are concerned with factual accuracy. Within the field of health care, the intellectual virtues aim at discovering the truth while the moral virtues aim at fidelity or trustworthiness. The importance of truth and of trustworthiness indicates why truthfulness is such an important virtue for health care professionals: to be (both professionally and morally) responsible, a health care professional must strive both to understand the truth and to convey it to her patients. She must, that is, demonstrate both accuracy and sincerity. She must also, however, be responsive to the particular needs, interests, and values of each individual patient: she must demonstrate receptivity.

One concern about a narrow focus on autonomy is that it can seem to obviate the need for receptivity. If I can fall back on the rule that I should tell the patient everything and then let the patient decide what to do, I no longer have to struggle as much with the difficulties of figuring out what would be the right thing to do for this patient, in this circumstance. I can rest easy believing that I did my job, without worrying about what would qualify as really caring for this person, here and now. Michael Fetters laments this consequence of the autonomy approach: “In the United States, physicians often turn to principles of autonomy to buttress their decisions—for example, ‘Well, it's their right to know’ or ‘The patient has to know; it's the patient's decision.’” 33 The issue is particularly salient for Fetters, an American physician who treats a high proportion of Japanese patients. That many of his patients are Japanese is relevant because of the prevalence of the view within the Japanese culture that, in cases of terminal illness, the physician should first inform the patient’s family rather than the patient himself, so that the family

33 Fetters, Section 9
can decide what the best course of action is and whether to disclose the diagnosis to the patient. In order to be responsive to his patients’ individual interests, Fetters uses what he calls an “advance directive for disclosure” to find out both how much information his patients want and how involved they want to be in decision-making. Kipnis argues for a similar approach, suggesting that, before beginning either testing or treatment, physicians initiate a conversation with their patients with something like the following inquiry:

    Here at _____ hospital we have many types of patient. Some want to know all about their illnesses, even if these are serious, and they want to make treatment choices for themselves. Other patients would just as soon let their loved ones receive all the information and handle all the decisions. We don’t want to make a mistake in the way we treat you. Can you help us to understand how involved you want to be in your care? If the patient’s response to this question is recorded in the chart, then the health care professionals who collaborate in the treatment of this patient are better equipped to care appropriately for the patient as an individual.

    The physician’s moral responsibility to respect patient autonomy, then, is to strive not for full disclosure, but rather for adequate disclosure. And, as should by now be clear, what qualifies as adequate will vary according to the patient:

    At issue is the extent to which a standard should be tailored to the individual patient—that is, made subjective. The subjective standard requires the physician to disclose the information a particular patient needs to know, if it is reasonable to expect the physician to know that patient’s informational needs.

    The caveat at the end of the quotation is important. Such an approach works only in contexts wherein it is reasonable to expect the physician to know (and the physician does

---

34 For more, see the series of articles discussing the role of family in informed consent decisions in Japan and the United States in *The American Journal of Bioethics*, 6 (1) 2006.
35 Kipnis, “Taking Families Seriously Enough” 22
36 Beauchamp and Childress 123
actually know, or have good reason to believe that she knows) the patient’s (non-medical) needs and interests. When the physician has inadequate knowledge of this sort, what is proposed as a subjective standard becomes no standard at all: it reverts to the paternalism of “whatever the physician thinks is best.” But where the physician does have an adequate understanding of the patient’s needs and interests, it is the physician’s responsibility to care for those to the best of her ability. A subjective standard is not the same as an arbitrary standard: it is a standard that is responsive to the subject of treatment. And such responsiveness is anathema to the rote rule-following that truth-telling implies.

Adapting one’s message, and the means of delivering one’s message, to the particularities of the intended audience can be done in a responsible manner, provided the utterer has a sufficient understanding of the audience. And refusing to cater one’s message to the audience, or refusing to understand one’s audience enough for such catering to be done responsibly, is itself morally suspect. The key when it comes to withholding information from a patient is to not presume that the patient would not want to learn some particular detail, but to let the patient’s own expressed values guide the nature and amount of the information that one reveals to the patient as well as the manner in which one reveals it. Remember that Buchanan points out that a physician’s responsibility is to make a “reasonable effort” to be understood, and Beauchamp and Childress understand veracity to include “the way the professional fosters the patient’s or subject’s understanding.” The most reasonable way to foster the patient’s understanding can, in some circumstances, include maintaining what I referred to in Chapter 4 as an “oblique” orientation toward truth.
So let us return to the Navajo case. “It is true,” Macklin points out, “that withholding information about the risks of treatment or potential adverse events in the future radically changes what is required by the doctrine of informed consent. It essentially removes the ‘informed’ aspect, while leaving in place the notion that the patient should decide.”\(^{37}\) Macklin considers this to be a lower standard of informed consent than the usual one, but she thinks the reduction worth it in order to advance responsible engagement in a multicultural context and the central concern for beneficence in health care. I am not as convinced that we should consider it a lower standard, so much as the implementation of a contextually sensitive standard. She concludes from the Navajo case that the physician’s obligation with regard to informed consent could best be phrased, “Carry out an informed consent discussion in a manner appropriate to the patient’s beliefs and understanding.”\(^{38}\) The description she gives of this obligation in practice reveals that this is not a particularly revolutionary idea; there is a sense in which this is already the approach that many probably use unobjectionably:

A heart surgeon would describe the procedures, risks, and benefits of bypass surgery in one way to a patient who is another physician, in a different way to a mathematician ignorant of medical science, in yet another way to a skilled craftsman with an eighth grade education, and still differently to a traditional Navajo.\(^{39}\)

Notice that, in order to act on this standard, the surgeon would need to have a certain amount of base knowledge about the patient in order to even begin to know how to go about catering the message.

\(^{37}\) Macklin 12
\(^{38}\) Macklin 12
\(^{39}\) Macklin 12-13
As is clear from Macklin’s description of the sort of responsiveness appropriate to informed consent discussions, the context in which health care professionals can be either justified or obligated to be strategic in the way they reveal the truth is not limited to dealing with patients from different cultural backgrounds. Another salient example occurs when patients are recovering from debilitating events such as a stroke or a spinal injury. In such cases it can be important to reveal the prospects for recovery in a way that calls to mind Buddhist upāya or the approach of the Upaniṣadic sage, strategically revealing the truth in a manner that fosters the patient’s ability to process it effectively and productively. Such “staged disclosure” and other similar approaches to care can be justified, Beauchamp and Childress claim, “despite their threat to trust between clinicians and patients.”

They clarify: “For prognosis, professional norms incorporate the therapeutic value of hope for patients, and virtues of compassion, gentleness, and sensitivity often displace veracity. These norms encourage disclosure of ‘bad news’ over time rather than all at once.”

They cite in support of this claim the example of a physician helping a patient rehabilitate after a stroke:

From the beginning the doctor knew that the patient was unlikely to recover significant use of his arm, and he offered caveats and uncertainty that did not fully match what he believed or felt. He stressed the limitations of prognostication, the unpredictability of recovery, and the need to give the brain a chance to heal. The patient received these answers well at the time, apparently preferring the physician’s “ambiguous statements about the future to the alternative judgment of the permanent paralysis he fears.” This indefinite, but caring and supportive, exchange continued, with the physician praising the patient’s progress in walking.

---

40 Beauchamp and Childress 291
41 Beauchamp and Childress 291. If they understand ‘veracity’ to be equivalent to ‘truth-telling,’ then I agree with them entirely. But if they understand it as equivalent to ‘truthfulness,’ then I would say the other virtues do not displace veracity so much as they constrain it.
and performing daily activities, despite residual weakness. After two weeks, the patient was enthusiastic about his progress and asked, “How about my arm?” The physician responded, “The arm may not recover as much as the leg.” Although this statement confirmed his fears, the patient focused on his overall progress. He had a strong hope that the physician might be mistaken, since he had repeatedly stressed his inability to prognosticate accurately.42

There is nothing in this story to indicate that the physician ever directly asserted something he believed to be false to the patient. Nevertheless, it seems clear that he did intentionally allow the patient to continue holding false beliefs about the prospects for recovery. And to allow someone to persist in holding a false belief when one could have easily corrected that belief fits some definitions of deception.43 But the story also indicates that the patient responded positively to the doctor’s vague pronouncements about his prognosis. If he wanted a more concrete prediction, it seems that he could have asked for or demanded it. In a sense, we can understand the doctor in this story as providing his patient not just with physical therapy, but also with mental (and even emotional) therapy in helping him come to a productive understanding of his present circumstances. This approach is perfectly in keeping with Buchanan’s insistence that the physician make a reasonable effort to be understood: bringing a person around to an understanding slowly can be more reasonable than dumping all the information on him at once.

A final example will demonstrate that there are even occasions where bringing a person around to a correct understanding can involve the literal assertion of what one believes to be false: a case of what I have called falsity being used in the service of truth. The case, discussed approvingly by Beauchamp and Childress, involves an elderly patient

42 Beauchamp and Childress 291
43 See, for instance, Carson, Lying and Deception.
who came to his physician to have an ulcer on his lip diagnosed. The patient, we are
told, had a lifelong fear of cancer, “which, for him, meant a shameful, painful, and fatal
disease that would spread inexorably.” A biopsy confirmed that the ulcer was indeed
cancerous, but that it was a very treatable form of cancer: “squamous cell carcinoma,
which would require only a short course of radiotherapy to cure, without any need for
surgery or even admission to the hospital.” When the patient asked whether he had
cancer, the physician told him that he did not. According to Beauchamp and Childress,
the physician justified his action on several grounds. First, he pointed to
the patient’s deep need for “effective reassurance.” Second, he argued
that it was “more truthful” to tell this patient that he did not have cancer
than to tell him that he did, because it would have been impossible to
inform him that he had a curable cancer “without leaving him with a false
and untrue impression” because of his enduring and unchangeable beliefs.
Third, addressing this patient and his concerns in his own language
expressed respect rather than paternalistic arrogance.

Of particular note for the present discussion is the physician’s second reason, that it
would be “more truthful” to deny that the patient had cancer because of the patient’s
idiosyncratic understanding of what ‘cancer’ means.

Macklin discusses what appears to be the same case, and provides some details
that Beauchamp and Childress leave out. Most important of these is that the patient and
physician had a long-standing relationship, which enabled the physician to be intimately
familiar with this patient’s deep and abiding fear of cancer. He was able to recognize his
patient’s fear as an irrational one, and thus as one that would be unlikely to be corrected

---

44 The case is from a letter published in The Lancet (June 11, 1994), and written by the physician in
question.
45 Beauchamp and Childress 292
46 Beauchamp and Childress 292
47 Beauchamp and Childress 292
48 See Macklin, pages 2-4. I say “appears to be” the same case because Macklin does not provide a
citation. But her description fits so closely the one given by Beauchamp and Childress that it would be
incredible if they proved to be distinct cases.
through argumentation or explanation. This helps to reinforce the point that the physician’s decision was, given the circumstances, the morally correct one. Macklin puts the point well:

This was a matter of the physician revealing to the patient the nature of his ailment, and describing it to him in a way that the patient would properly understand. In recognizing and being sensitive to the patient's fear of cancer, the physician showed respect for the patient's beliefs and values. The physician reasoned that this patient would misunderstand a diagnosis of cancer. Neither respect for persons nor beneficence mandates providing information that a patient would not fully comprehend.\(^\text{49}\)

Though she does not question the rightness of the physician’s decision in this case, Macklin does raise the issue of whether the physician was really truthful, given that he did not tell the truth. This question is more than the semantic quibble that Macklin dismisses it as, for there are legitimate concerns with allowing this patient to continue believing that he did not have cancer. Not only is there a risk of the patient giving an inaccurate medical history to any future health care providers he may have (a risk that is somewhat mitigated by the patient’s advanced age and long-standing relationship with the physician in question), but any of his immediate family members who are kept in the dark about his diagnosis of cancer also stand to give inaccurate family histories to their physicians.\(^\text{50}\) I would advocate explaining to the patient in question after the treatment was complete and he was fully clear of the cancer that he did, in fact, originally have cancer, but that it was one that he needn’t worry about. Hopefully, once there was conclusive evidence that his condition had been treated the patient would be more

\(^\text{49}\) Macklin 4
\(^\text{50}\) Perhaps, given the relatively benign nature of this particular form of cancer, the inaccurate medical histories would be a non-issue. I am not qualified to do more than speculate on this matter, but I am uncomfortable with the prospect of the patient being kept in a perpetual state of ignorance about his diagnosis.
receptive to the truth about the particular form of cancer that he had. Provided a follow-up along these lines, I would answer unequivocally that yes: the physician was being truthful in misleading his patient about the diagnosis of cancer, because he was trustworthy in what he said. His patient could reliably trust the beliefs that were generated as a result of hearing what the physician said.

These examples help to illustrate the complexity of the medical professional’s obligation to care for the patient. In such a complex environment it is naïve at best, and frequently pernicious, to assume that there is a simple, straightforward rule that will successfully guide action. Even a rule as seemingly innocuous as “always tell the truth” can, as we have seen, be problematic. Even though truth is central to the virtues associated with health care, caring for a patient is not as simple as “learn the truth, then tell the truth.” Successful caring requires the ability to use one’s own judgment in novel situations: it requires phronēsis, which is why Beauchamp and Childress note that discernment, one of the focal virtues of caring, “involves the ability to make fitting judgments and reach decisions without being unduly influenced by extraneous considerations, fears, personal attachments, and the like.” Even, I hasten to add, attachment to truth itself.

The importance of judgment and of maintaining trustworthiness when diverging from truth-telling means that there are risks in attempting it. It is easy to go wrong, and difficult to go right. One of the challenges with Aristotelian virtue ethics in general is that it is both difficult and risky to emulate the phronimos if you are not already one yourself (to say nothing of the difficulty inherent in correctly identifying the phronimos).

[51] Beauchamp and Childress 40
To protect oneself from going wrong, it is essential to strive to live up to three responsibilities: to understand the facts, to understand oneself, and to understand one’s audience. These are the Responsibilities of Truthfulness, and it is to these that we turn in the Conclusion.
Conclusion: The Responsibilities of Truthfulness

Becoming an ethically responsible speaker is not something that happens by accident or as a result of fortuitous happenstance, nor is such status attained simply by following a few practical rules. As with any other virtue, the virtue of truthfulness must be carefully and intentionally cultivated. We must be careful to guard against natural tendencies toward self-interest, self-deceit, wishful thinking, the desire to avoid confrontation, and a host of other traits that can discourage the discovery and communication of the truth. By articulating the three Responsibilities of Truthfulness I aim to provide a set of considerations that can aid in avoiding these tendencies. Someone who has taken care to understand the facts, to understand herself, and to understand her audience—as all three are described below—and who makes sure that what she says is informed by these three, will have gone a long way toward ensuring that her utterance is truthful and thus that she is being trustworthy in making it. This is not to say that the three conditions are sufficient to guarantee truthfulness, either at a particular moment or as a general disposition. But without all three, it is not possible to be truthful. Anything that appears to qualify as truthfulness without having taken the three Responsibilities into account is, to recall Vācaspati Miśra, merely the semblance of truthfulness, not truthfulness.

I call these the Responsibilities of Truthfulness because a serious communicator is responsible for accounting for them, and will typically be held responsible for failing to do so. Someone who claims to know what he does not actually have good reason to claim knowledge of is responsible for the consequences of his reckless assertion. Failing to care enough about the epistemic justification for what one says publicly and
authoritatively approaches what Frankfurt has identified as bullshitting, which is in some ways worse than lying. Similarly, to fail to calibrate one’s own self-image with others’ image of oneself, or to fail to consider the way one’s audience is likely to interpret what one says and the use to which such understanding will be put is to court misunderstanding and miscommunication. They are the Responsibilities of Truthfulness because to speak without having done one’s best to account for them is in a very literal sense to be irresponsible.

The first Responsibility of Truthfulness is to understand the facts. Understanding the facts involves both knowledge and receptivity: truthful speakers must take care to have true beliefs and must allow those beliefs to affect them and their understanding of the world. That is to say, one must not be simply a collector or chronicler of facts; one must instead be a synthesizer of information. Encyclopedic knowledge will help one win at the board game Trivial Pursuit, but by itself it will do little to contribute to living a good life. Knowing what is true and knowing what to say are seldom the same thing—especially when ‘what to say’ includes ‘how to say it’.

The responsibility to understand the facts is in many ways reminiscent of Grice’s maxims of Quality. His second maxim, “Do not say that for which you lack adequate evidence,” points to many of the same concerns. Of course the maxim needs some revision. There is nothing wrong with saying something for which you have insufficient evidence, even according to Grice’s account: the problem lies in asserting it or otherwise saying it so as to invite and inspire a greater level of confidence than is warranted. More importantly, however, as I pointed out in the previous chapter, the idea behind Grice’s category of Quality is that one’s contribution should be true—and this means that
sometimes asserting what you believe to be false (and thus lack sufficient evidence for) can be the most conversationally responsible way to make a true contribution.

Even though I allow for deception under the name of truthfulness, my basic sentiment is the same as what I take Grice to be trying to convey. In order to try to bring a person around to a productive understanding of the truth, it is essential to have a good idea of what the truth is. It is irresponsible to represent yourself as understanding what you do not, in fact, understand. Doing so invites others to trust where you know trust to be unwarranted. It amounts to offering to care for something that you know you are not in a position to care for. It is irresponsible to invite an obligation that you have good reason to believe you cannot fulfill.

This is not to say that there is something inherently culpable about speculation. We cannot always know, and sometimes we must speak on a matter about which we cannot claim the level of confidence that knowledge requires. It would be inappropriate, for instance, for a doctor to refuse to speculate about a patient’s prognosis, or for an attorney to refuse to speculate about a client’s chances of success if a legal action that is under consideration were to be pursued. It is essential, though, in such circumstances to make it clear that this is speculation and not an assertion of fact. To pass off one’s conjectures or inadequately warranted convictions as well-founded knowledge-claims is to fail in one’s ethical obligations as a speaker.

The importance of understanding the facts calls attention to the ethical import of epistemology as a field of study. Given the necessary role that knowledge plays in understanding, any account of understanding the facts presupposes answers to such questions as: What does it mean to know? What qualifies as adequate justification for
claiming to know? Which purported means of coming to know are legitimate and which illegitimate? These questions and more have been discussed and debated for thousands of years, in all philosophical traditions with which I am familiar, and I harbor no illusions that I will be able to settle them here. Fortunately, I do not think I need to. Different epistemological frameworks will answer differently the question of what it means to know, but they can all agree on the importance of not purporting to know what you do not know. I raise the issue here not to try to answer the questions, but rather to point out their relevance to the ethical considerations that have been occupying us in this project. While epistemological questions are interesting and valuable for their own sake, they are not mere intellectual curiosities: they have direct bearing on the ethics of speaking, one of the most fundamental ways in which we interact with one another.

Another important aspect of the responsibility to understand the facts is to recognize the limitations of one’s own knowledge or understanding. That one should take care not to speak as an authority about topics regarding which one is not an authority is not, I take it, a controversial claim. But this responsibility can be more difficult to live up to than it at first appears. When Socrates sought to understand the Oracle’s claim that no one was wiser than he, his inquiries led him to the realization that many who are wise in one arena think themselves to be correspondingly wise in other arenas about which they know much less. It is important to be like (this image of) Socrates in being aware of the limits of one’s knowledge, and not to presume or to claim that one knows what one does not. At the same time, it is important not to simply stop at the recognition of one’s own fallibility or limitations. In discussing the virtue that she calls “relational humility”, Vrinda Dalmiya points out that self-doubt can lead to a defeatist attitude unless it is
accompanied by increased confidence in the other, but when the two are mutually
entailing great epistemic gains can be achieved:

Relational humility works with these twin trajectories of epistemically
receding oneself while bringing others to the forefront. Without the
former, granting a voice to others simply leads to a cacophonous
multiplicity of epistemic perspectives. Without the latter, we end up in a
passive and stagnating acceptance of human fallibility. In a healthy
community of knowers, hitherto ignored perspectives progressively come
to light and are taken seriously because of the slackening of the
stranglehold of the centre. Self-ascription of ignorance therefore becomes
the heart of continued cognitive growth in a community. ¹

Understanding the facts is not just a matter of gathering up all the information one can
and then deciding how and when to distribute it to others; it is recognizing that others
have access to some information that I do not, and that if we work together we can both
come to understand more fully than either of us can alone or in competition with one
another.

Another point that is worth noting is that sometimes it is better not to know, and
thus that it is sometimes best not to ask. An obvious example of the importance of not
knowing occurs in medical research involving human subjects. Wherever feasible, such
research is typically conducted in a double-blind format so that those who interact with
the research subjects do not know which subjects are receiving the experimental protocol
and which are receiving the placebo. This is to ensure that those who are responsible for
measuring the results of the experimental drug (or whatever) do not unintentionally skew
the results of the study by, for example, treating the different groups of subjects
differently. While double-blind studies have little bearing on the issue of truthfulness in
what we say, their importance to scientific research shows that when you are interested in

¹ Dalmiya 12-13
an overall process being truthful it can be important for some people involved in that process to not be aware of the truth.

It can also be important not to know, or not to try too hard to find out, the truth when trying to build or maintain a relationship of trust. As discussed in Chapter 3, a key feature of trust is the discretion that is afforded the trusted. Someone who hires a private investigator to document a spouse’s business trip, or parents who demand that their teenage child submit to random drug testing, not only demonstrate a lack of trust, they also prevent trust from developing. These are extreme examples, but less extreme cases can have the same effect: in precluding the development of trust, they undermine the ability of all members of the relationship to become trustworthy, in both word and deed. Such inquisitorial behavior actually undermines a person’s ability to become a trustworthy speaker—to foster the virtue of truthfulness.

There is, though, a danger in taking the desire not to know too far. As Bernard Williams notes:

> Self-conscious pursuit of the truth requires resistance to such things as self-deception and wishful thinking, and one component of the virtue of Accuracy…lies in the skills and attitudes that resist the pleasure principle, in all its forms, from a gross need to believe the agreeable, to mere laziness in checking one’s investigations.²

With greater knowledge comes greater responsibility, and it is sometimes tempting to instead remain blissfully ignorant. This can be true in romantic relationships: a husband may choose not to ask questions because he would rather remain oblivious than face the unpleasant prospect of his wife’s infidelity and the difficult choices that follow from such behavior coming to light. The temptation to ignorance is also important with regard to

² Williams 125. The entirety of Chapter 6 is relevant to the present discussion.
the myriad social issues that affect us today. In many ways it is easier to live a
comfortable and guilt-free life if one remains ignorant of the arguments in favor of
vegetarianism, or of how veal calves are raised, or of the social and environmental
impacts of fossil fuel consumption, to name just a few examples. But we act with regard
to these things regardless of our knowledge. It is no less irresponsible to do so without an
awareness of the relevant facts than it is to speak as an authority about a subject of which
one is ignorant. The responsibility to know the facts extends to our actions as much as to
our utterances.

Just as there are many motivations to keep the truth hidden from oneself, so too
are there many motivations to keep the truth hidden from others—not all of which are
noble. This is the point at which the responsibility to understand the facts, which we
have been discussing, intersects with the responsibility to understand yourself, to which
we now turn. One of the reasons paternalism is such an insidious concern is that it is so
easy to convince yourself that you are acting for the benefit of another when in fact you
are acting purely (or mostly) out of self-interest. Often when we lie to others, for
instance, we tell ourselves that we are protecting the other from a painful truth when in
fact we are afraid of the consequences for ourselves if the truth is discovered.

This is one way in which knowing oneself is an essential component of
truthfulness: we must defend against self-deception. This is the case in an environment
wherein one must strive always to tell the truth; doubly so if I am correct that truthfulness
allows for deviation from the truth. Bernard Williams correctly argues that it is
appropriate to call Accuracy a virtue precisely because it must be cultivated against a
natural tendency toward wishful thinking and epistemic laziness. Humans have a natural
tendency to give disproportionate weight to those phenomena that support our
preconceived notions and to interpret ambiguous phenomena as also supporting our prior
ideas. This is called confirmation bias: we tend to notice what confirms our hypotheses
and ignore what disconfirms them; and it is one of the things that scientific method aims
to correct for.

The other sense in which it is imperative to know yourself is to be aware of how
others understand you—to calibrate your self-image with your audience’s image of you.
To refuse to consider others’ opinions of oneself amounts to a sort of practical solipsism.
As we will see, this does not mean that one must uncritically accept others’ opinions of
oneself, but a failure to account for these (potentially divergent) views will serve to
undermine one’s ability to be an effective communicator of information. Yet again, the
boy who cried wolf provides an illustrative example. Had he been sufficiently reflective,
he would have realized that blatant lying is not a viable long-term strategy because
people would start disbelieving him. At the very least, he would have come up with a
different means for soliciting help when the wolf actually came, aware that another cry of
“Wolf! Wolf!” would likely not incite the townsfolk to his assistance.

Other people’s perception is colored not just by our previous actions, but also by
the roles we are taken to play. For those roles one has chosen to take upon oneself, it is
important to recognize, embrace, and fulfill that role. For instance, parents and teachers
need to recognize and take seriously the huge impact they have on their children and
students just in virtue of their position of authority. Similarly, professionals—doctors
and lawyers, for example—must recognize their roles within a larger system and act
appropriately.
As I indicated, however, this does not mean that we must blindly accept whatever roles we find ourselves thrust into. Societally imposed gender roles are just one example of a role that a person may have good reason to want to disavow. This is an important issue that needs to be addressed, but it lies beyond the scope of the present essay to address it here. One point that I do want to make is that in order to intentionally disavow a role that has been imposed, it is necessary to recognize, first, that the role has been imposed, and second, what the imposition of this role means as far as how others will tend to interpret what one says and does. Alasdair MacIntyre says that understanding those roles that others take one to play is a necessary component of understanding oneself at all: “We enter human society…with one or more imputed characters—roles into which we have been drafted—and we have to learn what they are in order to be able to understand how others respond to us and how our responses to them are apt to be construed.”

Nancy Potter points out that being forced into a marginalized role often makes it more difficult to receive “uptake”—to be taken seriously as meaning what one intends to mean. Miranda Fricker deals with the same issue under the name of testimonial injustice, arguing that it is the responsibility of the audience to correct for the prejudices that contribute to testimonial injustice. She is correct, of course, but it is also important for the utterer to be aware of any such prejudicial tendencies: to account and correct for them when possible, or at least to enable a more coherent understanding of circumstances and outcomes when such correction is not feasible.

In order to calibrate one’s self-image with others’ image of oneself it is, naturally, necessary to understand those others. But image calibration is not the only sort of

---

3 *After Virtue* 216
responsiveness that knowledge of one’s audience enables—it also makes possible the responsible calibration of the utterance itself. A great illustration of this occurs in the Analects:

Zilu inquired, “On learning something, should one act upon it?” The Master [Confucius] said, “While your father and elder brothers are still alive, how could you, on learning something, act upon it? Then Ranyou asked the same question. The Master replied, “On learning something, act upon it.”

Gongxi Hua said, “When Zilu asked the question, you observed that his father and elder brothers are still alive, but when Ranyou asked the same question, you told him to act on what he learns. I am confused—could you explain this to me?”

The Master replied, “Ranyou is diffident, and so I urged him on. But Zilu has the energy of two, and so I sought to rein him in.”

Confucius’ behavior here demonstrates the crux of the idea that the responsible speaker caters the utterance to the needs, interests, and predilections of the particular audience. To do this well, of course, it is essential to actually understand your audience: the third Responsibility of Truthfulness.

In a sense, the idea that I am expressing here is so banal as to hardly warrant mentioning. It is obvious to anyone who has done any marketing that you must cater your message to your target audience. There is good reason for the difference between the commercials that air during Saturday morning cartoons and those that air during primetime on the Discovery Channel, or between the advertisements in Scientific American and those in Bazaar. Companies conduct extensive market research to ensure that they know their target audience well enough to cater their messages (and their products) successfully. And according to the Federal Trade Commission’s Policy Statement on Deception, whether a commercial advertisement qualifies as deceptive

---

4 Analects 11.22

249
depends not just on what is said, but also on who the target audience is: the same advertisement may be deceptive if targeted at children but not if targeted at adults, for instance.\(^5\)

Thomas Kasulis uses the example of a little league coach to illustrate how such catering infuses our lives beyond commercial advertisements.\(^6\) In his example, the coach tells one player to relax and just focus on letting the bat hit the ball, not worrying about how hard she swings; he tells another player not to worry about hitting the ball at all, but instead just to swing away. Kasulis then imagines a Kantian-inspired metaphysician who witnesses these exchanges and accuses the coach of lying: the two pieces of advice are contradictory, and so clearly he must have lied to at least one of his aspiring athletes. The metaphysician misses the point, of course. The coach was not lying, he was explaining to each of his players what they needed to understand so that they could improve from their individual starting points. Like Confucius, and like the Buddha, this coach is being truthful by responding appropriately to his audience, doing his utmost to ensure a productive understanding of what he has to say. The coach who is either unwilling or unable to respond to the particularities of different individuals, who tells everyone precisely the same thing regardless of relevant differences (in experience, understanding, desire, or ability), is the one we need to worry about. It is exceedingly difficult, but imperative, for those who are in any sort of a teaching or coaching position to be able to respond to the particular needs of individuals in this way.

\(^5\) http://www.ftc.gov/bcp/policystmt/ad-decept.htm
\(^6\) Kasulis x-xiv
This reveals again the importance of openness as it was discussed in the previous chapter. Gadamer says that we must be open to the text, must let the text bring to light our own prejudices by bringing us up short with regard to our own expectations. In order to understand a text, one must not approach that text with the belief that one already fully understands it. The same is true when in conversation with a person rather than engaging with a text: we must let the other speak to us, both literally and figuratively. I must be open to the possibility that my audience does not understand things in the same way I do, does not value the same things that (or in the same way as) I do. I must not simply impose my own view onto my interlocutor.7

Knowing and being responsive to one’s audience captures the difference between the doctor who chose not to tell his patient about the possible effects of a thyroidectomy and the one who chose not to inform his patient of a non-terminal cancer diagnosis. The first was responsive to his own idea of what would be best for the patient, but did not actually consider the patient’s perspective: he was not open to the patient. The second catered what he said to the way he knew his patient would understand it. The first was presumptuous and paternalistic; the second was responsively and responsibly truthful. The difference between these two doctors helps to illustrate an important concern with the responsibility to understand one’s audience. It can be difficult to know the facts because of both external and internal obstacles to truth-discovery (the world is resistant to my will, my beliefs are less so). Both of these limitations are increased when it comes to

7 One of the great values of relational humility, as Dalmiya describes it, is that it serves as a corrective against this sort of imposition and the testimonial injustice it leads to. The clear relevance that relational humility has to all three of the Responsibilities of Truthfulness shows that we are better off not thinking of them as distinct and independent achievements, but rather as interdependent and mutually entailing cognitive accomplishments, to be cultivated in tandem with one another.
the responsibility to understand another person. People lie, deceive, and otherwise mislead themselves and others about their own desires and interests, and there is less opportunity for independent verification about other people than there is about more mundane “objective” facts. We often see people as we want to see them, instead of as they really are. Our opinions of people tend to be resistant to modification—and the stronger our feelings, the more resistant they tend to be.

There is nothing really that we can do about this fact other than acknowledge it and continue to do our best to understand each other as well as we can. Here is one place where the ancient (both Pyrrhonian and Cārvāka) skeptical traditions may have something helpful to contribute. The value of recognizing that we cannot know with certainty can lead us to be more careful in the claims we make and less dogmatic in our assessments of ourselves and of one another. We must do our best to understand each other in order to care responsibly for one another, in order to be trustworthy in what we say and in how we act. A perfect understanding of each other is, of course, out of reach. But if we are aware of this, and if we are open about our limitations—and open to each other—our best may be good enough.

Before I bring this dissertation to a close, there is one more concern that I need to address. This concern is that my proposed analysis of truthfulness commits me to either an infinite regress or a vicious circularity, wherein trustworthiness is defined in terms of itself. This becomes apparent when we consider the following five claims that I seem to advocate:

1) Trustworthiness in speech is truthfulness.

2) Truthfulness consists in understanding the facts, oneself, and one’s audience.
3) **Understanding is a matter of knowledge and receptivity.**

4) **Any conception of knowledge presupposes a notion of truth.**

5) **Truth consists in trustworthiness.**

The problem with this set of premises, of course, is that it begins and ends with trustworthiness. So I seem to be saying nothing more than, ‘trustworthiness presupposes a notion of trustworthiness,’ and all this talk of truth, truth-telling, truthfulness, and lying is just an elaborate mechanism to distract from the man behind the curtain (even if unintentionally so).

My response to the regress/circularity objection is twofold. First, I do not take this set to constitute a vicious regress. The first premise begins not with trustworthiness, but with trustworthiness *in speech*: to begin there and end with trustworthiness amounts to saying that trustworthiness in speech is a species of trustworthiness. While this is not a particularly informative claim, neither does it commit one to an infinite regress. It merely reminds us that if we want to understand truthfulness we would do well to explore the notion of trustworthiness more thoroughly.

Second, statement (2) is simply false, and statement (5) is, if not incorrect, then at least misleading in its formulation. I do not take fulfilling the Responsibilities of Truthfulness to *constitute*, but rather to *contribute to*, being truthful. They are, as I said, necessary conditions rather than sufficient ones. The attitude that one will naturally cultivate and convey in striving to live up to the Responsibilities of Truthfulness is what is most important, and that attitude will be one of caring and compassion. Such an attitude will not preclude an utterer from deceiving someone, even intentionally, but it will ensure that such deception will only happen when there is good reason to believe that
it really is in the best interests of the one being deceived. As I argued in Chapters 1-3, I think there are good reasons for refusing to call such cases of beneficial deception ‘lies’, and as I argued in Chapter 4, I think there are good reasons for considering a person who engages in such behavior ‘truthful’.

I also do not take truth to consist in trustworthiness. Certainly reliability is an important aspect of truth. But trustworthiness, as discussed in Chapter 3, requires an amount of discretion on the part of the one who is trustworthy. And while I think our notion of truth needs to remain broad enough to account for the fact that a friend or a spouse can be true—that is, that truth can be predicated of people or of actions—it also needs to retain the ability to account for the fact that a statement (sentence, proposition, etc.) can be true. And statements are not the sort of thing that trustworthiness is properly predicated of. I am arguing that we would do well to remember that the ability to safely trust is the key to why we value truth, not that it is constitutive of truth.

The Responsibilities of Truthfulness do not provide a mechanism for guaranteeing that one will achieve truthfulness. They are guidelines, not rules. Human interaction in general, and linguistic interaction in particular, is simply too complex, malleable, and full of novel circumstances to reasonably expect concrete rules that provide a clear answer to questions such as, ‘What should I say?’ or ‘What should I do?’ in any particular situation. This should not be seen to indicate a limitation in my account. Ethics is not, and we should not strive to make it, as cut and dried as arithmetic. Early in the Nicomachean Ethics Aristotle insists that we should not demand greater specificity in our inquiries than our subject matter permits, and Kant’s Categorical Imperative is much better at ruling out unethical actions than it is at providing positive guidance regarding what one should
actually do in any particular situation. For one who genuinely seeks to be trustworthy, to speak ethically, the Responsibilities of Truthfulness will point you in the right direction and help you to avoid the dangers of self-deception, self-interest, and epistemic and conversational laziness. They could not do more than this—they could not, for instance, explain univocally exactly what utterance one should make in order to qualify as being truthful here and now—and yet remain true to the dynamic and multi-faceted nature of human linguistic exchange.
Works Cited


