

TRUTH IS EXPENDABLE: FOUNDATIONS
FOR AN EMPIRICALLY INFORMED
PHILOSOPHY OF TESTIMONY

by

Michael Wayne Hermon

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

The University of Utah

August 2011

Copyright © Michael Wayne Hermon 2011

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Michael Wayne Hermon
has been approved by the following supervisory committee members:

Mariam Thalos, Chair 9 June 2011
Date Approved

Stephen Downes, Member 12 June 2011
Date Approved

Ronald Mallon, Member 9 June 2011
Date Approved

Clifton McIntosh, Member 9 June 2011
Date Approved

Richard Fowles, Member 9 June 2011
Date Approved

and by Stephen Downes, Chair of
the Department of Philosophy

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

The goal of this dissertation is to offer an empirically informed evaluation of testimony as a source of knowledge. Epistemologists have assumed that testimony is a generally reliable source of true beliefs because human cognitive faculties would have evolved to be reliable at getting the truth. However, complementary evidence from the signaling theory and social psychology literature shows that testimony is a practical tool with a variety of nonepistemic functions, including forming and maintaining social relationships, coordinating group behavior, and prescribing conduct. Since the value of using testimony is very often independent of its accuracy, humans have evolved to expend as little resources on checking for accuracy as is necessary to satisfy their other needs. In other words, “truth is expendable” to humans trying to get along well in the world and with each other. This implies that testimony is a far less epistemically reliable source of information than philosophers have assumed, and although it is very often prudent to simply believe what people say, it is not epistemically rational to do so. At the end, I offer preliminary empirically informed prescriptions for judging the reliability of testimony.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	vi
CHAPTERS	
1. INTRODUCTION	1
The Goal	1
Chapter Summaries	4
PART I: THE PHILOSOPHICAL STUDY OF TESTIMONY	8
2. THE HISTORY OF THE EPISTEMOLOGY OF TESTIMONY	9
Skeptical Dismissals of Testimony.....	9
Hume's Reductionism	12
Reid's Nonreductionism	13
Radical Translation and the Reliability of Testimony	17
3. THE NATURE OF TESTIMONY	22
Formal Testimony and Natural Testimony	22
"Testimony" as a Suitable Target of Epistemic Evaluation	26
4. NONREDUCTIONISM	28
Nonreductionism and the Default Rule	28
Introducing the Nonepistemic Uses of Testimony	31
The Value of Shared Beliefs	33
Default Acceptance and Memory	35
5. REDUCTIONISM	37

Hume's Reductionism	37
Global Reductionism	38
Objections to Global Reductionism	39
Epistemic Heterogeneity and Nonreductionism	41
Statistical Local Reductionism	43
Permissive Local Reductionism	45
Conclusions	48
PART II: THE SCIENCE OF HUMAN COMMUNICATION	50
6. INTRODUCTION	51
7. THE EVOLUTION OF HUMAN COMMUNICATION	54
Introduction	54
From Inferential Translation to Strategic Signaling	55
Adaptive Signaling	60
The Evolution of Manipulation	61
Signaling in Contexts of Coincident Interests	62
Contexts of Noncoincident Interests	66
Conclusions	69
8. THE PSYCHOLOGY OF INFORMATION REGULATION	73
Introduction	73
Psychology and the Default Rule	75
Motivation and Information Bias	76
Ingroup Bias	81
Group Bias In Young Children	84
Self-image Bias and Manipulation	85
Confirmation Bias	88
Conclusions	90
PART III: A NEW EPISTEMOLOGY OF TESTIMONY	94
9. ACCURACY IS EXPENDABLE	95
Introduction	95
Conclusions: Epistemic Principles for Evolved Beings	96
Prescriptions: A Preliminary Empirically informed Philosophy of Testimony	100
Future Research	106
BIBLIOGRAPHY	108

ACKNOWLEDGMENTS

This dissertation would not have been possible without the unwavering support and confidence of Mariam Thalos. I will be forever grateful for her ongoing thoughtful encouragement. I would also like to acknowledge my darling partner, Ella, who's loving companionship and enthusiasm for my work provides me the motivation to develop and express important ideas.

I would also like to thank the other members of my committee: Steve Downes, Ron Mallon, Clif McIntosh, and Richard Fowles for evaluating this research, and the Department of Philosophy and Graduate School at the University of Utah along with the National Science Foundation for invaluable financial support over the years.

CHAPTER 1

INTRODUCTION

The Goal

The goal of this dissertation is to offer an empirically informed evaluation of testimony as a source of knowledge. Epistemologists have assumed that testimony is a generally reliable source of true beliefs because human cognitive faculties would have evolved to be reliable at getting the truth. However, complementary evidence from the signaling theory and social psychology literature shows that testimony is a practical tool with a variety of nonepistemic functions, including forming and maintaining social relationships, coordinating group behavior, and prescribing conduct. Since the value of using testimony is very often independent of its accuracy, humans have evolved to expend as little resources on checking for accuracy as is necessary to satisfy their other needs. In other words, “truth is expendable” to humans trying to get along well in the world and with each other. This implies that testimony is a far less epistemically reliable source of information than philosophers have assumed, and although it is very often prudent to simply believe what people say, it is not epistemically rational to do so.

The prevailing view is that true beliefs are essential to good decision making and therefore adaptive, while holding false beliefs can result in poor decision making and are

therefore maladaptive. It is therefore common to assume that humans would have evolved cognitive practices that are reliable at acquiring true beliefs. This story is prominent in the contemporary philosophical literature, specifically in the literature on the epistemology of testimony, where it is assumed that believing what other people say is a generally reliable way to acquire knowledge.

Epistemic errors on this view are always explained by either some malfunction in the traits responsible for judging information or result from some limitations on the reliability of human cognitive faculties. I argue for a more plausible explanation that epistemic errors often result from the prudent use of testimony that abandons good epistemic practices. For example, individuals depend on memberships in groups that require conviction in shared beliefs where the accuracy of those beliefs is less important than securing the benefits of membership. In cases like this, humans will ignore counter-evidence and accept the required beliefs. In cases like this, trading-off accuracy for the practical benefits of accepting testimony can lead to widespread epistemic errors.

In their optimism about the reliability of testimony, epistemologists have suggested norms for judging the reliability of testimony that are permissive of credulity. The majority of epistemologists of testimony think it is very often rational to simply believe what people say unless there is a good reason not to, and they insist that this practice often results in the hearer coming to know what they accept. If testimony is primarily a practical and not an epistemic tool, then credulity is not a reliable epistemic practice.

At the end I, will offer prescriptions for judging the reliability of testimony that are informed by the signaling theory literature that has identified specific requirements

for stable signaling systems. The context in which testimonial interactions take place is significant to the reliability of testimony in specific ways. For example, stabilizing valuable signaling when signal production is cheap requires that there be effective enforcement of honesty. If there is no enforcement, this is a good reason to think that the testimony is epistemically unreliable.

In addition, the social psychology literature offers descriptions of actual human practices used when processing information and identifies various motivations humans have in accepting and offering testimony. A useful strategy for evaluating testimony is therefore to try and determine the motives of a speaker to believe what they claim and their motives in offering the testimony. If these are nonepistemic motives such as trying to get the hearer to behave in specific ways, this would be a good reason to think that the testimony is not epistemically reliable. The empirical literature is therefore a guide to prescribing improved strategies for evaluating testimony.

I want to alert the reader to the fact that I have arbitrarily selected terms to use when describing the participant of dyadic testimonial interactions. When I am talking about human testimonial interactions, I will use the terms “speaker” and “hearer” of testimony, to refer to what I will call the “sender” and “receiver” when I am talking more generally about signaling. It should therefore be clear when I am talking about human communication that involves the use of testimony and signaling more generally. Recognize that the consistent use of these terms is a convenience and does not imply that all of testimony is spoken or heard. There are certainly ways that humans communicate that produce beliefs that do not involve speech or hearing. For example, testimony, as I will define it in Chapter 3, can also be offered through writing and gestures.

Chapter Summaries

What follows in Part I is a critical survey of the philosophical literature on the epistemic status of testimony. In the next chapter, I introduce the history of the philosophical study of testimony as a source of knowledge, and display the historical origins of the current framing of the debates within the epistemology literature.

Contemporary interest in the epistemology of testimony has focused on a distinction in brief remarks made by David Hume and Thomas Reid on the epistemic status of testimony. The primary problem with the epistemology of testimony, the so called “epistemological problem of testimony,” has attempted to answer the question, how is it that humans can very often come to know what they accept from testimony? This problem clearly assumes the general reliability of testimony as a source of factually accurate beliefs. The two solutions to this problem have been to follow the tradition of Hume and claim that the reliability of testimony is inferred from past instances of testimony providing the truth, and the competing position in the tradition of Thomas Reid that testimony is a direct source of knowledge like perception or memory.

In Chapter 2 I also introduce the reader to W.V.O. Quine's approach to understanding the emergence and persistence of shared meaning. I introduce Quinean semantics because this tradition is often cited in favor of the view that the use of testimony would not be stable and persist unless the vast majority of testimony was true. I will be contrasting Quine's theory of translation with the approach from the evolution of signaling I later introduce in Chapter 7, and argue against a priori arguments that testimony must be epistemically reliable.

Chapter 3 is a short chapter in which I address the nature of testimony. Included

is a critical examination of Coady's theory on the nature of testimony in which I reject defining testimony as a type of speech act that meets specific epistemic requirements. Instead, I offer a broad definition of testimony to be the target of epistemic evaluations. "Testimony," for the purposes of this dissertation, will be any communication that results in new beliefs.

Chapter 4 and 5 examine the key debate in the epistemology of testimony between so called "reductionism" and "nonreductionism." These theories offer various solutions to the epistemological problem of testimony by describing the requirements on a hearer of testimony to be epistemically justified in accepting that testimony. To clarify, a hearer of testimony is "epistemically justified" in accepting testimony if upon accepting that testimony they would come to know what they have been told if it is true. This is important because throughout the dissertation, I will be contrasting "epistemic justification" from prudence, which is a justification for accepting testimony because of the practical value of doing so.

In Chapter 4, I critically examine "nonreductionism," which is the view in the tradition of Thomas Reid, that testimony is a direct source of knowledge analogous to perception or memory. I will argue for separating the various sources of beliefs and examining their reliability independently. In this chapter, I point out various differences between testimony as a source of beliefs and other sources such as memory or perception.

In Chapter 5, I examine various types of "Reductionism," which in general is the view, in the tradition of Hume, that requires that a hearer of testimony have positive evidence that the testimony is reliable. Here I distinguish between "local" reductionism, which is the view that a hearer of testimony must have positive reasons to believe each

instance of testimony is reliable, and “global” reductionism, which is the view that a consumer of testimony can have good positive reasons to think testimony is a generally reliable source of beliefs.

By the end of Part I, I will have offered reasons to reject both theories and will have displayed the presuppositions of the epistemological problem of testimony. In Part II, I show that the the presupposition, that testimony must be generally reliable, and that credulity very often leads to knowledge, are false.

Part II of the dissertation moves from the philosophical study of human communication to empirical findings on human informational biases and the evolution of human communication. In Chapter 7, I introduce the literature on the evolution of signaling, and argue that human communication evolved primarily because of the benefits of traits that increased the Darwinian fitness of participants, a function that only sometimes requires that communication be true. Chapter 7 concludes with the observation, contrary to Quineian semantics, that nothing in the evolutionary description of the origins and stable use of testimony requires that testimony be epistemically reliable. I argue that this presupposition of the epistemology of testimony, that testimony is a generally reliable epistemic tool, amounts to a just-so story about the function of testimony that is inconsistent with good evolutionary thinking.

Chapter 8 is a survey of the social psychology literature that shows humans will abandon good epistemic practice in favor of satisfying other psychological and social needs. Humans regularly believe similar individuals while rejecting the claims of members of others groups, and are prone to believing what will improve the practical results of their performance, irrespective of accuracy. Part II of the dissertation

concludes with a number of observations about human testimonial practices that provide evidence that testimony is far less reliable than epistemologists have assumed.

Part III of the dissertation uses the empirical literature to draw conclusions about the epistemic reliability of testimony. I argue that for humans, factual accuracy is expendable because practices are better adapted that improved practical outcomes even at the expense of accuracy. This has significant implications for the epistemic status of testimony that the reliability of testimony is highly context-dependent and depends on the motives of the participants. In the penultimate section of the dissertation, I offer new prescriptions for judging the reliability of testimony that are informed by the discussion of signaling contexts in Chapter 7 and the discussion of motivations in Chapter 8. In the last section, I suggest a number of new research projects within an empirically informed philosophy of testimony.

PART I

THE PHILOSOPHICAL STUDY OF TESTIMONY

CHAPTER 2

THE HISTORY OF THE EPISTEMOLOGY OF TESTIMONY

Skeptical Dismissals of Testimony

Given the importance of believing what people say in human life, it is surprising that very little has been said about the epistemic status of testimony in the history of philosophy. The few historical references to the so-called, “epistemology of testimony,” go back to Plato, Descartes, and Locke, who all expressed skepticism over the epistemic reliability of testimony. In all three cases, their position was that testimony is an inferior source of knowledge compared to the ideal epistemic methods they favored. For example, in the *Meditations*, Descartes says:

So, too, I reflected that we were all children before being men and had to be governed for some time by our appetites and our teachers, which were often opposed to each other and neither of which, perhaps, always gave us the best advice; hence I thought it virtually impossible that our judgments should be as unclouded and firm as they would have been if we had had the full use of our reason from the moment of our birth, and if we had always been guided by it alone. (Descartes, 1641, p. 117)

Descartes view is that only the “full use of our reason” produced the “unclouded and firm” judgments characteristic of good epistemic practice, whereas, the advice of teachers fails to be a consistently good source of knowledge, and may be no better than being guided by emotions or “appetites,” which he clearly derides as an irrational affect that

sullies our beliefs, and results in only opinionated reactions instead of knowledge. Locke went on to voice a similar concern:

We may as rationally hope to see with other men's eyes as to know by other men's understanding...The floating of other men's opinions in our brains makes us not one jot the more knowledge, though they happen to be true. What in them was science is in us but opiniatrey. (Locke, 1961, p. 58)

Locke's contention is that epistemic justification is lost during the transmission of beliefs from one individual to the next, presumably because the speaker's justification for their own belief is not transferred during the act of testimony. Contemporary theorists have been quick to deny this by claiming that when a claim is offered as testimony, the act of testifying provides *prima facie* evidence for the claim being communicated. This is because an act of sincere testimony includes a tacit commitment by the speaker to its accuracy. Nonreductionists, who I will discuss in Chapter 4, argue that a “fundamental” trust in speakers is rational, because when a speaker knows what they say, this epistemic standing adheres in a way that allows even credulous hearers to receive that knowledge.

However, for Locke, “science,” the reliable epistemic method, required inductively inferring generalizations from raw sensory data, and this inferential strategy owed nothing to believing the say-so of others. Descartes, too, identified good “science” with the “full use of reason,” which was a practice of individual minds, and independent of a social justification for beliefs.

In both cases, good science was being distinguished from the speculations of the common man, rigor was being championed over common opinions, and “science” favored over “opiniatrey.” There was a demand to replace the most common source of the beliefs for the folk with an internal, logical rigor. Descartes and Locke are expressing

an epistemic elitism. The mundane activities that dominate normal life did not meet their standards of a proper “science.” It is understandable that the influence of the new science of Newton as an epistemic ideal, and the vast number of superstitions and cultural errors it supplanted, would lead to a mistrust of everyday epistemic practice dominated by individuals voicing their opinions on matters. It is even more the case today that the claims of science are increasingly inconsistent with everyday beliefs and methods. Getting the facts seems instead to require training, talent, and resources, but even when these are present, factual accuracy is a difficult goal to achieve.

Contemporary philosophers have universally rejected these views on the grounds that they imply a very radical skepticism. Neither Descartes nor Locke seem to admit the ubiquity of testimony as a source of beliefs. Most contemporary epistemologists agree with the thesis that all beliefs held by humans are derived from the say-so of others, because the interpretive framework humans rely on has its source in early development and learning. Therefore, if Descartes and Locke were correct, humans would know almost nothing.

For the next couple hundred years, British and American philosophers committed to the “positivist” movement of radical empiricism focused almost exclusively on perception and inference as a legitimate source of knowledge. The grand project was to show how high-level scientific claims, such as physical laws, could be inferred from observational statements. Testimony was consequently ignored in favor of the view that good epistemic justification came from inferences, from indisputable foundational beliefs, and was therefore radically individualistic.

However, contemporary philosophers are keenly aware that science itself is

fundamentally dependent on testimony, and much scientific practice is regulated by judgments over communicated information. For example, the practice in the publication of scientific papers of having multiple blind referees review the submissions is a special type of testimony regulation (Zollman, 2009). In addition, science is now recognized as a social endeavor requiring many researchers acting in coordinated ways, which may be influenced by various norms of communication (Zollman, 2007). The characterization of the scientist as the isolated researcher in the laboratory making discoveries is a misleading myth.

Hume's Reductionism

Surprisingly, the philosopher often cited as the father of the epistemology of radical empiricism, David Hume, recognized that testimony was an indispensable source of information, and he offered an alternative to Locke's skepticism.

Hume broke with Locke, and commented in his *Enquiry Concerning Human Understanding* (1748) that:

There is no species of reasoning more common, more useful, and even necessary to human life, than that which is derived from the testimony of men and the reports of eye-witnesses and spectators... It will be sufficient to observe, that our assurance in any argument of this kind is derived from no other principle than our observation of the veracity of human testimony, and of the usual conformity of facts to the reports of witnesses. (Hume, 1977, p. 74)

This position, that the reliability of testimony must be inferred from past experiences of “conformity” between what a speaker has said and observations of the accuracy of those assertions, has more recently been labeled “reductionism” in the epistemology of testimony. This is the view that hearers can only acquire knowledge from testimony by

“reducing” the justification for believing what is said to a justification from a non-testimonial, more clearly reliable source of knowledge. Hume's suggestion is that non-testimonial sources, specifically perception and induction, allow an individual to have a degree of justification for believing testimony equivalent to the degree of justification conferred by believing what can be induced from past experience.

Whether this reduction is performed on a “global” scale or must be performed for each specific speaker, or testimonial context, is a matter of debate. “Global” reductionism argues that testimony is proven generally epistemically reliable based on reflection on the widespread reliance on, and success of, believing testimony in the past. More commonly, reductionists argue for a “local” justification of testimony, either by having evidence for the reliability of individual speakers, individual acts of testimony, or by having evidence for the reliability of testimony in specific contexts (see Falkner, 2000).

Reid's Nonreductionism

Far more common in contemporary approaches to testimony has been to reject reductionism in favor of viewing testimony as a direct source of knowledge. The historical expression of this view is due to Thomas Reid, who is regarded as the historical counterpart of Hume, and provides the first expression of the competing position, called “nonreductionism,” in the contemporary debate. Reid, like Descartes, also considers the contribution of the testimony of caregivers to early childhood learning; but unlike Descartes, Reid assumed that learning from parents, even prior to a child's cognitive development, must result in knowledge. If this were not the case, then humans would

know far less than they think they do, both as children, and as adults reliant on their early childhood education. Reid says:

The wise author of nature hath planted in the human mind a propensity to rely upon human testimony before we can give a reason for doing so... If children were so framed as to pay no regard to testimony or authority, they must, in the literal sense, perish for lack of knowledge. I believed by instinct whatever they told me, long before I had the idea of a lie, or a thought of the possibility of their deceiving me... And although this natural credulity hath sometimes occasioned my being imposed upon by deceivers, yet it hath been of infinite advantage to me upon the whole; therefore, I consider it as another good gift of Nature. (Reid, 1983, p. 281)

Reid seems to offer a pseudo-theological argument similar to the “benevolence” argument in *Meditation IV* (1641) that Descartes uses to deduce the general reliability of human perceptual faculties. Descartes argues that it follows from the existence of a benevolent God that human perceptual faculties are generally reliable. The argument goes that, since every person simply believes what they perceptually experience as true, then if perceptual faculties failed to be reliable, God, who designed those faculties, must either be deceiving us about the world, or must have created humans with fundamentally faulty perceptual faculties. Given that God is benevolent, Descartes argues, God would neither systematically deceive us, nor would he create us with faculties that make systematic errors. Therefore, Descartes concludes, human perception must be generally reliable in just the way people commonly presume.

Reid seems to think that testimony can be justified in a similar way, assuming that the immense benefits of being an audience to testimony cannot be an accident, but is instead a gift of the “wise author of nature.” Note that Reid is convinced of the reliability of testimony even when it is accepted with naive credulity, or when it is just believed out of “instinct,” because without testimony being generally reliable, we would “perish for

lack of knowledge.”

This “wise author of nature” type of argument may seem quaint, but it is surprisingly similar to contemporary arguments that attempt to prove the reliability of testimony, and more generally, the reliability of biological traits. This type of argument proceeds from an assumption about the function of that trait, and then concludes that the trait must reliably perform that function to explain its existence and persistence in the population.

In the contemporary epistemology of testimony, it is dogmatically assumed that language is for spreading true beliefs, and that therefore, the “wise author of nature,” or evolution, would guarantee the reliability of this capacity. Otherwise, it is argued, there would be no explanation for the existence and persistence of the use of testimony. This is, however, merely a type of adaptationist story telling widely regarded as suspect in biological reasoning. Chapter 7 of this dissertation examines the literature on the evolution of language to see if there are good reasons to think that the biological function of human language is to reliably transmit factually accurate beliefs. Even if this were the case, authors like Ruth Millikan (1993) have argued that the trait may fail to presently reliably perform this function.

There is, however, a key additional assumption of Reid's optimism about the general reliability of testimony that will dog us throughout, and which we must address in the end. Reid is expressing what Jennifer Lackey (2008) has recently called the “Infant/Child Objection,” that young children lack the cognitive capacities required to make the inductive inferences Humeian reductionism requires. If Hume is correct, we would have to concede that children, if they are in fact underdeveloped in this way,

would not know what their caregivers tell them. Almost all contemporary epistemologists have proceeded from the assumption that in mundane cases where caregivers tell their children facts, and those facts are true, the child comes to *know* what they are told. The implication is that this optimism would extend to beliefs that adults accept credulously.

We will explore this assumption more in Chapter 4, but here recognize the implications of assuming this. If knowledge can be acquired by credulously accepting testimony, even when the hearer would be incapable of detecting errors, then all credulous acceptance of testimony, that was known by the speaker and is true, must count as knowledge. This would seem to imply that it is rational, or a good epistemic practice, to simply believe what people say unless there is reason not to. This view is the most common among contemporary epistemologists of testimony, and dangerously threatens prescribing norms that will lead to widespread gullibility. Consequently, this optimistic view about the “wise author of nature” will be a primary target of what follows.

Having said this, I want to point out that Reid, Hume, and Descartes have expressed, even in the short quotes I have provided, a fact about testimony that I will develop throughout the dissertation: that our motivation to credulously accept testimony results from the practical necessity of believing others. I agree with Reid that testimonial interactions manage the practical and social necessities of human affairs. What is questionable is whether testimony must reliably transfer true beliefs to do so, or whether it might be possible that our affairs may be managed, or even improved, by believing testimony that is unconfirmed or even false.

Reid and the nonreductionists are not alone in thinking that the practical necessity

of testimony in human affairs implies that testimony must be generally true. This is the “wise author” intuition laid bare. Hume equally assumes that nature has provided humans with a reliable truth-securing practice. Returning to what I already quoted, “There is no species of reasoning more common, more useful, and even necessary to human life,” and consequently, it must be the case that very often, we can be “assured” of the “conformity” of testimony with the facts. Again, this move has been made from reflecting on the *practical* importance of human communication, which seems to motivate Hume to infer that testimony must be generally *epistemically* reliable.

However, no argument is given that factual accuracy is necessary for practical success, and it may be that when accuracy is irrelevant, it would be inadvisable to expend the resources to check for accuracy. If correct, it would seem more accurate to say that without credulously accepting the testimony of caregivers, human children would “perish for a lack of *information*,” and not “perish for a lack of *knowledge*.”

Reid's view has become the primary contemporary competitor to Hume. Labeled “nonreductionism,” this is the view that the say-so of others has an epistemic status equivalent to direct sources of beliefs, such as perception or memory, and therefore does not require that the hearer have positive evidence for the reliability of a speaker in order to gain knowledge.

Radical Translation and the Reliability of Testimony

One contemporary source of confirmation for the “wise author of nature” assumptions about the function and reliability of testimony has been the pervasive view among twentieth century philosophers of language that shared stable semantic meaning

requires generally true communication. This view owes its philosophical origins to reflections on W.V.O. Quine's (1960) famous thought experiment on the possibility of "radical translation." How, it is asked, can an individual infer the meaning of an expression with no prior knowledge of the new language, and only the behavior of the speaker as an indication? Answering this question is supposed to uncover the fundamental logic of the emergence of language. The basic problem that radical translation must address is that an individual contemplating what another individual means cannot assign meanings to a speaker's utterances without knowing what the speaker believes, whereas, at the same time, one cannot identify the speaker's beliefs without knowing what the speaker's utterances mean.

Donald Davidson (1967, 1973), following Quine, suggests that the hearer must interpret the speaker's beliefs by comparing the accompanying actions of the speaker, in the empirical context, to what the hearer would be thinking in such cases. The interpreter must use a "principle of charity" to maximize agreement with the speaker by assuming that what the speaker is saying is true. Language must be, as Tyler Burge (1993) puts it, merely a system of "content preservation," where the function of the system is to faithfully transfer content, allowing the hearer to attend to the same fact as the speaker.

The story goes that, having formed a perceptual belief about the world, the sender then performs an ostensive behavior that can be perceived by the recipient. The natural salience of the act, for example, pointing at what the sender is talking about, allows the hearer to consult their prior beliefs about the object, and infer that the sender must believe what they would believe in the same situation. The meaning of that behavior is then inferred to have the content of the prior belief of the hearer. Content has been preserved

between the mental state of the sender and the mental state of the hearer through an act of “agreement.” This view implies that communication requires fundamental trust on the part of the recipient (this is what Davidson's claim of “charity” amounts to), and that semantic meaning, therefore, would be destabilized by the inability of the observer performing that translation to read the mind of the communicator by employing the tacit principle of charity.

Notice, however, that this view does not imply that communication is in fact true, only that the participants agree that it is true. The claim that instability would ensue if a significant amount of communication was false is incorrect. What this view implies is only that communication would be impossible if a hearer could not identify what a speaker says as having the content of some prior belief. It is possible that there could be shared systems of false beliefs that are stabilized by some value they have that is independent of factual accuracy. For example, it might be that there are beliefs that it is valuable to share whether they are true or not. Consider the variety of competing world views that are inconsistent with one another. At most, one of those sets of beliefs could be true, and very likely they are all false (Downes, 2000, p. 441), yet humans may very likely gain social and practical benefits from sharing cultural beliefs with other members of their social group.

Latter responses to the possibility of radical translation argue that the relationship of agreement is too strong, because it rules out the possibility of a hearer understanding a claim with which they disagree. H.P. Grice (1969, 1975) and his colleagues argued instead that the logical relationship between the beliefs of the speaker and the prior beliefs of the hearer should be one of logical relevance and not identity. Other theorists,

notably Sperber and Wilson (1995), have agreed that a hearer can infer the meaning of an utterance by recognizing its *relevance* to the truth of prior beliefs, even if the utterance is inferred to hold an evidentiary logical relationship to some prior belief, or counts against those beliefs.

The difference between the Davidsonian and Griceian stories is that where Davidson claims that shared meaning requires fundamental trust, Griceians argue that this trust need not be “blind” (Sperber et al., 2010, p. 363). Instead, hearers can exploit practices that make epistemic judgments about what is said based on various logical relations to what is already believed. Interpretation does not require agreement, but instead requires inferring the intentions of the speaker to affect the hearer's thoughts and actions.

The effect of these theories on the epistemology of testimony has been to reinforce the assumption that, first, language is exclusively for the transmission of content, and second, that shared meaning and language could only exist and persist if it did so reliably. What I wish to point out is that these theories from the philosophy of language rely on transcendental arguments. Both are premised on a reflection about what seems like the obvious function of a part of human mental life, that language produces beliefs, and go on to infer something about the reality of language, that language would be unstable unless it reliably accomplished the goal of getting people accurate beliefs. However, this transcendental argument is merely another just-so story about the function of a biological practice that, when combined with confidence in the “wise author of nature,” generates optimism regarding the reliability of the practice.

In Part II of the dissertation I will challenge the “wise author of nature” and

transcendental arguments in two ways: First, these arguments are worse than adaptationist stories when they postulate the function of language without even taking seriously their biological origins. Many epistemologists and philosophers of language have ignored the fact that traits evolve because of their contribution to an organism's fitness, and it is unclear what the relationship is between the factual accuracy of information and an increase in fitness that results from the use of information systems. Second, as already mentioned, just because a human trait seems to have a specific function today does not mean this is the function responsible for its emergence and persistence. Traits that reliably performed some function in the past may fail to reliably perform that function in a new environment, and the informational environment has been changing rapidly.

We will often return to comment on the impact of this history, because it has framed the contemporary debate that follows, and is the origin of numerous assumptions that deserve scrutiny. We now proceed to the contemporary debate, and begin with C.A.J. Coady's description of how we should define and demarcate testimony from other speech acts. For purposes of pinning down the subject matter of our epistemic investigation, I stipulate a much more general definition of "testimony" as our target. I will then proceed to a description of the theories of "nonreductionism" versus "reductionism," and will conclude this part of the dissertation with a brief examination of Lackey's "dualist" theory of the requirements for testimonial knowledge.

CHAPTER 3

THE NATURE OF TESTIMONY

Formal Testimony and Natural Testimony

Much of the contemporary interest in the study of testimony as a source of knowledge was stimulated by C.A.J. Coady's work, *Testimony: A Philosophical Study* (1992), and so we will begin an examination of the recent debate with Coady's theory of testimonial knowledge. We begin not with a theory of what is required for interactions to be considered epistemically reliable, but instead, with Coady's theory about how we should demarcate genuine testimony from other kinds of communicative acts.

I begin with Coady's analysis for two reasons; first, because it has been a common practice to substitute an analysis of what testimony is with an examination of when, as humans, we accept testimony as reliable; and second, because Coady's confidence that "testimony" is a very common, reliable type of speech act, falls at the extreme of optimistic confidence in the "wise author of nature."

Coady thinks that "natural testimony," the kind of common speech act he considers is typical of informative communication, is an extension of "formal testimony," which is the say-so of individuals called upon to attest to true information in a formal institutional setting. An example of formal testimony would be when someone is asked

to provide true information before a court of law or board of inquiry. Coady observes that cases of formal testimony are characterized by an assumption that the “witness” has some special expertise on the facts in question, and that the truthfulness of the witness offering formal testimony is explicitly demanded.

Coady distinguishes testimony in these settings from other types of speech acts, such as offering a closing statement, by pointing out that an attorney offering a closing statement is not expected to be making claims as evidence for the facts, nor are the standards of truthfulness explicit as in the case of the witness. Coady presumes that formal testimony is just an institutionalized version of the informative speech acts that make up a large portion of human communication, and that speech acts of this sort are extremely common, saying, “On my account the scope of testimony is very wide and testifying is not an arcane procedure restricted to the law courts but a very fundamental act we engage in many times a day” (Coady, 1992, p. 54). Coady then goes on to characterize testimony in the following way:

So we may take it that the conventions governing the speech act of testifying ('natural testimony') are specifiable as follows:

A speaker *S* testifies by making some statement *p* if and only if:

- (1) His stating that *p* is evidence that *p* and is offered as evidence that *p*.
- (2) *S* has the relevant competence, authority, or credentials to state truly that *p*.
- (3) *S*'s statement that *p* is relevant to some disputed or unresolved question (which may, or may not be, *p*?) and is directed to those who are in need of evidence on the matter. (Coady, 1992, p. 42)

Notice that requirement (1) and (2) are both epistemic conditions. They claim that this “very common” type of speech act both provides evidence for the hearer that the claim is true, merely by virtue of the speaker so testifying, and that “natural testimony” is offered

by a source that has genuine authority. Given these stipulations, any speech act that meets requirements (1) – (3), and was in fact true, would be, merely in virtue of satisfying the requirements for being “testimony,” a reliable source of knowledge.

Notice that for testimony to be a reliable source of knowledge on Coady's account, the hearer need not do any “work” to verify the reliability of the speaker. Specifically, the hearer need not perform any checks of the source's reliability since the act of testifying itself supplies the evidence of its accuracy. This is the move that Locke rejected by claiming that “even when true,” the reception of beliefs from other people reduces “science” to “opiniatrey.”

If, as Coady suggests, natural testimony is a common version of formal testimony, then the hearer can simply assume that the speaker is considered to be offering authoritative testimony on information on which they have “special expertise.” That an individual testifies is evidence that they know what they are claiming and therefore evidence that the claim is true.

The claim that this kind of testimony is “very common” displays Coady's commitment to common speech acts being generally reliable, and this implies an informational environment where hearers can come to know what they hear without doing any work. The problem, then, for anyone concerned with judging the reliability of what someone says, would be to identify whether they are confronted with a genuine instance of “testimony” instead of some other communication that does not qualify as this rarefied practice.

I think the most interesting question in the epistemology of testimony, and the one being answered by this dissertation, is, “when should an individual, oriented towards

getting the truth, believe the testimony of others?” For Coady to both stipulate that “testimony” is epistemically reliable, while also claiming that testimony is a common practice, seems to justify widespread credulity. If declarative speech acts are generally testimony, as Coady seems to suggest, then the solution to the question is to simply believe what everyone declares. Certainly, whether requirement (2) holds of any given act is a matter that needs proof and requires significant understanding of the context of communication and the expertise of the source.

The problem is diagnosed by Lackey and Fricker (1995) as, “confus[ing] the *metaphysics of testimony* with the *epistemology of testimony*” (Lackey, 2007, p. 16, her emphasis). Coady, they claim, has mixed up what testimony “is” with whether human communication, more broadly construed, reliably results in true beliefs. I would put it less obscurely by saying that the way to judge the reliability of what people say is by attending to key signals of factual accuracy, and not by attempting to judge if what is said is “testimony.” What I want to know, and what I hope epistemologists are trying to illuminate, is when it is justified to believe what other people say, and how to avoid false testimony. For the consumer of testimony, Coady has provided no test for accuracy and instead only encouraged confidence in credulity, confidence that would only be warranted by a good survey of the epistemic practices of speakers.

The situation is even worse, however, because clearly, if the hearer could know that conditions (1) - (3) were met, they would have very secure grounds for believing what was claimed, and, if the claim turned out to be true, would have acquired knowledge (on almost any analysis of knowledge). The problem is that in almost any normal case of being an audience to testimony, the competence of the speaker is not guaranteed, and

whether the speaker is competent is exactly what the conscientious consumer of information needs to know. How are these facts to be ascertained? Astonishingly, the dominant view in the contemporary debate, “nonreductionism,” assumes that requirements (1) - (3) are met a large percentage of the time, and it is rational for the hearer to assume this without employing any tests. In the same way that memory is a “credentialed” source of knowledge, so is testimony. We will investigate this argument in more detail in the next chapter where I will conclude that we should evaluate the epistemic merits of various sources of beliefs independently.

“Testimony” as a Suitable Target of Epistemic Evaluation

My reaction to Coady is not to argue for another “correct” theory of what testimony is. The choice of demarcations between types of speech acts is not an important epistemological question. It may be possible to analyze “testimony” as the term applies in a context of its natural use, but this is not the topic of the dissertation. My goal is not to characterize what “testimony” is, but to offer advice for when people should believe what they hear.

I will therefore merely give a principled stipulation of what testimony is to clarify the target of the epistemic evaluations I will make later. The first principle is that I want to avoid Coady's error in demarcating testimony epistemically, and instead, demarcate testimony in an epistemically neutral way. That is, I do not want to require that for an act of human communication to be testimony it must satisfy epistemic standards of reliability, allowing for the epistemic evaluation of the reliability of testimony to be independent of any metaphysical investigation into different kinds of human

communication.

The second principle is that testimony should include any human communication subject to epistemic analysis, so that the evaluations I will propose at the end apply to the broadest set of interactions possible. This principle does impose some constraints on the subject of our concern, since the topic of classic analytic epistemology has been the status of beliefs. Other types of knowledge such as the assessment of skills or familiarity with people and places has been largely ignored. I propose we therefore pick out the target of evaluation as any human communication where the receiver is judging whether to form the belief that what is communicated is the case. “Testimony,” is all human communication whose content can be believed by an audience (one or more other individuals).

I do not mean to include beliefs that might be formed as a result of what is communicated that are not beliefs about what is communicated. For example, I might form the belief that someone is an idiot based on what they say, but this is not a belief in the accuracy of what they said, so we should limit our target to only those beliefs that contain the content of what is communicated. This is not dissimilar from theories of reductionist like Fricker (1995, p. 396-397) and Audi (1997, p. 406) who demarcate testimony as “tellings generally” or “people telling us things.” We can now move on to a discussion of our first theory that directly addresses the reliability of testimony.

CHAPTER 4

NONREDUCTIONISM

Nonreductionism and the Default Rule

Coady's theory and the claims of nonreductionists agree that testimony is a generally reliable “direct” source of knowledge. By “direct” it is meant that, similar to perception or memory, a hearer of testimony can be epistemically justified in simply believing what they hear without identifying any positive evidence for what they hear.

The question that both reductionists and nonreductionists are debating is, “how do people acquire justified beliefs from testimony” (Lackey, 2011, p. 73)? Both assume that this is not only possible but very common. Nonreductionism is the view that a hearer can be epistemically justified in believing what they are told even if they have no good reasons to think that the speaker is reliable, just as long as they have no good reasons to think that the speaker is unreliable. So, for instance, Matthew Weiner concludes that, “(w)e are justified in accepting anything that we are told unless there is positive evidence against doing so” (Weiner, 2003, p. 257), and Burge writes that, “(a) person is entitled to accept as true something that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so” (Burge, 1993, p. 467).

I have mentioned that this is the most common view about the epistemic status of

testimony and also the one I will attack the most vehemently. As evidence that this position is widely held, the following is an incomplete list of adherents, including, Austin (1979), Robert Audi (1997), Alvin Goldman (1999), David Owens (2000), Ernest Sosa (2006), and many other prominent epistemologists.

A popular way that the epistemology of testimony has been framed is in terms of a “default rule.” The so-called, “epistemological problem of testimony,” asks how can it be that an individual can be epistemically justified (i.e., justified in a way that if a belief is true it would count as knowledge), while following what has been called the “default rule.” The default rule is a norm that states that it is rational for an individual to simply believe what they hear unless they have good reason not to. The view is that hearers of testimony can be epistemically justified in believing what they hear by “default,” and by this they mean that acceptance is automatic and error checking is only required if there is some good reason to think the speaker is unreliable. To say it another way, the “default rule” makes it epistemically justified to simply believe people without having any positive reason to think they are reliable. As Lackey (2008) puts it, nonreductionism does not require that the hearer do any “work.”

As with Coady's theory, nonreductionism places all of the epistemic “work” for guaranteeing the accuracy of testimony on the speaker, and comes very close to the same requirements on epistemically legitimate acceptance that Coady required for a speech act counting as genuine testimony. Coady labeled speech acts as “testimony” if the speaker was a genuine authority in making the claim. Nonreductionism also requires that for a hearer to acquire knowledge, the source must have known what they are claiming, even though the hearer need not know this to acquire knowledge. Unlike Locke, the

presumption is that the act of testimony itself is sufficient justification for the belief in question in the same way that the act of seeing some state-of-affairs is sufficient for believing that the world is as perceived.

In addition, nonreductionists assume that speakers do know what they testify to most of the time, presumably because, as discussed in the philosophy of language literature, the stability of language requires generally honest (and here I mean factually accurate) communication, along with fundamental trust on the part of the receiver.

Nonreductionism argues for two claims that will be challenged in Part II of the dissertation, but that I preview here. They assume that humans very often do just “default” believe and simply accept what they are told without positive evidence for doing so, and that it is rational to do so, implying that this practice very often leads to knowledge. The theory of nonreductionism integrates both a descriptive and a normative claim into an argument for the general reliability of testimony. However, the descriptive claim, that humans default believe while monitoring for defeaters, is often not supported by empirical evidence, but is more often just regarded, based on reflection, to be an obvious fact. Is it not the case that children and even adults very often simply believe what they are told?

In more detail, “default believing” involves automatically accepting what is heard, while monitoring for defeating reasons. That is, there is an additional automated process that awaits signs of unreliability in a speaker, such as insincerity, deceit, or a lack of authority. When these signs are “monitored,” the automatic acceptance is defeated, and rejection or additional judgments are possible. If nonreductionists did not assume that people process information in this way, then their arguments for the claim that following

the default rule is rational would fail to sanction actual practice, but the stated goal of the epistemology of testimony is to describe how it is that people are justified in believing what they hear. So, the ability of nonreductionism to answer this problem relies on it being the case that people really do very often process testimony in a way described by the default rule.

This is a poor way to engage in an inquiry into the epistemic status of a source of information. Not only is the reliability of actual practice already assumed, presumably because of some simple reflection on the function of testimony, and an argument based on assuming the “wise author of nature,” but in addition, philosophers have relied on mere reflection to describe actual human practice.

The goal of the rest of this chapter is modest. What follows is an attempt to merely call into question the assumptions of the epistemology of testimony and nonreductionism. It will be sufficient if the reader no longer thinks it is obvious that the use of testimony persists because it generally produces true beliefs, and that humans actually process information in a way described by the default rule. The real challenge based on the empirical literature appears in Part II, and provides the new account of actual practice and the considerations important to judging reliability, which will serve as the new axioms for the epistemology in Part III.

Introducing the Nonepistemic Uses of Testimony

Let us start with some considerations that make it possible for testimony to serve functions that are independent of the transmission of true beliefs. Example 1: Testimony could exist so that individuals can share beliefs and goals to facilitate affective bonding

between individuals. For example, you may like someone and be more inclined to be amicable towards them because you have the same political or religious beliefs. Jack says, “Republicans are hopelessly corrupt,” and this act of testimony influences Mary to like Jack more because she agrees. Example 2: Testimony is necessary for sharing the beliefs and goals required for cooperative action. Take for instance the conquistadors who, if they had not all believed in the riches of Eldorado, would not have sailed to the new world, or the commitment to ending the evil of Hitler that motivated the Allied troops of WWII, or the faith in a mandate from God that spurred on the Crusades. Example 3: Testimony is necessary for humans to alert other individuals of our beliefs and goals to facilitate planing. “I’ll be at the theater at 8:00 pm.” Example 4: Testimony is necessary for offering altruistic support to a partner in an interdependent relationship. “You can do it honey, you’re the best in your class.”

The point is that the relationship between the success of testimony in facilitating these other outcomes and the success of testimony in transmitting factually accurate beliefs is complex, and depends on the specific use in a certain context. It is not obvious that the value of using testimony that explains both its emergence and persistence is solely that it allows the transmission of *true* beliefs. It is at least possible that these non-truth-related outcomes may often be more important than factual accuracy.

Let me diagnose for a moment what nonreductionist might be thinking, and then I will make a very important point about the value of testimony. I have already stipulated that our discussion is about the type of human communication, “testimony,” that results in the reception of new beliefs. So, there is no disagreement about the function of testimony to transmit beliefs. Typically, the “good” evaluative property associated with

beliefs is “truth,” so it would be forgivable to conclude that, for the use of testimony to persist, it must reliably generate *true* beliefs.

The Value of Shared Beliefs

However, this overlooks a key difference between testimony and other sources of beliefs. Specifically, it overlooks that testimony not only results in new beliefs but it can result in *shared* beliefs. What I am proposing is that it is possible that the “good” outcome, that explains the existence and persistent use of testimony, may in some cases at least, be the production of *shared* beliefs, and not *true* beliefs. Burge argued that testimony is for content preservation, and this may often be the case. What he failed to realize is that content preservation may be enough in many circumstances to justify communication, regardless of the factually accuracy of what is communicated.

Consider another key argument for nonreductionism. Many nonreductionists have tried to explain how beliefs that result from testimony are justified by arguing that testimony has the same epistemic status as other more direct sources of knowledge, specifically perception and memory. The current trend is to view testimony as most epistemically similar to memory (see Burge, 1993, 1997; Schmidt, 2006). Individuals do not come to believe what they remember on the basis of having good positive reasons to think any specific instance of remembering is reliable, and yet, they assume, humans certainly can come to know what they remember.

People do not evaluate everything they see and remember; they seem to automatically accept what they see and remember unless they have good reason not to, that is, they accept by “default.” If children can come to know testimony in the same

way, then it is tempting to think that testimony has the same epistemic status. Memory has been the most widely used faculty for comparison because memories have propositional content and can be seen as instances of telling ourselves things that previously happened. Memory is almost like a kind of internal testimony.

However, memory and testimony are fundamentally different in a key way that may have influenced their function, and their epistemic reliability. Both perception and memory are faculties that individuals use to collect information; the result of both are personal beliefs. Testimony, on the other hand, has an intrinsically social character. By this I mean that, unlike memory or perception, humans tell other people what they do to change the mind of the hearer, specifically to make the hearer think in the way the speaker is currently thinking.

As stated before, this fact may bring a significant challenge to the intuition that the “good” outcome of testimony is solely to transmit true beliefs. Instead, the “good” outcome of testimony may often be to get the hearer to think in a similar or compatible way. Recall the examples of possible nonepistemic functions of testimony, including transmitting the necessary beliefs and goals required for coordination, bonding, and planning. In these cases, the benefits of shared beliefs accrue not only to the hearer but also to the speaker, and this is important to explain the reason a speaker may be willing to incur the cost of producing the testimony. Speakers reveal what is in their interest and this makes testimony very different from the impartial environment accessible through perception or even the contents of memory.

Default Acceptance and Memory

Before we move on, I want to quickly say something about the descriptive claim made by nonreductionism. Do people process testimony by employing a practice described by the default rule? Again, I just want to challenge the dogma that this is so. Clearly, this is an empirical question about human psychology, and as such, should not be answered by mere common reflection. If social psychology and social economics have proven anything in the last 40 years, it is that the workings of the human mind are very different than they seem to the individual user.

As evidence, consider an example relevant to our discussion of nonreductionism: memory. Increasingly, experimental evidence has shown that even memory can be highly unreliable, as in instances of memory “biases.” Two common biases include the use of the availability heuristic and memory construction (Tversky & Kahneman, 1974). Use of the availability heuristic occurs when an individual judges the likelihood of an event by remembering similar events in the past. Since only a selection of experiences are remembered, memory is an unreliable test for an events likeliness. For example, people will tend to remember events that are slightly out of the ordinary making them, on recollection, believe that those types of events are more common than they are.

Memory construction occurs when an individual comes to remember an event that did not occur usually because they heard about the event from someone else. Bruck and Ceci (cited in Melnyk, 2007) performed a study in which they read preschoolers stories of fictitious happenings and then questioned them about those events a week later. In this study, 58% of preschoolers reported remembering the events as if they had actually experienced those events, and many would even embellish the stories with vivid

recollections of what happened.

What this shows is that what humans take on reflection to be clear cases of reliable belief formation have been shown, with more rigorous empirical observations, to be mistaken. For this reason, I submit, the question about how humans actually process testimony should be left to the psychologists. I will survey this literature in Part II.

CHAPTER 5

REDUCTIONISM

Hume's Reductionism

The historical roots of reductionism come from Hume's claim, already quoted, that the reliability of what other people say can be, “derived from no other principle than our observation of the veracity of human testimony, and of the usual conformity of facts to the reports of witnesses” (Hume, 1977, p. 74). Contemporary epistemologists have labeled this type of practice a “reduction,” because the individual justifies the “veracity” of testimony by using nontestimonial sources of epistemic justification. For Hume, the specific sources of nontestimonial justification are the observation of the facts to which prior testimony has “conformed,” and inference, specifically an induction from a sample of reliable testimony to the reliability of a present instance of testimony. Succinctly put, humans can come to know what they are told when they have good inductive evidence from testimony in the past being reliable.

The claim that humans must inductively infer the reliability of testimony is however a vague requirement, since it is unclear what breadth of testimony can be justified by a single induction from past experience. For example, Hume could be claiming that each instance of testimony must be justified by a separate inductive

inference that takes as its premises a sample of past observations of successes at being told the facts by that individual in a similar context. This is the most common view of reductionists in the contemporary literature, that each instance of testimony must have a good positive reason for thinking it is reliable, and is known as “local reductionism” (for examples see Fricker, 1995; Falkner, 2000, discussed below).

Alternatively, the reliability of testimony in general might be inferred from a much broader sample of observations of factually accurate testimony. At the other extreme is the view that we can have “global” justification for all testimony based on a sample of all past testimony that allows an inference to the general reliability of testimony. There are other possibilities as well, including inferring the reliability of the testimony offered by a specific individual or inferring the reliability of testimony in a specific type of context, each from a different representative sample of observed successes. What follows is a critical examination of each of these types of reductionism.

Global Reductionism

If Coady and nonreductionists are the most liberal and optimistic about the general reliability of testimony, “global” reductionists are only minimally more discerning. Although this theory is not widely held, it is instructive to point out its similarities to nonreductionism, along with the shared reasons for rejecting these theories. A global reductionist, who thinks we can be generally credulous, but only because testimony has been observed to be a generally accurate source of information, like the nonreductionists, accepts that human testimony is generally reliable. They differ in believing that some positive reasons are required for accepting this, presumably reasons

that all humans acquire early in their experiences with testimony (see Adler, 2007).

Most theorists favor nonreductionism because of the intuitions that small children and even credulous adults regularly get knowledge from testimony. It seems counter-intuitive that an individual must provide some inductive justification for every piece of testimony they accept. Since people do not, in fact, regularly make a justifying inference, reductionists must concede that humans know very little of what they have been told. The benefit of global over local reductionism is that an individual need not go to the effort of justifying each instance of testimony, because the justification of each instance comes from a “standing” justification of testimony generally. By “standing,” I mean that the collection of the sample, and the inference that concludes that testimony is generally reliable, could be performed only once in the past and yet apply to judgments about the reliability of any future instance of testimony.

Objections to Global Reductionism

There are significant problems for global reductionism, often brought by non-reductionists, but which, I will argue, count against both theories. First, it is unclear whether any individual has the breadth of experiences of factually accurate testimony to justify testimony generally (see Lackey, 2011, p. 76). What would seem to be required for testimony to be justified generally would be that the individual had witnessed the conformity of testimony to the facts in a very broad range of situations on the whole breadth of topics humans communicate about, so as to have a sample of experiences that meets inductive standards of being representative of the set of all acts of testimony.

Coady writes:

it seems absurd to suggest that, individually, we have done anything like the amount of field-work that (reductionism) requires...(M)any of us have never seen a baby born, nor have most of us examined the circulation of the blood nor the actual geography of the world...nor a vast number of other observations that (reductionism) would seem to require. (Coady, 1992, p. 82)

The problem raised by Coady, Lackey, and colleagues, is that no individual human has the sample of experiences required to premise a cogent induction from their past experiences of the reliability of testimony to the conclusion that testimony is generally reliable. The available sample will always be unrepresentatively small and biased.

Another way “local” reductionists have framed the same problem of collecting a representative sample is to point out that testimony does not look like a homogeneous “kind,” with respect to making epistemic judgments. Certainly, the epistemic status of testimony varies by context. Fricker writes:

looking for generalizations about the reliability or otherwise of testimony... as a homogenous whole, will not be an enlightening project. Illuminating generalizations, if there are any, will be about particular types of testimony... (Fricker, 1994, p. 139)

Introducing a term of art, let me phrase this objection as observing that testimony is “epistemically heterogeneous,” which means that simple observation reveals that within the set of all cases of testimony, there is significant variability as to the epistemic status or reliability of different instances of testimony. There are instances or kinds of testimony that are reliable and others that are obviously unreliable.

Testimony is epistemically heterogeneous because the expertise of the individual speaker can vary significantly and because features of the context, such as the relationship between the speaker and the hearer, are relevant to the epistemic status of testimony. Consider the cases of someone reporting the time on their watch in

comparison with the case of a used car salesman making claims about the reliability a vehicle. Certainly the epistemic status of these reports are too varied, and dependent on properties of the context, for both cases to be included in a sample that would justify testimony as a whole. The typology of testimonial contexts I introduce at the end of this dissertation is an attempt to describe this heterogeneity in a way that can improve epistemic judgments about any given instance of testimony.

Epistemic Heterogeneity and Nonreductionism

Let us now return to the nonreductionist who, like Coady, simply accepts that credulity, even naive credulity (where, like a young child, the hearer has not developed good error checking practices), can result in the hearer knowing what was said. To be precise, the hearer only comes to know what they accept when the speaker knew what they are claiming, and when the hearer has no good reason to think the speaker is unreliable. These are the additional requirements for nonreductionists, but are assumed to be satisfied in a high proportion of cases. How can nonreductionists, like Coady, deal with the epistemically heterogeneous nature of testimony?

Coady and colleagues cannot claim that testimony is universally reliable while still taking seriously their own objection to global reductionism, because recognizing the epistemic heterogeneity of testimony would seem to be a good reason to be concerned about the reliability of any given instance of testimony. Recall that nonreductionism has a “defeater” monitoring condition that assumes humans can pick up telltale signs of inaccuracy that become good reasons to reject testimony. For nonreductionists, a hearer is only justified in believing what they hear if they have no good reason to suspect that a

given instance of testimony is unreliable. Knowing that the reliability of testimony is highly variable and context-dependent would seem to qualify as a standing reason for skepticism; after all, ignoring this fact and credulously accepting everything people say, politicians, psychics, and salespersons included, seems highly gullible.

My claim is that attending to the fact that the reliability of testimony is highly variable and context-dependent, facts that we are all usually aware of, the individual would have a good defeating reason to be skeptical. If this is correct, then any consumer of testimony with a realistic understanding of its actual fallibility would be a defeater-defeater, or a good reason to think that although the epistemic status of testimony is highly variable, none-the-less the instance of testimony under evaluation is reliable. This of course is exactly the requirement demanded by local reductionism.

Moreover, notice the following perplexing fact about nonreductionism. If the individual did not attend to the epistemic heterogeneity of testimony, that is, they simply ignored the obvious fact that the reliability of testimony is highly variable, then they would be aware of no good defeating reason, and if the speaker knew what they testified too, the hearer would come to know what they hear. If correct, the epistemic status of a source of information, in this case testimony, improves because of the epistemic agent's ignorance of epistemically relevant facts about the nature of that source, specifically that it is highly variable. Of course, ignoring relevant epistemic facts should not make an epistemic agent more justified in their beliefs.

This is an especially surprising admission for nonreductionists, since their entire justification for the general reliability of testimony follows from widely held assumptions about the nature of testimony. The assumption that testimony would only exist and

persist if it was generally reliable seems in direct contradiction to Coady's admission that testimony is epistemically heterogeneous. Notice that the seeming inconsistency between optimism in the reliability of cognitive practices and the observed fact of widespread error is exactly the problem this dissertation is intended to solve and one Coady seems to ignore.

The moral to the story is that good epistemic practice must involve the consumer of testimony discerning epistemically reliable from epistemically suspect testimonial contexts, instead of practicing a faithful ignorance of the epistemic heterogeneity of testimony. However, the next theory of “local” reductionism asserts the need for positive evidence for each instance of testimony, so there are no “standing” justifications, even for discernible contexts.

Statistical Local Reductionism

“Local” reductionism requires that an individual have positive reasons to think a speaker's testimony is reliable in order to be justified in believing any instance of testimony. This is the most popular type of reductionism, since it avoids the objections to the global variety, and it requires practice that is astute to the topology of the testimony's epistemic heterogeneity (examples of local reductionism can be found in Adler, 2002; Falkner, 2000; Fricker, 1994; Lehrer, 2006). Local reductionism, like reductionism in general, has come in a number of varieties that disagree about the degree of evidence that justification requires. At the extreme of strict prescriptions for consumers of testimony is the view that justification requires that a hearer have a statistically representative sample of past successes of testimony to justify each instance

of testimony. By this I mean that the consumer of testimony would have to meet the standards of induction described in canonical statistics by meeting the standards of sample size and representativeness required of cogent inferences from observed properties of a sample to properties of a population. I will refer to this version of local reductionism as, “statistical local reductionism.”

The advantages of statistical local reductionism are that it requires consumers of testimony to be astute to the epistemically heterogeneous status of various instances of testimony, and that it requires humans to practice in a way commonly endorsed as a canon of ideal rationality. It is immediately obvious, however, that this strict requirement has rather severe implications: First, humans would in fact know almost nothing, because they have never performed the cogent inductions required. Second, prescribing that individuals practice statistical local reductionism would not be useful since almost no one could or would practice statistical local reductionism.

Humans would know almost nothing because although they are massively dependent on testimony for what they believe, they have never inferred the reliability of speakers in the required way. Second, imagine trying to teach people to only believe what is justified by the rules of statistical sampling that appear in a statistics textbook. Humans neither have the time nor the training to make such inferences, and, even if they did, it is dubious whether they have stored the accurate observations of testimony being accurate in the past that would apply to just the context of testimony in question that would make up a representative sample.

This problem is made worse by the following observation that much testimony is not first-hand, but has been passed on through a chain of speakers from some original

source. In cases like this, is it sufficient to only infer the reliability of the immediate speaker on the assumption that they have passed on true information in the past, or would the hearer instead have to identify the original source and infer their reliability from some sample of observations about the original source's reliability? Again, it seems unlikely that any individual has or could do this regularly.

Permissive Local Reductionism

Local reductionists, like nonreductionists, agree that testimony must very often be a reliable source of factually accurate information and are therefore uncomfortable with the radical skepticism that would accompany statistical local reductionism. In response, local reductionists have argued for less strict requirements. For example, Elizabeth Fricker remarks:

In claiming that a hearer is required to assess a speaker for trustworthiness, I do not mean to insist, absurdly, that she is required to conduct an extensive piece of MI5-type “vetting” of any speaker before she may accept anything he says as true... My insistence is much weaker: that the hearer should be discriminating in her attitude to the speaker, in that she should be continually evaluating him for trustworthiness throughout their exchange, in the light of the evidence, or cues, available to her. This will be partly a matter of her being disposed to deploy background knowledge which is relevant, partly a matter of her monitoring the speaker for any tell-tale signs revealing likely untrustworthiness. (Fricker, 1994, p. 149-150)

It is unclear, however, that this concession is sufficient to either allow much testimony to be justified, or offer a more employable practice for improving performance at judging testimony. This is still a prescription that would require considerable effort and skill.

Surely, Fricker must insist that this kind of discrimination be done correctly and that the individual does not evaluate in a way that mistakes the significance of evidence

and cues, or is based on illegitimate stereotypes and biases. If humans, as I will argue in Part II, are heavily influenced by biases that diverge from good epistemic practice, then as in the case of statistical local reductionism, Fricker must admit that a large proportion of what humans have accepted from testimony would not be justified. This might be acceptable to Fricker, however, who is more suspect than nonreductionists of the reliability of testimony.

The real question of interest then would be what specific strategies can be described that are employable for improving evaluations of testimony. All we have been given is to consider “evidence, or cues,” by “deploy(ing) background knowledge which is relevant,” and “monitoring the speaker for any tell-tale signs revealing likely untrustworthiness.” These prescriptions would be unreliable if in employing background knowledge people are simply polarized to believe what is similar to what they think they already know, and if they regularly fail to apply evidence correctly, or if they are poor lie detectors. In Part II, I will provide evidence that actual human practice is susceptible to each of these failings. Good epistemic practice in Fricker's view would therefore require significant training. My goal in Part III will be to describe some general principles about testimonial interactions that offer relatively easy checks for accuracy based on identifying the general context.

I would like to point out one other fact about Fricker's suggestion that will be relevant to a better understanding of the nature of testimony. Notice that what Fricker requires seems to fly in the face of common standards of etiquette, as it would be very common for an individual to take offense at realizing that a friend is allocating significant energy to checking for the reliability of what they say. The hearer could of course

evaluate in secret and not reveal their skepticism, but this is not what people usually do. After all, testimony, as I will argue, is the medium by which social groups are formed and maintained, and therefore skepticism and dissent are very often seen as offensive attitudes only suitable for an audience in conflict with the speaker. This is why questioning what someone says is so often met with hostility, and is considered tantamount to calling the speaker a liar, because questioning the claims of others rejects solidarity in favor of social divisiveness.

In Fricker's defense, she can require only that the hearer of testimony perform this active epistemic checking when they are primarily concerned with factual accuracy, and she may admit there is nothing "irresponsible" if an individual cares less about the factual accuracy and more about getting along well in the world, and with other individuals. I am sympathetic to prescribing additional accuracy checking when an individual is oriented towards epistemic accuracy, but think this is better done as a caution against being misinformed, instead of as a prescription for generating positive evidence of reliability. In addition, I will argue that the key to ease of checking should focus on checking the context to identify possible nonepistemic motives for testifying that might result in testimony being unreliable.

Paul Faulkner recognizes the possibility, crucial to my thesis, that humans may often have practical instead of epistemic motives for their use of testimony. Faulkner writes:

Given that a speaker's intentions in communicating need not be informative and given the relevance of these intentions to the acquisition of testimonial knowledge...[i]t is doxastically irresponsible to accept testimony without some background belief in the testimony's credibility or truth. (Faulkner, 2000, p. 587)

It is unclear, however, whether this is a necessary or a sufficient requirement on having epistemic justification. Faulkner therefore only requires that an individual have some background beliefs that offer evidence of the speaker's credibility and not that each specific speaker's testimony be justified by induction.

What I take Falkner's suggestion to be is that a hearer must have reason to think the speaker is attempting to offer factually accurate information, or, more accurately, that the speaker is telling the hearer things in order to provide true information and not, for example, to manipulate the hearer, generating solidarity for cooperation, trying to emotionally bond with the hearer, or any other among many nonepistemic uses of testimony. What we need is a more nuanced prescription for judging the various uses of testimony.

Conclusions

It is very common for individuals to just reject needing any evidence, or good reason, to think the sources of their beliefs are reliable. For philosophers concerned with the epistemic status of testimony, this should rightly generate concern over what accuracy can be expected. The correct response, however, is not to succumb to optimistic assumptions that humans must know much of what they are told because of their massive dependence on testimony, along with considerations like the inescapable credulity of childhood.

Instead, we must recognize that credulity may be common because there is value to agreement as a practical and social necessity, and not because humans do or have

existed in an informational environment of generally true communication. In Hume's words, there is no past experience of getting information that seems, “more common, more useful, and even necessary to human life,” than testimony, but, as with contemporary epistemologists, he assumes that the “usefulness” of testimony is derived from its factual accuracy. This assumption, however, fails to take seriously the evolutionary origins of human communication, and testimony's role in the social nature of the human species.

PART II

THE SCIENCE OF HUMAN COMMUNICATION

CHAPTER 6

INTRODUCTION

In the last section, we saw that philosophers concerned with the epistemology of testimony assume a number of facts about the function of testimony, how humans process testimony, and about the epistemic status of testimony in a broad set of common situations. In this part of the dissertation, I will challenge the just-so story that the usefulness of testimony is to offer true beliefs, and challenge the assumption that human practices evolved to efficiently share the facts. My goal is to replace a priori theories and common reflections with contemporary theories about the biological origins of human communication and the psychology of human practice.

The questions, how did human communication emerge and what stabilizes the persistent use of testimony by humans, are questions about the natural history of human communication, and cannot, therefore, be answered by reflection and a priori theorizing. Instead we must examine human communication as a capacity made possible by successful traits that allow for the learning of linguistic practices, traits that have their origins in gene and cultural co-evolution.

The question, how do humans produce and process testimony, is clearly a question about human psychology. Rigorous empirical experiments are therefore the only

available technique for collecting the evidence required to provide an accurate description of both the complex practices and motivations of speakers offering testimony and the remarkably flexible practices of hearers processing testimonial information. In the last few decades, social psychologists and behavioral economists have uncovered significant information about human judgment that is both inconsistent with the assumptions of the epistemology of testimony, and provides a more nuanced account of when and why people say and accept what they do.

Steve Downes following David Hull described a similar commitment when he said, “I ... intend to make sure that the discussion of truth and selection is informed by evolutionary biology and not merely the collective imaginations of philosophers” (Downes, 2000, p. 426; Hull, 1989). I extend this commitment by making sure that the description of actual human practices that will be the target of epistemic evaluations comes from rigorous experiments on human behavior, and not just the collective imaginations of philosophers.

In Chapter 7, I will challenge the first two claims, that language exists and persists because it distributes true beliefs, and that it reliably does so. My goal will be modest, since I only require demonstrating that communication *could* evolve because of benefits conferred that do not require that testimony be generally true. In fact, the biological story of how human communication emerged and why it persists makes no mention of “truth” at all. If correct, the assumption of the epistemology of testimony that testimony must very often be factually accurate will be false. In addition, the epistemological problem of testimony that asks how it is that testimony is reliable will have a false presupposition. It may be that testimony is only reliable in a narrow range of contexts, and so the more

useful question will be, when is it the case, if ever, that testimony should be considered reliable to a truth-seeker.

The reliability of beliefs offered as testimony depends on the actual practices of humans who have formed beliefs from testimony and then pass those beliefs on. If it was found that humans regularly diverge from good epistemic practices when deciding what testimony to accept, this would be evidence that much testimony is unreliable, at least in the contexts where these divergences are detected. The best way to identify the practices of humans in various contexts would be to examine empirical experiments that provide evidence for a description of human regulatory practices.

In Chapter 8, I will survey the social psychology of information regulation. First, I will describe experimental evidence that challenges the claim that humans “default” believe. If you recall, the epistemological problem of testimony is often framed as the problem of describing why following the default rule, to credulously believe testimony while monitoring for defeaters, is rational. If humans do not employ practices described by the default rule, then this is an uninteresting question since a solution would tell us nothing about humans.

This will be the first step in offering an empirically informed and nuanced account of how humans react to testimony. What the psychological literature shows is that humans have a very complex set of practices that, in many cases, are surprisingly ill-suited to avoiding epistemic errors. Instead, humans are very flexible in their attention to factual accuracy, and this flexibility allows them to trade-off the effort for good epistemic practice in favor of satisfying other important needs. It appears as if an individual's practical and social success can even be impeded by excessive attention to the truth.

CHAPTER 7

THE EVOLUTION OF HUMAN COMMUNICATION

We are anything but a mechanism set up to perceive the truth for its own sake. Rather, we have evolved a nervous system that acts in the interest of our gonads, and one attuned to the demands of reproductive competition. If fools are more prolific than wise men, then to that degree folly will be favored by selection. And if ignorance aids in obtaining a mate, then men and women will tend to be ignorant.

– Ghiselin (1974, p. 126)

Introduction

This chapter argues that the emergence and stability of human communication, and language specifically, does not *require* that communication be generally factually accurate. It is not only possible but plausible that the use of testimony evolved and persists because it provides benefits that may be independent of or even impeded by traits tailored solely to spreading the truth.

I will be telling a story about the evolution of signaling that has developed in the Signaling Theory literature over the last few decades. This story is a contemporary biological alternative to the a priori philosophical theories of Quine and colleagues on the origins and nature of human communication. I will also take note of important lessons about the relationship between factual accuracy, and the other fitness and utility

enhancing benefits of the use of language.

From Inferential Translation to Strategic Signaling

David Lewis' (1969) use of signaling games to investigate the emergence of shared meaning represents the transition from reflection on Quine's thought experiment of the radical translator, to the biological approach to meaning as a strategic convention. Lewis investigated the origins of shared meaning by considering a two-player game where a sender had various strategies for behaving in overt ways as a reaction to various states of the world, and a receiver practiced a strategy for reacting in specific ways to the various behaviors. What Lewis demonstrated was that, if some regular combination of signal and response was mutually beneficial to both the sender and receiver, then the behaviors involved constituted a stable equilibrium of game strategies. That is, the signaling and receiving strategies are such that no player could deviate from those strategies and do better.

In the semantic tradition of Quine and his followers, the “meaning” of a propositional form of communication was the “intention” of the proposition, or what the proposition was “about.” Propositions were about what was in the content of a declarative sentence. As already described in Chapter 2, language was seen as a tool for sharing this intentional content, and required that a hearer infer the content of the corresponding beliefs of the speaker from their ostensive behavior and the prior beliefs of the hearer.

My only goal in summarizing this view is to compare it to more recent biological descriptions of meaning, starting with Lewis' conventions in a signaling game. For

Lewis, meaning was not the intentional content of a proposition but instead the behavioral regularity between a sender and receiver that produced a shared increase in utility. What made a signal more than mere noise was that through the regularity of signaling behavior and reaction, there was a benefit conveyed that explained the emergence and persistence of the interaction. “Meaning” in this sense did not exist independently of a social interaction with mutual benefits. Meaning was not, for example, a representative feature of language that was then passed through a tool for “content preservation.”

If the utility benefits of coincident sender and receiver strategies are interpreted as increasing the reproductive fitness of participants, then the simple Lewis signaling game offers a plausible description of how communication evolves. The basic phenomenon described by Signaling Theory involves an organism that is regularly changed by specific environmental conditions, where that change in constitution or behavior modifies the activity of another organism. If another organism is affected by this change in behavior of the first, then it is possible that the change in the receiving organism might increase the fitness of both organisms. When this occurs, the regular change in the first organism is called a “signal” because it prompts a mutually beneficial interaction between the two organisms. If the sender-side and receiver-side strategies are either heritable or could be learned, presumably through reinforcement, then the signal/reaction regularity can go to fixation as a signaling system.

Theorists attempting to explain the evolutionary emergence of meaning, notably Maynard Smith, Ruth Millikan, and Brian Skyrms, have endorsed the view that the “intentionality” of propositions is just the evolutionary benefit of that communication

form. Maynard Smith writes,

In biology, the use of informational terms implies intentionality, in that both the form of the signal, and the response to it, have evolved by selection. Where an engineer sees design, a biologist sees natural selection. (Maynard Smith, 2000; also see Skyrms, 2010, p. 43)

The “aboutness” of a proposition is its historic contribution to the Darwinian fitness that is responsible for its continued use. Bryan Skyrms (2010) gives a more technical description of biological meaning in terms of the types of information contained in a signal.

There are two kinds of information in a signal that contains all of the “meaning” of the signal, and both types of information introduce changes in probability assessments for the sender and receiver. First, there are the probabilities that each of a set of states of nature was observed by the sender. This first information is the “content” of the signal, and contains information about the relationship between signaling behavior and the observed states of the world that produce them. Second, there are the probabilities that a receiver will perform any of a set of acts in response to the signal. This second information is the regular reactions of receivers to the signal. Together, these two types of information describe sender and receiver strategies within a signaling game (see Skyrms, 2010). The “intention” of a signaling interaction is the increase of Darwinian fitness that accrues from the regular use of mutually beneficial signaling strategies. If the use of these strategies is heritable or can be passed on through instruction, then their use can reach an equilibrium throughout a population as a stable signaling system.

Clearly, the ability of humans to use a specific set of signals as a natural language is not an evolved trait, but is learned. Only the ability to *learn* languages has evolved, a

benefit that has allowed for great flexibility in signaling behavior and also, as we will see in the next chapter, great flexibility in the practices used for processing signals.

A goal of this chapter is to drive home the significance of the fact that testimonial traits and practices evolved and are therefore explained by their contribution to the fitness of participants and not primarily by a truth-spreading function. To extend this point to even learned human language, I want to point out that the learning process is itself regulated by reinforcement of linguistic practices. Canonical theories of reinforcement learning, such as Roth-Erev or Bosh-Mosteller reinforcement, have been shown by Skyrms (2010) to behave like Darwinian selection regimes. The intention of a propositional form of human communication can be understood as the contribution of that form to the utility of the participants, which accounts for the reinforcement that explains its continued use. My point is that even the practices used in communication that are learned would exist and persist because they were practically valuable (i.e., offered utility), and not solely or primarily because they spread the truth.

The more complete story of the evolution of testimonial communication is that, at some point, traits for fixed signaling strategies gave way to more flexible traits for learning languages. The specific signaling and response strategies are then stabilized by their utility benefits through reinforcement learning during development. It has been suggested by Lachmann (2001) that this shift occurred when human predecessors moved from costly signals that are prevalent among other animals, to the very cheap signals of combinatorial language available to humans. I will wait to discuss this more until I have surveyed the significant literature on the relationship between signal cost and the benefits of signaling.

The goal of Signaling Theory has been to determine when selection regimes will produce an equilibrium of adaptive strategies for both sender and receiver under various conditions. When both sender and receiver use signaling and response strategies that result in fitness increases that outperform any available invading strategy, then a signaling system has emerged. It is the selection of a signaling system that explains the emergence and persistence of signaling behavior, including human language and the use of testimony.

According to this story, “meaning” is a property of all stabilized biological interactions, and these interactions are necessary for all fitness-enhancing, symbioses, cooperation, and so on, both between what are typically considered individual organisms (e.g., two birds producing mating calls) and also between what are typically considered the constituents of organisms (e.g., cells within the immune system producing and responding to protein concentrations).

In the next section, I address one complication of explaining human language as a signaling system. As Jason Noble points out, “The chief problem of a Darwinian account of human speech,... is the apparent level of altruism involved” (Noble, 1998, p. 1). The objection is that humans, unlike other animals, use a complex combinatorial language that is very cheap to produce. As I will discuss, this is difficult to explain evolutionarily because the stability of signaling systems is dependent on participation producing an adaptive advantage for both sender and receiver. If signal production costs are very low, then misleading signaling can be advantageous. However, if signals are misleading, then they destabilize the systems because receivers would do better by ignoring the signal or deviating from their current strategy. What I will describe are the conditions that would

allow for stable signaling that benefited both receiver and sender and the requirements for stable signaling in a low cost signaling regime.

Adaptive Signaling

Because the stability of signaling systems relies on signaling behavior increasing the participant's fitness, the central question of signaling theory has become, under what conditions would selection produce signaling systems where the sender would incur the cost of signal production while providing valuable signals? In this literature, the term used to refer to valuable or misleading signals is that the signal is, "honest," or "dishonest." However, this use would be very misleading for my purposes because the goal is to reinforce that the signaling is selected for its adaptive advantage and not its factual accuracy. Therefore, I will break with the use of the terms "honest" and "dishonest" referring instead to talk about "valuable," "valueless," or "harmful" signaling.

The first point to consider is that signals that fail to benefit the recipient will destabilize the system by making defection on the part of the recipient a more advantageous strategy. Consequently, harmful signaling would lead to instability in the system as receivers who accepted harmful signals would be at a selective disadvantage. If a sufficient proportion of receivers begin to ignore signals, then the benefits to the sender through modifying the receivers' thoughts and behaviors will fail to outweigh the cost of signal production.

Certainly, factually accurate signals may offer valuable benefits in this way by providing the recipient with additional accurate information with which to make

decisions; however, the question is why senders would offer hard-won and costly-to-produce signals to benefit other individuals with whom they compete. This type of altruistic scenario is inconsistent with signaling that results from natural selection.

It would be surprising, then, that communication would evolve as an epistemic tool used for passing on truths. A system like the one proposed by Burge for “content preservation” is evolutionarily unstable, because there is no explanation for why anyone would offer testimony in the first place. A key insight of the co-evolutionary hypothesis is that sender costs can be explained by the benefit to the sender by influencing the actions of the receiver (see Noble, 2000). Language use would not be altruistic in the way described in the thought experiments of Quine and colleagues.

The Evolution of Manipulation

Dawkins and Krebs (1978) argued that the capacity for persuasion and the capacity to detect the value of signals should co-evolve, a suggestion that dominates current thinking about signaling systems. As in other “red queen” scenarios such as predator/prey or pathogen/host co-evolution, better adapted persuasive skills increase selection pressure on the capacity to check for errors, while improved vigilance puts additional selection pressure on the capacity to “manipulate” the receiver.

“Manipulation” in signaling does not always involve the morally suspect human behavior of trying to mislead other people. The use of the concept of “Manipulation” throughout the rest of this dissertation means what it does in signaling theory, which is the neutral modification of another individual's thoughts and actions. Technically, “manipulation” means altering the likelihood of the various actions available to the

receiver as a result of the receiver receiving the signal. Every genuine signal, therefore, “manipulates” the receiver, because any act that is a communicative act must produce a reaction in the receiver. This is true for the human use of testimony as well.

I defined testimony in Chapter 3 as any communicative act that produces beliefs in the hearer. This belief-modifying function of testimony is a type of “manipulation.” It should be pointed out, however, that when humans offer testimony to change the mind of someone, they most often do so because of the additional effects on the future behavior of the hearer. Humans tell each other things to “try” to change each other's minds, so as to “try” and change each others behaviors. I have placed “try” in scare quotes because it is very often the case that humans are not subjectively aware that they are being manipulative. It is only in some cases that humans reflect on what they want to say to bring about a specific modification in the thoughts and actions of the audience.

Whatever additional explanations can be offered for incurring the costs of signal production, they all require the manipulative efficacy of the signal. If there is no manipulation, then there is no benefit possible, as the signal will have had no effect. To say it a different way, any signal that fails to manipulate is not a genuine signal, but only noise. A story that accounts for signal production costs will, therefore, describe how the manipulative effects of the signal have benefited the sender in a way that makes the cost of signal production worthwhile.

Signaling in Contexts of Coincident Interests

Contexts of coincident interests are ones where the success of the recipient benefits the sender. In these contexts, there is a straight trade-off between the cost of

signaling and the increase in fitness that results from benefiting the receiver. There are a variety of ways that a sender can come to have coincident interests with the receiver, and each provides one among many explanations for beneficial signaling.

First, coincident interests can occur through kin selection effects (Hamilton, 1964). If the receiver is closely related to the sender, then increases to their fitness are simultaneously increases to the senders. A very important type of context of coincident interests for humans are contexts where signaling allows for coordinated or cooperative activities. If two individuals working in a coordinated way can amplify their fitness beyond what each could achieve independently, then the cost of signals that manage such coordination may be worth it. In the next chapter, I will discuss the psychology of “ingroup” and “outgroup” biases. An “ingroup” is a set of individuals who depend on each other for success. Members of ingroups are individuals with coincident interests.

It may also be that humans signal to produce a type of cooperative activity that no members of the group are aware of. Hunters that signal to each other with the intent to coordinate for the purpose of flushing and killing an animal are involved in reciprocal manipulation for purposes of explicit coordination. However, individuals in many systems can aggregate to accomplish shared goals without knowing that they are doing so, or even being aware of what those goals are.

There is a type of dynamic system called a “distributed autonomous system” that is characterized by constituents acting independently, often in probabilistic or random ways, and, over time, “self-organizing” themselves into functional wholes that efficiently direct resources to do work. By “self-organize,” I mean that there is no executive control or “organizing” that directs the constituents. These systems organize themselves by

imitation, where the partial success of some constituent in performing some action results in other actors picking up and following their “trail.” In doing so, resources are concentrated towards the completion of the task. In this way, the earlier, seemingly random efforts “self-organize” into directed allocation of resources for accomplishing that task.

For example, when ants forage for food, they begin by running in random directions away from the nest until an individual finds food and then returns that food to the nest, leaving behind a pheromone trail. Other ants who, just by chance, cross the trail, are induced to follow the trail to the food source and then return to the nest, and, in doing so, strengthen the trail for others to follow. In this way, very simple autonomous actors behaving in initially random ways structure the overall foraging system to efficiently collect food (Holldobler & Wilson, 1990).

The immune system works in a fundamentally similar way (Segel & Cohen, 2001). Antibodies for most previously encountered pathogens circulate in similar quantities until a pathogen enters the body. When, by random chance, a correctly configured antigen encounters the pathogen, the antibody reacts by activating the production of pathogen-specific killer cells. In this way, a previously undirected set of immune cells can become an efficient directed immune response. With no executive control, the immune system has its resources allocated towards the efficient elimination of the specific threat.

In another example, Melanie Mitchell (2001) and her colleagues have produced a computer program that uses randomly generated minicomputational devices, called “Codlets,” that, over time, structure a work space to compute analogy problems. When

given the problem, if ABC is to ABD, then MNO is to what, and the computer can correctly compute “MNP” (and much more complicated examples). Mitchell's contention is that the human brain might actually work as a distributed autonomous system to solve complex problems through the self-organization of simpler computational constituents.

The revolutionary activism currently being witnessed in the Middle East may result from a similar type of self-organization. Autonomous actors reacting to environmental cues organize into aggregates capable of efficiently accomplishing very difficult political goals. There is an “expression” cascade that concentrates political resources to topple an entrenched autocracy.

With so many examples of distributed autonomous systems in nature, and such a varied set of phenomenon across scales and function, it is plausible that evolution commonly selects for constituents that can aggregate in this way. Each constituent is benefited by the dramatic increase in effectiveness available as a result of resource aggregation.

My conjecture is that testimony, and human communication in general, is a medium that allows for humans to aggregate into functional wholes by imitating individuals who display partial success. It may even be that whole societies and cultures have emerged in this way. If this is correct, then part of the story of the evolution of language is that it provided the medium by which societies and cultures could emerge from the aggregation of simpler constituents through imitation and reinforcement. The individuals involved may not even know what goal their aggregated resources are being put to, or, as in the case of many social causes, the expression cascade can become

bureaucratized by vanguards, or a set of members taking over executive control.

In this section, I have consolidated numerous properties of human social environments that produce coincident interests between senders and receivers that account for signaling costs. It seems clear, however, that humans communicate in situations where there is no coincident interest, and we therefore require further explanations for how signaling in these contexts could be stabilized by mutual benefits.

Contexts of Noncoincident Interests

A context of noncoincident interests is one where the sender is not benefited by the receiver being benefited. This kind of situation is typical of competitive interaction, and is, therefore, typical of the interaction of individuals undergoing selection. Humans, however, communicate with unrelated individuals who they may never meet, and when individuals are in competition. How can the stability of communication in these contexts be explained?

Researchers have identified a number of ways that valuable signaling can be stabilized in contexts of conflicting interests. For example, valuable signaling can be stabilized by a bet on reciprocity, so that the immediate cost of signaling must be absorbed with only a possibility of future benefits (Trivers, 1971).

Benevolence can also be stabilized by an increase in the intrinsic cost of signal production, as long as the cost is correlated with some quality being communicated. The most common example in the literature is of the peacock, whose flamboyant tale signals the ability of the bird to “waste” resources on its production. This was first explained by the “handicapping” metaphor of Zahavi (1975, 1997) and Grafen (1990), who proposed

that resource-intensive signals displayed quality in a way similar to knowing a horse's or a golfer's handicap. How elaborate the tail is directly corresponds to the ability being signaled.

Veblen (1899) likened this type of signaling to conspicuous consumption in humans, where wealth and status are signaled by prominently displaying wealth and signs of status, the acquisition of which require sufficient wealth and status. Of epistemological importance is that these signals will tend to be more accurate, in a factual sense, because they are directly correlated to the qualities being displayed.

An example from human linguistic interactions is where expending the additional effort to produce detailed, well-argued, or technical descriptions is indicative of education or training, and therefore factual accuracy. For example, if an acquaintance offers a very detailed and consistent description of how a nuclear reactor works, this is direct evidence that they have special knowledge of how a nuclear reactor works. Although this is some evidence for the factual accuracy of testimony, it is not conclusive. Merely pedantic communication, as, for example, in the elaborate descriptions of astrologers, need not correlate with factual accuracy.

In addition, there are two extrinsic, or externally imposed, sources of increased signaling cost that operate in tandem to stabilize beneficial signaling. The first is effective vigilance against harmful information. If receivers can reliably detect harmful signals, then the cost to produce such signals will be lost by the sender if the receiver can detect and resist being manipulated. Notice that the receiver-side detection capacities that co-evolved with manipulative techniques would have evolved, not to judge factual accuracy per se, but to judge the beneficial or harmful nature of signals. They would

have evolved to judge benevolence or malice as opposed to truth or misinformation. In Chapter 8, I will introduce evidence that real human practices are more tuned to judging the benefits of accepting testimony than judging its factual accuracy.

Working along with effective vigilance to increase harmful signal costs is the ability to report on an individual's reputation. When receivers are good at detecting and ignoring malevolent signals, the cost of misleading signals can be massively increased by a system of effective punishments for malevolent or negligent signalers (see Searcy & Nowicki, 2005; Lachmann, 2001, for a formal treatment). Human language is very often used for reporting on the reliability of others, while labeling someone a “liar” or “user” can have very dramatic social effects.

Punishing perpetrators of harmful persuasion is necessary both for stabilizing beneficial “manipulation” in contexts of conflicting interests, but also for the development of complex combinatorial language. Lachmann, Szabolcs, and Bergstrom (2001) have shown that signaling behavior that evolved because of mutual benefit for senders and receivers could not employ the arbitrary signals used in human language unless enforcement reliably increased signal cost.

In the handicapping case, signal cost was sensitive to the quality being displayed, but human language uses very low cost signals. Saying something takes little effort, and the effort required is becoming vanishingly small with advances in information technology. Some other influence must stabilize valuable signaling. The claim of Lachmann et al. is that the signal receiver must impose a social cost when signaling cost is low. This means that the transition to cheap, arbitrary learned signals relied on enforcement. This enforcement is commonplace in teaching children to speak by simply

praising and disagreeing with proper and improper use of words and phrases.

Of epistemological importance is the observation that, in contexts where signals are cheap and there is little or no accountability for harmful signals, there is nothing in the context of communication to prevent harmful manipulation. In Part III, I will use this fact as an important diagnostic principle for identifying contexts where testimony may be harmful. However, harmful signaling and false signaling are not coextensive. There are cases where a sender can harm a receiver by telling them the truth, for example, when someone is made passive to being controlled by reiterating their actual failures and damaging their self-esteem.

Conclusions

What I have provided in the preceding discussion is an alternative to philosophical theories about the origins and persistence of the use of human language. I have offered very little in the way of arguments that the evolution of signaling story is correct. What I am committed to is just this: if we take seriously the biological origins of the traits used in human communication, then something like the evolutionary account found in the signaling theory literature is required, and, most importantly, the evolution of signaling implies no requirement on testimony that it be generally true.

Richerson and Boyd (2005) have claimed, however, that the relationship between the value of human communication and its factual accuracy is one where accuracy is generally necessary to explain the benefits of language. Their argument is that language and culture evolved to allow the flexibility required for humans to live in the wide variety of environments that they have learned to occupy. The story is that the evolution of

genetically acquired adaptations could not keep up with the rapid migration of humans into new territories. The abilities to survive in any given new environment had to be learned and rapidly passed on to offspring during development. Language, on this account, is an adaptation for passing on information about the specific physical environment that cannot be acquired through genetic inheritance. This is an example of a valuable use of testimony, navigating the physical environment, that would require accuracy, and if this story described the only function of language responsible for its evolution, then it would be reasonable to think that testimony would have evolved to be generally true.

My claim is that although language may have evolved because it played a significant role in allowing humans to teach knowledge necessary to survive in one of a variety of environments, it is plausible to think there are many other valuable uses of testimony as well. I have already described numerous uses of testimony, such as producing solidarity in social groups and manipulating behavior, that are independent of reporting on the environment. For example, if people use language to form coalitions by having others attest to beliefs required for membership, then it is plausible to think the uses of testimony are varied and may or may not require accuracy.

Whether the benefits of communicating requires factually accurate testimony is therefore a question that must be answered for each type of context where communication occurs. This is because the relationship between the need for the truth and the value that testimony provides is complex. There may be contexts where testimony being true is necessary for that testimony to produce the intended benefits (e.g., when reporting on the local physical environment while undertaking a cooperative

endeavor), but there are also contexts where accuracy is either irrelevant or may impede the production of the intended benefits (e.g., when telling a friend they are more competent than they are to pump up their confidence).

There are many important lessons from signaling theory that will inform the epistemological discussion to come, and so, in conclusion, I summarize them here. First, nowhere in the explanation of human communication as a signaling system is truth mentioned at all. The behavioral regularities involved in communication are stabilized because they regularly increased the fitness or utility of participants, and were therefore selected for or had been reinforced during development. The claim that the benefits of signaling could not arise unless signaling consistently produced true beliefs is far from guaranteed and certainly not true in general.

Second, if the evolutionary story is correct, humans would not tell other humans things to simply pass on information. In one way or another, the goal of telling is to “manipulate” the thoughts and actions of the audience. It is consistent with this fact that humans will very often be unaware of the manipulative effects of their testimony. In addition, the cost of offering testimony can only be explained if the effect of manipulating the audience regularly outweighed the cost of signal production.

Also, since response behavior co-evolves with sender behavior, it should be expected that since sender behavior is manipulative, response behavior is tuned to judging the desirability of going along and being modified by the signal. The regulatory traits used by humans confronted by testimony would be tuned to judging the benefits of acceptance, and not solely or even primarily the factual accuracy of the testimony.

Lastly, there is “balance” required between the effectiveness of the capacity to

persuade on the speakers side and the capacity to avoid being duped on the receiver's side. If one of these capacities, persuasion for example, becomes more effective faster than new variations and selection can increase the effectiveness of the other trait, we can expect language to be used in malicious ways. To balance the use of misleading testimony when the cost of saying things is very cheap, and especially in contexts of conflicting interests (e.g., competitive contexts), effective vigilance combined with enforcement is required. If either is absent or insufficient, the manipulative use of language could result in widespread misleading testimony.

CHAPTER 8

THE PSYCHOLOGY OF INFORMATION REGULATION

[I]f (as Dawkins argues) deceit is fundamental in animal communication, then there must be strong selection to spot deception and this ought, in turn, to select for a degree of self-deception, rendering some facts and motives unconscious so as not to betray – by the subtle signs of self-knowledge of the deception being practiced. Thus, the conventional view that natural selection favors nervous systems which produce ever more accurate images of the world must be a very naive view of the mental evolution.

- Trivers (2006, p. xx)

Introduction

In the last chapter, I argued that there is nothing in an evolutionary explanation of human communication that requires testimony to be generally factually accurate, even if in specific contexts the value of communicating would require accuracy. The implication of the fitness-enhancing origins of animal communication is that the traits responsible for producing and processing signals would not necessarily be tuned to judging factual accuracy, and humans would deviate from good epistemic practices if some other practice was more beneficial. In this chapter, I survey extensive studies of actual practice that involve just such deviations.

Part of my goal in this chapter is to explain why humans deviate from ideal epistemic practices, so let me begin by listing three types of answers available. Some

epistemic errors may result from the malfunction of what would otherwise be a well-functioning system for making epistemic judgments. The individual being observed may have a mechanism responsible for their practices that does not operate in the way that the vast majority of other individual's mechanisms operate, and this deviation is grounded in a physiological defect or damage to the mechanism.

Another possibility is that deviations are best explained as limitations on cognitive systems that were selected for because they were the best design that emerged from the natural causes of cognitive variation. It is clear that evolution does not result in perfect traits, but only the best ones available. Cognitive systems can also make errors because they operate in a context of limited information, and because of the need to be cognitively “miserly,” and process information with as little effort as will get the job done.

I will argue for a third, and what I consider more plausible, explanation that very often deviance from good epistemic practice is no error, but a prudent tradeoff for securing beneficial goals other than factual accuracy.

We begin with the common observation of epistemologists that humans very often simply believe what they are told. This practice was described by the “default rule,” which states that it is rational to automatically believe what we are told unless there is good reason not to. I will examine the evidence for and against thinking that humans do “default” believe, and then go on to sketch a much more nuanced description of actual human practice that has been described by social psychologists and behavioral economists.

Psychology and the Default Rule

The view I am arguing against, that comprehension requires acceptance, was not only assumed by philosophers, but given preliminary empirical support by the experiments of Gilbert et al. (1990, 1991, 1993). Subjects were shown unrecognizable statements that included Hopi words to control for prior knowledge, and then had the truth value of the statements flashed on a screen. In the experimental trial, an audible tone was played at the moment the truth-value was revealed. It was found that this distraction disrupted the ability of individuals to remember when statements were false but did not affect a hearers ability to remember when a statement was true. The interpretation of this result was that believing requires no additional effort and must occur during comprehension, while “tagging” and remembering a claim as false must require some extra effort to activate and operate an additional conscious error checking process.

However, more recent experiments, notably Hasson et al. (2005) and Clement (2004), have shown that automatic acceptance does not occur when statements are informative. In the first Hasson experiment, as in Gilbert's, tones were used to produce cognitive load, but during the experimental trial, subjects where shown statements that had been pretested as being informative. The pretesting involved asking subjects to rate how informative a statement was, such as, “this person walks barefoot to work,” about an anonymous 30-year-old man. The result was that for statements ranked as highly *uninformative*, the ability to remember the falsity of statements decreased as in the Gibson experiments; but for statements that were ranked as highly informative, subjects had no additional difficulty remembering which statements were false.

Acceptance, it seems, can be nearly effortless when statements are irrelevant, and

involves no awareness of a judgment of accuracy. When, however, a statement is informative, individuals do not “automatically” believe in this way and can make judgments of accuracy without first accepting what they comprehend. This suggests that a statement's informativeness, specifically the relevance of the content to another person, can alter a bias to simply believe what is said, prompting the individual to expend additional effort when making a judgment about person-relevant statements.

What the Hasson study demonstrates is that the processing, specifically the amount of effort spent to consider and remember its accuracy instead of being credulous, is sensitive to the relevance of information to judgments about other people and their environment. Humans seem to use the most efficient processing, and simply believe what they hear, when information is received in a context of no practical or social importance. This means that informational processing involves more than a judgment to accept or reject the content of a claim, but also includes a judgment about resource allocation based on an assessment of the practical and social importance of the information. The social psychological literature on attitude change has confirmed this repeatedly, while uncovering a typology of motives that provide a more nuanced description of this practice.

Motivation and Information Bias

In a study conducted by Lundgren and Prislin (1998), it was confirmed that at least three competing motivations strongly affect information processing in a social context. Lundgren and Prislin prepared subjects by telling them three different goals of the research study they were participating in with the intention of priming one of three

motives in the subjects. The three motives were towards defense of their own opinion, agreeableness towards the opinion of another person, or towards factual accuracy. In one case, the subjects were told that the study was to examine critical thinking skills; in another, that it was to study skills of interpersonal rapport; and in the third, that it was a survey of opinions on a relevant issue. The goal was to make subjects more motivated by accuracy in the first case, agreeableness (i.e., the impressions of another person) in the second, and defense of the subject's opinion in the third. Cialdini and Trost (1998) described these three motives as “managing the self-concept, building and maintaining relationships, and acting effectively” (Wood, 2000, p. 541), respectively, and this tripartite typology has become a standard in the social psychology of attitude change (in this literature, belief formation and modification is a primary type of “attitude change”).

The method was that after being primed, each subject was asked for their initial opinion on a debatable issue provided by the researchers: whether the school should increase tuition 30% in the upcoming year. Each participant was also informed of another confederate's opinion on the same issue. As a control, all subjects were told that the other individual was strongly in favor of the tuition increase. In addition, each subject was then given a number of different opportunities to collect relevant information on the merits of the tuition increase in the form of pamphlets and literature that equally represented both sides of the debate.

The results showed that subjects primed to be motivated towards accuracy selected arguments to read on both sides of the issue and tended to alter their opinions towards an attitude more neutral between the confederates and their own. Additional studies have shown that individuals motivated by accuracy can detect more and less

plausible arguments and will favor the evidence provided by good arguments over implausible ones (Chaiken et al., 1989) This is important because these results confirm that humans can practice better, even if not ideal, epistemic techniques when motivated to do so.

Moreover, subjects primed to be more defensive of their own opinions tended to only read arguments that seemed congenial to their prior point of view, and became more polarized, having their prior opinion strengthened (the effects of a defensive motivation are confirmed in Frey, 1986; Lundgren & Prislín, 1998). Subjects that were primed for agreeableness tended to read arguments in favor of the reported opinions of the confederate, and had their opinions moved closer to those of that individual.

In addition, subjects were later interviewed in private and it was found that their modified attitudes towards the topic were highly robust. Of special interest was the finding that the attitudes of individuals who had been primed for agreeableness were more persistent than the attitudes of the other groups. This suggests that the motive for maintaining relationships or being agreeable affects information regulation more strongly than being motivated by protecting one's own opinions or by a desire for accuracy.

Kruglanski (1989, 2004) has also produced a typology of responses to information that depend on the receivers psychological and social motivations. Kruglanski introduces four attitudes that affect information processing, including a high need to simply believe what people say (i.e., “close” on information), or “freeze” on what is already believed, to a high need to reject what people say or refuse to have an opinion.

For example, Kruglanski summarizes experiments where subjects that were primed for agreeableness (what he calls a need to “avoid nonspecific closure”) would

remain ambivalent on an issue even when they held prior beliefs and were confronted by relevant evidence (Kruglanski, 1989, p. 38-39). In two selection experiments by Peri, Kruglanski, and Zakai (1986), it was shown that subjects can be manipulated into avoiding choosing between a set of alternatives while also abandoning a significant amount of information processing. It was more important to seem congenial and diplomatic to the group than assert strong opinions, and individuals actually came to reject that an issue could be settled, instead focusing on reasons for skepticism towards all sides.

In another context, where many individuals held and expressed similar opinions, a situation commonly described as an “echo chamber,” many researchers, including Kruglanski (2004, p. 111-130) and Sunstein (2003), have observed that beliefs become polarized far beyond what available evidence would warrant. This “group polarization” phenomenon has been implicated in fanatical conviction, such as terrorist bombers, but is also common in social settings such as sporting events, religious gatherings, and political rallies.

Both being motivated towards ambivalence and the motivations that produce group polarization result in practices that deviate from the good epistemic practices observed by Lundgren and Prislin. In the first case, relevant information is discounted or ignored, and in the second, conviction rises dramatically even with no additional independent evidence being identified. Could these practices be explained as either a malfunction of an otherwise successful epistemic mechanism or the heuristic-like efficient processing of bounded cognition directed at the truth?

The malfunction explanation seems unlikely because the phenomenon described

seems, if not universal, at least very common. If it were the case that the practice of becoming polarized, for example, was a malfunction, then some plausible story would need to be told about how the vast majority of humans commonly malfunction in this way. More plausible would be the claim that these phenomena suggest the use of efficient heuristics (see Gigeranzer, 2008). Humans might process information by employing a rule like, “have conviction in what people say,” because most people speak the truth. What Lundgren and Prislun and Kruglanski have observed is the use of a variety of heuristics that are each activated by different motivations that depend on the social context, and some of these heuristics are better at detecting factual accuracy, by correctly weighing evidence, for example, than the others. When an individual is motivated by a desire for another person to have a good impression of them, a heuristic something like, “be open to new ideas, believe what 'that' person tells you,” might be activated. When the same individual is motivated to defend their opinion, a heuristic something like, “freeze on your prior opinion and disregard offhand what people tell you,” might be activated. Practices described by the use of these heuristics deviate from good epistemic practice because, as already mentioned, they fail to appropriately take into account available evidence, and lead to gullibility in the first case and dogmatism in the second. What these studies show is not that humans are incapable of practicing better strategies for judging accuracy, but that in certain contexts, humans “choose” not to.

It is far more probable, as the psychologists insist, that humans have evolved to substitute good epistemic practice for information regulation that helps “build or maintain relationships,” or “manage the(ir) self-concepts,” which influences behavior and practical success. In addition, there is a plausible explanation for why this would be the case, and

it draws on the evolutionary story from the last chapter. Humans who did better by acting on false beliefs would be selected for over individuals who insisted on the truth.

Unwarranted narcissism, for example, may be a very valuable attitude when trying to attract a mate. And, more than any other animal, humans massively depend on their relationships with other humans for their well being. It would therefore be no surprise that humans would evolve practices for dealing with information tuned to forming and maintaining beneficial social relationships, even if those practices conflict with the deployment of the best epistemic practices available to them.

Ingroup Bias

What I argued for in the last section was that, when confronted with some source *S* making a claim *p* in a context *C*, the processing of that information is not well described by the receiver questioning, “is *p* true?” Instead, the receiver comes closer to asking, “should I accept *p* from *S* in *C*?” And the answer is not merely influenced by epistemic considerations, but the needs of the receiver vis-a-vis their relationship with *S*, their needs to maintain a positive self-image, and possibly other competing motivations. Any one of these needs, or all of them, can contribute to the individual's judgments about accepting or rejecting *p* in *C*.

For example, the Lundgren and Priskin type studies suggest that whether a human is credulous depends on whether they are motivated towards building and maintaining social relationships. Whole disciplines in social psychology comprising vast numbers of empirical studies have suggested a “social adaptationist model” (Fiske, 2000, p. 305) of information regulation. The conjecture is that humans evolved to regulate effective social

coordination in face-to-face interactions (Caporael, 1997). Fiske writes:

At core, people are motivated to maintain affiliations and bonds with others,...Having understood (or having the sense of understanding) a person on whom one depends, one gets along by going along, that is, by reflecting the other person's beliefs. (Fiske, 2000, p. 306)

An “ingroup” is a group of individuals who depend on one another to secure important practical outcomes, and cooperation among members of ingroups requires some shared beliefs and goals. We should therefore consider that humans are biased to believe the relevant beliefs necessary for participation in important social groups, and expect that humans would be biased toward preferentially accepting the testimony of members, or good prospective members, of ingroups.

At the same time, an individual who is not dependent on the success of her audience, that is, has independent or conflicting interests, can benefit from manipulating the audience in a harmful way without incurring the cost of harming a member of an ingroup. In addition, knowledge of reputations could be limited to a familiarity with individuals with whom an individual has had dealings. It would be unsurprising, then, if humans evolved to efficiently categorize outsiders from members of ingroups. This can be efficiently accomplished by cueing on easily recognized superficial signs of group membership, specifically signs of similarity and difference.

This is what additional studies have suggested, that people, within milliseconds, identify members of their ingroups by gender, race, and age (Banaji & Hardin, 1996; Zarate & Smith, 1990). Once the judgment is made that the individual is a member of an ingroup, they may be motivated towards maintenance of that relationship. If motivated in this way, as the Lundgren and Priskin study suggests, people will be biased to only

consider confirming evidence for that individual's testimony and have their opinion moved closer to that of the speakers. Other experiments, notably by Dovidio and Gaertner (1993) and Fiske (1998) show specifically that people respond more quickly and more positively to ingroup members than they do to outgroup members (see Fiske, 2000 for a review).

In addition, once a speaker has been judged as part of an ingroup, there is a motive to individuate that person and see them, not in stereotyped categories, but as an individual (see Fiske & Depret, 1996, for references). This is necessary for the improved understanding of the person required for successful coordination. In doing so, they will be exposed to more of the beliefs and testimony of that individual and, in a very literal way, come to share the same attitudes as other ingroup members.

As a refinement of my commentary on the Hasson (2005) study, these experiments have been shown that additional cognitive resources are allocated to ingroup members. However, these resources are not always directed at factual accuracy but could be used for “tagging” and remembering what other ingroup members believe are false. This is important because the remembering of prohibitions and taboos can be important to maintaining good social standing.

At the same time, when confronted with superficial signs of an individual being different from themselves, humans are biased to apply negative stereotypes, consistent with the detection of a threat. As explained in the last chapter, a lack of social dependency with another individual results in a context without the coincident interests that stabilize valuable signaling. Consequently, humans would have good reason to be skeptical of people who display signs of failing to be a member of their ingroup and who

may mislead a hearer with social impunity.

Of course, this adaptation would have been more appropriate in a tribally organized human history where individual success was wholly dependent on tribal success and the greatest threat to a tribe was other tribes of humans. As Caporael (1997) and colleagues have conjectured humans evolved to regulate effective social coordination in face-to-face interactions in a tribal environment.

Group Bias in Young Children

Before we continue, I wish to draw attention to additional studies that demonstrate complex regulatory practices even in children as young as 3. One of the reasons epistemologists in the tradition of Thomas Reid have assumed people very often “default” believe is that more complex regulatory practices are unavailable to young children. In fact, significant evidence suggests that young children develop the ability for complex judgments that regulate acceptance of testimony at a very young age, and many of those practices also fit the social adaptationist model.

Some of the more striking studies include Mascaro and Sperber's (2009) study that observed a bias by three-year-olds to listen to the testimony of informants that act benevolently and have been described by others as having high moral character. In a similar study by Clement, Koenig, and Harris (2004), it was shown that young children would also prefer informants that both appeared and were labeled as more competent. As Sperber et al. (2010, p. 372) comments, “(e)arly epistemic vigilance draws on some of the capacities used in selecting partners for cooperation, which includes moral evaluation, monitoring of reliability, and vigilance towards cheating” (see Cosmides & Tooby, 2005;

Harris & Nunez, 1996). Sperber et al. (2010), however, label these capacities for “epistemic vigilance,” and ignore the other motivations that affect cognitive processing, preferring to claim that deviations from good epistemic practice can be explained by heuristics that cue on signs of benevolence or group membership because these properties would have normally coincided with factual accuracy.

What I favor is an inversion of the view of Sperber et al. that selecting partners for cooperation is an indirect way of improving judgments of accuracy. Instead, it seems more likely, given the wealth of evidence on display, that nonepistemic motives will override accuracy concerns unless factual accuracy is important to successful cooperative social interactions. Going to the effort to confirm the truth is in the service of cooperative success instead of expending the effort to identify suitable cooperative partners being in the service of getting the truth.

That suitability judgments may be sometimes effective at weeding out the truth would be expected, since the reliability of an individual to know and share the truth is an important measure of their desirability as a cooperative partner. However, the evidence for the inversion I favor is that the empirical literature confirms repeatedly that, in contexts where an individual is not primed to be specifically concerned with accuracy, humans will abandon collecting and judging evidence, in favor of a preferential credulity towards members of the groups on which they depend.

Self-image Bias and Manipulation

I have provided a survey of experiments that show that humans will ignore evidence in favor of maintaining valuable relationships. Next, I want to introduce

additional studies that confirm that humans will fail to employ good epistemic practices in favor of maintaining a positive self-image. Although some of the experimental evidence surveyed below does not address the processing of testimony directly, each of the studies shows that humans routinely come to believe what is false by ignoring evidence or being biased towards information.

There is more than anecdotal evidence that humans trade off concerns over accuracy for the practical benefits of maintaining consistently unrealistically positive self-assessments. Haselton and Buss (2000) showed that men overestimated the sexual intent of women even in the face of clear and available evidence that the women were not interested. In a related study, Steele (1998) found that women tended to underestimate the familial investment intentions of men and so delay sexual contact, even when confronted with clear evidence that the man was committed. In other studies of self perception, it was shown that humans consistently hold unrealistically positive assessments of their own intelligence, honesty, persistence, originality, amicability, and ironically, their reliability at accurate self-assessment (Friedrich, 1996; McKay & Dennett, 2009, p. 32; Pronin, Gilovich, & Ross, 2004; Pronin, Lin, & Ross, 2002). As a comedic example, in a 1997 U.S. News and World Report poll, readers were asked whether various personalities, including themselves, would get into heaven. Nineteen percent responded that O.J. Simpson was getting in, 66% were optimistic about Oprah Winfrey, while only 79% stated that Mother Teresa would pass the Pearly Gates, yet 87% of respondents said they would be accepted.

Self-image bias is so widespread that it could not be a malfunction of the human cognitive system. It is also implausible that it is a natural limitation of cognitive design,

since humans seem to possess the ability to better process evidence when motivated to do so. McKay and Dennett (2010) argue, and I agree, that maintenance of an unrealistically positive self-image is an example of what they describe as “adaptive misbelief,” or the holding of *false* beliefs that would increase individual fitness. If self-image bias is not an efficient heuristic for getting the truth, then what explains the willingness of individuals to hold significant numbers of false beliefs, ignoring relevant evidence, about something as important as their own competencies?

I agree with McKay and Dennett that self-image is intimately connected to an individual's motivation, confidence, and subsequent behavioral success. If an individual believes they are better than they are, or more attractive than they are, they may attempt to acquire resources, attract mates, or accomplish other feats that they would not attempt otherwise. If the cost of failure is low, then the repeated attempts, even if they only have a small chance of success, may result in an overall advantage.

This is another example of what Triver's observes in the opening quote, that the benefits from holding false beliefs can depend on the individual thinking they are true. If this is common, then an individual's expressed conviction alone is not good evidence for the accuracy of their testimony. Conviction would often serve as a motivational property of beliefs rather than correlating with epistemic success. Evolution, it would seem, actually selected for flexibility in cognitive practices to abandon good epistemic practices in favor of useful biases.

Confirmation Bias

A second type of study that is related to the self-serving bias is the widely studied phenomenon of “confirmation bias.” This bias is described by Nickerson as the “ubiquitous” practice of the, “seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (Nickerson, 1998, p. 175). Other researchers have claimed that this practice affects everyone, regardless of age or other objective measures of intelligence (see Stanovich, 2004), and is therefore not a malfunction of the human cognitive system.

There are at least two different functional explanations for the persistence of confirmation bias in humans that have been offered in the literature. Mercier and Sperber (2010) argue that confirmation bias seems to occur when an individual is engaged in reasoning in an attempt to produce a compelling argument. That is, confirmation bias is a practice associated with the sender-side traits used in manipulating hearers with testimony. The reason, they argue, why humans persist in being biased towards their own beliefs, and look only to identify positive evidence for existing beliefs, while ignoring or disregarding counter evidence, is that this practice would allow them to be more convincing.

Kruglanski (2004), in work I have already cited, offers a different, but also practical, effect of confirmation bias. Kruglanski argues that humans need to have conviction in beliefs to overcome ambivalence and make the decisions required for decisive action. Because of this fact about human motivation, humans will tend to hold on to, or “freeze” on existing beliefs, even in the face of good counter-evidence, far longer than would be prescribed by ideal epistemic practice. This explanation is very

similar to the one offered for self-serving bias, that what humans choose to believe may have more to do with guaranteeing that they are properly motivated than it has to do with knowing the truth.

Mercier and Sperber's argument closely reflects again the contention in the Triver's quote, that what convictions a human holds determine that individual's ability to "manipulate" others in a beneficial or self-serving way. The story I want to tell is as follows: Humans are heavily affected by the perceptions of others towards them, including the critical assessment of their reputation, and other signs they may display of their suitability as a cooperative partner. It would be highly advantageous, then, for people to portray themselves as more likable, competent, or successful, than they actually are. The problem is that doing so poses a great risk of being found out and punished if an individual *knowingly* lies or portrays himself in misleading ways (see von Hippel & Trivers, 2011). Individuals who were therefore capable of the self-delusion necessary to testify with genuine conviction to what is false or dubious would have a significant advantage by avoiding the risk of displaying the, "subtle signs of self-knowledge of the deception being practiced."

Notice that in human culture, there are two very different reactions to an individual who communicates what is false, and these reactions depend on whether the individual simply made an epistemic error, or whether the individual knowingly lied. The first individual is usually forgiven, and the second is labeled as a liar deserving of moral censure. Humans do seem to have the capacity to convince themselves of some fact in order to convincingly persuade others of that fact (e.g., it is a common strategy for effective bluffing in poker to just "decide" that you have different cards than you do).

Human evolutionary history would therefore involve selection for cognitive practices capable of producing conviction in falsehoods and convenient beliefs.

Conclusions

The last two chapters have done more than display alternatives to the assumptions of the epistemologists of testimony. Many of the observations about the evolution of signaling and the psychology of information regulation have significant implications for the epistemic reliability of testimony. The most startling observation is that, far from evolution being expected to produce epistemically reliable cognitive faculties, selection would instead favor the advantage of the flexibility to abandoned good epistemic practice when other nonepistemic motives take precedent.

The rest of this conclusion is a summary of the suggestive scientific findings from Part II that are important to judgments about the epistemic reliability of testimony. In general, the reliability of testimony will depend on the practices and motives in play when the belief being testified to was accepted by the speaker. It is an important feature of testimony that humans pass on what they have come to believe. It is therefore necessary to know what the motives of a speaker were in accepting testimony and whether their source was sufficiently vetted for their motivations in testifying. The new story about how humans engage in testimonial interactions is as follows:

Humans use testimony to manipulate other individuals often in benevolent and cooperative ways but sometimes in self-serving and harmful way. Because providing misleading information can greatly benefit the speaker, the use of valuable testimony must be stabilized by interdependency in social groups and effective enforcement of

benevolence. Humans offer testimony in a social context where the relationship with the hearer determines the meaning of the information provided. It is not the case that human speakers act as devices for simply passing on true information, like an encyclopedia or textbook.

Humans tell each other things to coordinate in cooperative relationships and this is sometimes but not always accomplished by providing valuable true information. Humans offer testimony to signal their suitability for continued mutually beneficial interactions and relationships, and this is most often in the form of self-reportings that serve to display the individual's reputation. One way that reputation can be increased is by providing reliable, relevant factually accurate information, but many other characteristics can be desirable in social interactions. Humans can use testimony to manipulate their audience in favorable ways, and sometimes these tactics are self-serving and harmful to the hearer.

Humans are biased to accept or reject testimony based on their need to use testimony to satisfy certain cognitive and social ends, and human receiver-side practices are very flexible in moving between these various concerns. Among these motivations, there seems to be at least three that result in different biases and produce different judgments about what information to accept. Individuals can be motivated by a desire for accuracy, by a desire to defend their prior opinions, or to be amicable towards the source of information. This is not an exhaustive typology and an individual can be motivated in multiple ways and in varying degrees.

When an individual is defensive of their opinions it could be that those opinions constitute a self-perceived identity that must be stable for the individual to be comfortable, or it may be that those opinions are important to express with conviction in

order to manipulate hearers effectively. Therefore, when an individual is primarily motivated to defend their opinions, they will tend to be aggressive in expressing their points of view, be dismissive of counter-evidence, have polarized conviction in their beliefs, and see dissent as an attack on their character and indicative of poor character on the part of the dissenter.

Individuals who tend towards demanding accuracy will require evidence for what they hear even when etiquette and amicability would demand agreement. Truth seekers will be skeptical even when conviction is required for decisive action, and identify themselves with a commitment to epistemic rigor. Truth-seekers may become frustrated with what they take to be unsupported opinion and speculation.

Lastly, when individuals are motivated by a need for amicability, which some evidence suggests produces the most robust attitudes, they may become credulous and even gullible, relying on the benevolence of the source. If asked to decide, they may become uncomfortable with disagreement, or debate, and insist that there is no way to decide one way or another. An amicable individual will abandon prior conviction in favor of a more comfortable ambivalence, but may be incapable of summoning the conviction necessary to make decisions.

It may be that different individuals tend in general to be more motivated by some concerns and less over others. These general tendencies may be learned and molded during development through reinforcement. For example, someone who has been subjected to harmful conflict over ideas may be motivated by amicability and be uncomfortable with debate. Someone who has a need to have control over others may be primarily motivated by defensiveness and aggressively demand acceptance of their

beliefs by others, while judging dissent as a threat.

If any part of the preceding story is true, then these findings are significant for epistemic judgments about testimony, since all testimony will be offered by individuals who have formed their beliefs based on the variety of motivations discussed. Next, I will offer a new epistemology of testimony that will be informed by the science in the last two chapters.

PART III

A NEW EPISTEMOLOGY OF TESTIMONY

CHAPTER 9

ACCURACY IS EXPENDABLE

Introduction

In this last part of the dissertation, I revisit the reliability of testimony out of a desire to offer preliminary suggestions for helping us, as humans, avoid what I think are widespread factual errors. What norms I propose will only be useful to those among us who want to know the facts and avoid false beliefs, even when an insistence on accuracy comes at the expense of getting along well in the world. In other words, this section is for the truth-seeker motivated by a concern over accuracy.

As philosophers and scientists who make a career out of trying to be accurate, it may be tempting to think that there is some social, practical, or moral mandate on being epistemically rigorous, and being careful to avoid passing on misinformation, but this is not natural for humans. While this noble insistence on the truth is important for individuals who make a career out of describing reality, it can be unrealistic for individuals in everyday circumstances. What follows then are preliminary empirically informed principles and prescriptions for humans making common judgments about who to believe. My goal is not to prescribe ideal practices or norms that should be enforced in specific epistemic contexts such as during scientific investigations.

So what is the epistemological status of testimony? The answer is complex since it will vary depending on features of the context and the motives of the individuals involved. What is not the case is that a truth-seeker should assume that testimony is generally reliable, nor that it would be a sufficiently discerning practice to passively await the monitoring of good reasons to reject what is said. Instead, it must always be in the truth-seeker's mind whether the testifier believes what they do to satisfy other social or psychological needs, or whether the testimony must be true to be beneficial. We now examine the relationship between sharing the truth and the other uses of testimony.

Conclusions: Epistemic Principles for Evolved Beings

Let me now make explicit the conclusion of everything that has come before: The relationship between the role of epistemic reliability and the practical value of testimony is inverted from what has been commonly assumed. It is not the factual accuracy of testimony that explains its value and persistent use, but instead it is the numerous benefits of communication that sometimes require that testimony be factually accurate.

Epistemologists have assumed that testimony is an epistemic tool, and it is the knowledge produced that provides the benefits that explains its use. In contrast, the evidence from Part II strongly suggests that it is the multiple fitness and utility enhancing benefits of testimony that explain the evolution and persistence of the use of testimony. Therefore, when testimony gets an individual knowledge, that accuracy must be explained by the nonepistemic benefits of communication that required accuracy.

For example, if one of a pair of friends traveling to a restaurant testifies to the location of the restaurant, it is because they want to help both arrive at the restaurant, and

it happens in this case that accuracy is necessary to achieve those ends. What is not the case is that the passenger is testifying because they want the friend to have the truth and it just so happens that the location of the restaurant is a piece of useful information.

Evolved consumers of information are practical first and only concerned with accuracy when necessary.

With this inversion in mind, let me introduce new epistemological principles that contrast with the common optimistic presumptions of the epistemology of testimony. Any truth-seeking consumer of testimony should base their evaluation of the accuracy of testimony on the following principles:

Principle 1: “Truth is Expendable.” If all of the benefits of the use of testimony would accrue without testimony ever being true, then humans would never expend the effort to be concerned about the factual accuracy of testimony.

Principle 1 follows from the facts that human communication evolved, and that checking for factual accuracy requires significant resources. In a case where the benefits of interacting are unaffected by accuracy, the additional epistemic effort required to check for the truth would be a waste of resources and therefore be maladaptive. It would therefore be a selective advantage to forgo effort spent on checking for accuracy if accuracy was irrelevant to success. Therefore, the traits responsible for regulating what a hearer accepts would be better adapted if they could flexibly ignore considerations of accuracy when having the truth is unhelpful. This is exactly what the psychological studies have shown, that humans will simply accept what they hear when accuracy is irrelevant and abandon good epistemic practices to meet other needs.

To say it a different way, the “default” practice of humans is to expend minimal effort towards checking for accuracy unless making an error would be costly. I conclude that it is this fact that explains widespread human credulity towards testimony and not, as epistemologists have assumed, that communication evolved in a reliable informational environment. The “default rule” from the epistemology of testimony, that it is generally *epistemically* rational to be credulous, is false. However, the human practice of being credulous will often be *prudent*.

Principle 1 is not a norm but a claim about human nature; however, it implies a significant fact about the reliability of testimony. Since humans testify to what they have heard, except when they report on first-hand experiences or their own mental states, the truth-seeker must check whether the speaker had a concern over accuracy in forming their belief or whether they may have accepted what they believe for nonepistemic reasons.

If correct, Principle 1 is a standing defeater to all instances of testimony, because knowing that humans will deviate from good epistemic practice unless doing so is harmful is a good reason to think that testimony may be unreliable. Humans form the vast majority of their beliefs through testimonial interactions that evolved for prudence over accuracy, and consequently, only the practical value of those beliefs is in any way guaranteed by the origins and persistent use of testimony.

An implication of the fact that truth is expendable is that a conscientious truth-seeking consumer of testimony must minimally discern that they are in a context where the value of communicating requires that it be true, and would do well to be sensitive to motives and competencies of the speaker that would likely make testimony unreliable.

The claim of nonreductionists that humans can regularly and credulously come to know what they accept is false. It will often be *prudent* to simply accept what one is told, but because testimony is a practical tool and accuracy is expendable, detecting the truth requires confirming that the value of testimony depends on factual accuracy, and that the speaker is offering valuable testimony.

The truth-seeking consumer of testimony should therefore assume that the claims of speakers are merely used prudently unless there is a good reason to think that prudence required accuracy. To restate the inversion again, it is prudence that sometimes requires accuracy and not accuracy that is always prudent. This fact suggests a second normative principle that can be used as a test:

Principle 2: “Accuracy Serves Prudence:” The offering of testimony is only evidence that the claim made is true if the benefits to participants from the testimonial interaction requires that the testimony be true.

Because of the expendability of truth, any testimony that satisfied some practical function that did not require that it be factually accurate must be assumed to have been formed by an individual using poor epistemic practices that ignore accuracy in favor of satisfying other practical needs. Any belief that is formed using a poor epistemic practice is not epistemically justified, and therefore cannot be known the speaker, and any testimony that is not known by the speaker cannot be passed on to the hearer as knowledge.

When I say that to be epistemically justified requires knowing that the benefits of the testimonial interaction “requires” that the testimony be true, I mean that if the testimony was false, then either the participants would fail to secure the nonepistemic benefits of interacting or some harm would be caused by the error. Minimally, then, a

truth-seeking consumer of testimony must check whether the testimony in question would be valuable only if it was true. This makes the reliability of testimony context-dependent in a clear and testable way. The truth-seeking consumer of testimony can judge whether, if false, the benefits of the interaction would be lost.

What I am arguing for is a type of reductionism in the epistemology of testimony because every instance of testimony requires that the truth-seeker do some work to epistemically justify accepting testimony. Minimally, the truth-seeker must confirm that the nonepistemic benefits of using testimony requires that it be true.

Prescriptions: A Preliminary Empirically informed

Philosophy of Testimony

In this penultimate section of the dissertation, I discuss possible empirically informed prescriptions for detecting the accuracy of testimony. More than anything else, I would like what follows to be an example of how the science in Part II provides information necessary for identifying effective strategies for improving epistemic judgments. Nonreductionists prescribed the very efficient heuristic of following the default rule, but as we have seen, this is not an epistemically reliable practice. Reductionists made vague prescriptions to use background beliefs and be sensitive to signs of deception, but these prescriptions did not address the specific reasons why testimony is frequently unreliable. We are now in a position to reject the a priori optimism that led to prescribing the default rule and make more nuanced empirically informed prescriptions.

A first test for detecting factual accuracy that is suggested by the social

psychology literature would be to detect the speaker's motives, both in offering testimony and in holding the belief communicated. Does the speaker believe what she/he are claiming because of the value of knowing the truth or does the belief satisfy some other nonepistemic need such as maintenance of a positive self-image or membership in an important social group? In addition, judging the reliability of testimony seems to require that a hearer determine whether the testifier is communicating the belief out of a benevolent attempt to provide information that must be true to be useful or whether the factual accuracy of what is claimed is irrelevant to informing more successful future conduct.

An example of a case where testimony is most likely to be true would be when a friend offers information about the local physical environment to aid in navigating that environment, and when the speaker has directly witnessed what they claim. The example of someone telling directions, pointing out the tracks of an animal, confirming the location of artifacts (e.g., mom saying there is milk in the fridge), or commenting on the properties of objects (e.g., that stove will burn you), are good examples.

The truth-seeking consumer of testimony would be well advised to accept testimony of this sort once if they are aware that the function of the testimony is to provide true information about the local environment. Acceptance in these cases would not be cases of complete credulity since the hearer has made a judgment prior to acceptance, the judgment that the speaker is offering information about the physical environment witnessed first hand, to help the hearer navigate the environment. If any of these conditions is not met, a hearer should be less confident in the accuracy of the testimony.

However, even in these most epistemically reliable cases of testimony, the conscientious truth-seeking consumer of testimony must be sensitive to signs that the speaker is, for example, playing a joke or trying to sabotage them, and giving them false information. It has been recently shown in relation to eyewitness testimony in courts of law that testimony is shockingly unreliable (see Loftus, 1979). The hearer must also be sensitive to less nefarious causes of misinformation, such as when the speaker is either incompetent at detecting or remembering the facts accurately.

As suggested by the signaling theory literature, if the speaker has genuine coincident interests with the hearer, then the speaker harms themselves by misleading the hearer. The implication is that a preliminary test for accuracy depends on judging the *benevolence* of the speaker. Speakers that do not have a hearer's best interest in mind, because, for example, they are in competition, should be assumed to be epistemically unreliable unless there are other good reasons to think they are testifying to the truth. However, judging benevolence is not enough to determine if an instance of testimony is true, because there are often cases where factual accuracy is irrelevant or even an impediment to the nonepistemic value of testimony. It must also be that the value of testimony requires that it be accurate, as in the case of navigating the environment.

I recognize that determining the motives of the speaker in believing what they do and their motives in testifying to those beliefs can be difficult facts to determine. It requires mind-reading and understanding the relationship between accuracy and usefulness, and may require skill at detecting deception. Humans would need to be taught how to improve these skills because, as it seems, they will abandon good epistemic practice whenever it is prudent to do so. In addition, since human practices evolved for

face-to-face interactions with members of ingroups, without training, they will often be more credulous than is warranted. However, just alerting hearers to the practical nature of testimony may go a long way towards instilling a healthy epistemic skepticism that would reduce errors.

Consider another common type of testimony that should be considered *unreliable*. When people communicate standards of good conduct or communicate behavioral imperatives, the prescriptions of this section of the dissertation included, there is a clear nonepistemic function to modify the behavior of the hearer. A conscientious truth-seeker should consider this nonepistemic goal as counting against the epistemic reliability of what is said.

In other words, the speaker should be seen as having a “conflict of interest” where the nonepistemic interest, altering the behavior of the hearer in desirable way, is in conflict with offering true testimony. A conflict of interest occurs when the fiduciary responsibility to a client or patient is not met because the individual providing the service or care acts in their own conflicting interests. In the case of testimony, a conflict of interest occurs when there is a self-interest that conflicts with the accuracy of testimony. Because of the evolution of manipulation, this is common, if not universal, in human communication.

In academics, there are disclosure laws that require researchers to state that they have no other interest but to produce and communicate factually accurate theories. Defendants in courts of law are also asked to take a vow that they are testifying to the facts instead of trying for some other desirable outcome. What I have attempted to show is that in normal human communication, there is regularly a conflict of interest, because

humans use testimony for its practical value, which may conflict with a concern over accuracy. Conflicts of interest are good reasons for judging that testimony is unreliable.

The signaling theory literature also suggests that a truth-seeking consumer of testimony can judge that likelihood of testimony being true by determining the cost to the speaker of producing the testimony. If testimony is very cheap to produce, then misleading testimony is more likely. If there is a significant cost to producing testimony, then it is more likely to be beneficial, but again, there must be an additional epistemic judgment about whether factual accuracy is required for the benefits to accrue.

A very important and relatively easy test for determining the cost of distributing information is to identify if there is any enforcement of a responsibility on the part of the source to provide factually accurate information. A hearer of testimony can ask themselves whether the speaker would be harmed if the claim was found out to be false.

For example, compare the extensive effort on the part of researchers to discover, and bring to prominence, scientific theories, in comparison to some individual distributing a conspiracy theory or political slander through a bulk e-mail. In the scientific case, the cost of research is evidence that it is more likely to be valuable, and since the value of new science is explicitly that it is accurate, genuinely scientific claims will be more likely to be true. In addition, a scientist who publishes faulty data risks their reputation and career. A sender of an anonymous bulk e-mail has neither expended considerable resources on determining the facts nor are they accountable for factual errors.

It is important, however, to remember that enforcement is only as good as error detection within the informational community. This is important because there are cases

where punishments may be harsh but detection is very unlikely or impossible. I might go to hell or be considered immoral if I get my religious beliefs wrong but there is no way to confirm that they are false in a way that would result in me paying the price. In contrast, a scientist publishing results is under the scrutiny of other scientists who perform blind review and will replicate their experiments. In science, these are important practices to guarantee accuracy. The most important technique for enforcement in common communication is to alert other people to a source's reputation. Norms that are tolerant to speakers making false claims, such as a commitment to “agreeing to disagree,” may reduce conflict, but make inaccuracy more likely.

I want to admit again that these tests can be difficult to perform, and so this section offers heuristics that might improve epistemic judgments. The prescription for the truth-seeker is just this: If the speaker is known to be benevolent, has no conflict of interest, and accuracy is necessary for accruing the benefits of communication, then testimony has a high likelihood of being accurate. If testimony was very expensive to produce and/or there is reliable and effective enforcement, then it also has a higher chance of being valuable, and if that value would not accrue unless it was factually accurate, it has a higher chance of being true. Testimony that clearly satisfies other nonepistemic needs, where its accuracy is independent of its value, when production is cheap and when there is poor enforcement, should be considered epistemically unreliable.

However, it must be remembered that for a person trying to get along well in the world, being a truth-seeker has its costs. Checking for the truth takes effort that may be better spent in pursuing practical ends, can damage an individual's suitability for social interactions and relationships, and can impede effective practices of manipulation that

would otherwise produce valuable agreement in other individuals. However, alerting people to these facts should in itself go some way to improving their epistemic judgments.

Future Research

In this last section, I describe additional research projects that would be part of a larger empirically informed philosophy of testimony. The first project would be to develop a much more complete account of what was started in the last section by developing more accurate and efficient heuristics for avoiding epistemic errors.

A second project would be to understand the relationship between a speaker's expressed conviction in their beliefs and the accuracy of their testimony. There is an interesting fact about human psychology that humans believe testimony, and therefore think it is "true" to gain the practical value of having conviction. If you will recall, Kruglanski (1989, 2004) described conviction as the "closing" and "freezing" on beliefs and explained the need to do so as necessary for committed action. Sperber and his colleagues argued that conviction has a rhetorical function to make testimony more convincing.

These results imply that there is a divergence between the strength of conviction, and the likelihood that the person who is strongly committed knows the truth. It seems that conviction is not a generally reliable guide to accuracy, but is instead a reliable guide to the personal value of holding the belief in question. This would be an important result since expressed conviction is very often used to convince hearers of the accuracy of testimony, and because strong conviction may even be a sign that the belief has more to

do with the testifiers psychological well-being, and is therefore false. Notice that people often substitute conviction for good evidence when trying to convince others. A common reaction to the challenge, “how do you know that?” is “because I just know (spoken with great emphasis)!”

The third project would be to offer a more plausible explanation for the widespread epistemic errors that produce disagreement, and cultural variation in human belief systems. Since many of those nonepistemic needs are satisfied by sharing important cultural beliefs, it is no surprise that humans would widely disagree, and defend their psychologically and socially important convictions, and this would lead to variation and disagreement between individuals that have grown up in different testimonial environments.

BIBLIOGRAPHY

- Adler, J. E. (2002). *Belief's own ethics*. Cambridge, Mass. London: MIT Press.
- Audi, R. (1997). The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly*, 34(4), 405-422.
- Austin, J. (1979). Other minds. In *Philosophical papers* (3rd ed., pp. 44-84). Oxford: Oxford University Press.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7(3), 136-141. doi: 10.1111/j.1467-9280.1996.tb00346.x
- Burge, T. (1993). Content preservation. *Philosophical Review*, 101, 271-300.
- Burge, T. (1997). Interlocution, perception, and memory. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 86(1), 21-47.
- Caporaol, L. R. (1997). The evolution of truly social cognition: The core configurations model. *Personality and Social Psychology Review*, 1(4), 276-298. doi: 10.1207/s15327957pspr0104_1
- Chaiken, S., Wood, W., & Eagly, A. H. (1996). Principles of persuasion. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles*. (pp. 702-742). New York, NY US: Guilford Press.
- Cialdini, R. B. (1987). *Compliance principles of compliance professionals: Psychologists of necessity*. Paper presented at the Social Influence: The Ontario Symposium. 1986. Ontario, Canada.
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vols. 1 and 2, 4th ed., pp. 151-192). New York, NY US: McGraw-Hill.
- Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, 19(4), 360-379. doi: 10.1111/j.0268-1064.2004.00263.x

- Coady, C. A. J. (1992). *Testimony : A philosophical study*. Oxford, New York: Clarendon Press; Oxford University.
- Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *Evolutionary Psychology Handbook* (pp. 584-627). New York: Wiley.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17, 304-323.
- Davidson, D. (1973). Radical Interpretation. *Dialectica*, 27, 314-328.
- Dawkins, R., & Krebs, J. (1978). Animal signals: information or manipulation? In J. Krebs & N. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (Vol. 1). Oxford: Blackwell.
- Dennett, D. C. (1991). *Consciousness explained* (1st ed.). Boston: Little, Brown and Co.
- Descartes, R. (1641). *Meditations on first philosophy*. In Haldane & Ross (Eds.), *The Philosophical Works of Descartes* (Vol. i). Cambridge: Cambridge University Press, 1967.
- Dovidio, J. F., & Gaertner, S. L. (1993). Stereotypes and evaluative intergroup bias. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception*. (pp. 167-193). San Diego, CA US: Academic Press.
- Downes, S. (2000). Truth, selection and scientific inquiry. *Biology and Philosophy*, 15(3), 425-442.
- Faulkner, P. (2000). The social character of testimonial Knowledge. *The Journal of Philosophy*, 97, 581-601.
- Fiske, S., & Dupret, E. (1996). Control, interdependence and power: understanding social organization in its social context. *European Review of Social Psychology*, 7, 31-61.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology, Vols. 1 and 2 (4th ed.)*. (pp. 357-411). New York, NY US: McGraw-Hill.
- Fiske, S. T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: evolution, culture, mind, and brain. [Article]. *European Journal of Social Psychology*, 30(3), 299-322.
- Fredrich, J. (1996). On seeing oneself as less self-serving than others: the ultimate

- selfserving bias? *Teaching of Psychology*, 23(2), 107-109.
- Frey, B., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding out. *American Economic Review*, 87, 746-755.
- Fricker, E. (1994). Against gullibility. In B. K. Matilal & A. Chakrabarti (Eds.), *Knowing from words* (pp. 125-161). Dordrecht: Kluwer Academic.
- Fricker, E. (1995). Telling and trusting: Reductionism and anti-reductionism in the epistemology of testimony. *Mind: A Quarterly Review of Philosophy*, 104(414), 393-411.
- Fricker, E. (2006). Testimony and epistemic autonomy. In L. J. & E. Sosa (Eds.), *The epistemology of testimony*. Oxford: Oxford University Press.
- Ghiselin, M. (1974). *The economy of nature and the evolution of sex*. Berkeley: University of California Press.
- Gigerenzer, G. (2008). *Rationality for mortals : How people cope with uncertainty*. Oxford ; New York: Oxford University Press.
- Gilbert, D. T. (1991). How mental systems believe. *Psychologist*, 46, 107-119.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601-613. doi: 10.1037/0022-3514.59.4.601
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65(2), 221-233. doi: 10.1037/0022-3514.65.2.221
- Goldman, A. I. (1999). *Knowledge in a social world*. Oxford New York: Clarendon Press; Published in the U.S. by Oxford University Press.
- Grafen, A. (1984). The ecology of information. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology* (pp. 62-84). Cambridge: Blackwell Press.
- Grice, H. (1969). Utterer's meaning and intentions. *Philosophical Review*, 78, 147-177.
- Grice, H. (1975). Logic and conversation. In C. P. & M. J. (Eds.), *Syntax and semantics, Vol 3: Speech acts*. New York: Seminar Press.
- Harris, P., & Nunez, M. (1996). Understanding of permission rules by preschool children. *Child Development*, 67(4), 1572-1591.

- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1), 81-91. doi: 10.1037/0022-3514.78.1.81
- Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science*, 16(7), 566-571. doi: 10.1111/j.0956-7976.2005.01576.x
- Hull, D. (1989). A function for actual examples in philosophy of science. In M. Ruse (Ed.), *What the philosophy of biology is* (pp. 309-322). Dordrecht: Kluwer Academic Publishers.
- Hume, D. (Ed.). (1977). *An enquiry concerning human understanding*. Indianapolis: Hackett Publishing.
- Hölldobler, B., & Wilson, E. O. (1990). *The ants*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Kruglanski, A. (1989). *Lay epistemics and human knowledge*. New York: Plenum Press.
- Kruglanski, A. W. (2004). *The psychology of closed mindedness*. New York, NY US: Psychology Press.
- Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, 19(5), 448-468. doi: 10.1016/0022-1031(83)90022-7
- Kruglanski, A. W., Freund, T., & Bar-Tal, D. (1996). Motivational effects in the mere-exposure paradigm. *European Journal of Social Psychology*, 26, 479-499.
- Lachmann, M., Szabolcs, S., & Bergstrom, C. (2001). Cost and conflict in animal signals and human language. *PNAS*, 98(23), 13189-13194.
- Lackey, J. (2006). Learning from words. *Philosophy and Phenomenological Research*, 73(1), 77-101.
- Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford: Oxford Univ Pr, 2008: Oxford Univ Pr.
- Lackey, J. (2011). Testimony: Acquiring knowledge from others. In A. I. Goldman & D. Whitcomb (Eds.), *Social epistemology: Essential readings* (pp. 71-91). New York: Oxford University Press.
- Lehrer, K. (2001). Testimony and trustworthiness. In J. Lackey & E. Sosa (Eds.), *The*

- epistemology of testimony* (pp. 145-159). Oxford: Oxford University Press.
- Lewis, D. (1969). *Convention: A philosophical study*. Cambridge: Harvard University Press.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556-567.
- Lewis, D. (1974). Radical interpretation. *Synthese*, 27, 337-344.
- Locke, J. (1690). *An essay on the human understanding*, abridged and ed. T. John. London: Everyman Press, 1993.
- Loftus, E. (1979). *Eyewitness testimony*. Cambridge: Harvard University Press.
- Lundgren, S. R., & Prislun, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin*, 24(7), 715-726. doi: 10.1177/0146167298247004
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, 112(3), 367-380.
- Maynard Smith, J. (2000). The concept of information in biology. *Philosophy of science*, 67(2), 177-194.
- McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6), 493-561. doi: 10.1017/s0140525x09990975
- Melnyk, L., Crossman, A. M., & Scullin, M. H. (2007). The suggestibility of children's memory. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The Handbook of eyewitness psychology, Vol I: Memory for events* (pp. 401-427). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Mercier, H., & Sperber, D. (Forthcoming). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Science*.
- Millikan, R. (1993). *White queen psychology and other essays for Alice*. Cambridge: MIT Press.
- Mitchell, M. (2001). Analogy-making as a complex adaptive system. In L. S. & I. Cohen (Ed.), *Design principles for the immune system and other distributed autonomous systems*. New York: Oxford University Press.
- Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology*, 2, 175-220.

- Noble, J. (2000). Co-operation, competition and the evolution of pre-linguistic communication. In J. Piotr, Gmytrasiewicz, & N. Michael (Eds.), *The emergence of language among evolved beings* (pp. 40-61). Cambridge: Cambridge University Press.
- Owens, D. (2000). *Reason without freedom: The problem of epistemic normativity*. London ; New York: Routledge.
- Peri, N., Kruglanski, A., & Zakai, D. (1986). Interactive effects of initial confidence and epistemic motivations on the extent of informational search. *Unpublished manuscript, Tel Aviv University*.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vols. 1 and 2, 4th ed., pp. 323-390). New York, NY US: McGraw-Hill.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781-799. doi: 10.1037/0033-295x.111.3.781
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. doi: 10.1177/0146167202286008
- Quine, W. V. (1960). *Word and object*. Cambridge: Technology Press of the Massachusetts Institute of Technology.
- Reid, T. (1983). Essay on the intellectual powers of man. In E. Beanblossom & K. Lehrer (Eds.), *Thomas Reid's inquiry and essays*. Indianapolis: Hackett Publications.
- Richerson, P. J. & Boyd, R. (2005). *Not by genes alone : How culture transformed human evolution*. Chicago, University of Chicago Press.
- Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton: Princeton University Press.
- Skyrms, B. (2010). *Signals: Evolution, learning and information*. New York: Oxford University Press.
- Sosa, E. (2006). Knowledge: Instrumental and testimonial. In J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 116-123). Oxford: Oxford University Press.
- Sperber, D. (1985). *On anthropological knowledge : Three essays*. Cambridge Cambridgeshire; New York: Cambridge University Press.

- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25(4), 359-393. doi: 10.1111/j.1468-0017.2010.01394.x
- Sperber, D., & Wilson, D. (1995). *Relevance : Communication and cognition* (2nd ed.). Oxford, UK ; Cambridge, Mass.: Blackwell.
- Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago ; London: University of Chicago Press.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 21: Social psychological studies of the self: Perspectives and programs* (pp. 261-302). San Diego, CA US: Academic Press.
- Stevens, L. E., & Fiske, S. T. (1995). Motivation and cognition in social life: A social survival perspective. *Social Cognition*, 13(3), 189-214.
- Sunstein, C. R. (2003). *Why societies need dissent*. Cambridge, Mass.: Harvard University Press.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.
- Trivers, R. L. (2006). Forward to Richard Dawkins' the selfish gene *The Selfish Gene: 30th anniversary edition* (pp. xix-xx). Oxford: Oxford University Press.
- Tversky, A., & Kahneman, D. (2005). *Judgment under uncertainty: Heuristics and biases*. Northampton, MA US: Edward Elgar Publishing.
- Veblen, T. (1899). *The theory of the leisure class*. New York: Dover.
- von Hippel, W. and R. Trivers (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(01), 1-16.
- Weiner, M. (2003). Accepting testimony. *Philosophical Quarterly*, 53(211), 256-264.
- Wood, W. (1999). Motives and modes of processing in the social influence of groups. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 547-570). New York, NY US: Guilford Press.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51, 539-570. doi: 10.1146/annurev.psych.51.1.539

- Wood, W., & Kallgren, C. A. (1988). Communicator attributes and persuasion: Recipients' access to attitude-relevant information in memory. *Personality and Social Psychology Bulletin*, *14*(1), 172-182. doi: 10.1177/0146167288141017
- Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. New York: Oxford University Press
- Zahavi, A. A. (1975). Mate selection-a selection for a handicap. *Journal of Theoretical Biology*, *53*(1), 205-214.
- Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, *8*(2), 161-185.
- Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, *74*(5), 574-587.
- Zollman, K. (2009). Optimal publishing strategies. *Episteme: A Journal of Social Epistemology*, *6*(2), 185-199.