DO WE REALLY CARE WHETHER OUR BELIEFS ARE TRUE?: AN EXAMINATION OF THE ARGUMENTS FOUND IN CHAPTER FIVE OF STEPHEN STICH'S *THE FRAGMENTATION OF REASON*

by
Shawn Warren

A Thesis submitted to the Department of Philosophy
of Saint Mary's University
in partial fulfillment of the requirements
for the Degree of Master of Arts
Saint Mary's University, Halifax, Nova Scotia
November, 1996

Abstract

Do We Really Care Whether Our Beliefs Are True?:
An Examination Of The Arguments Found
In Chapter Five Of Stephen Stich's
The Fragmentation of Reason

by

Shawn Warren

November, 1996

This thesis is a critical examination of the arguments found in Chapter Five of Stephen P. Stich's book, *The Fragmentation of Reason.* Generally speaking, Stich's book is designed to deconstruct some significant aspects of philosophical thought commonly found in the style of philosophy referred to as analytic, particularly analytic epistemology. His goal is to both reveal how analytic philosophy and epistemology fail to provide adequate guidance in evaluating and improving our cognitive reasoning strategies, and to ultimately offer a philosophically pragmatic methodology which he feels is superior in this respect. In the process of this deconstruction, Stich criticizes two important features of analytic epistemology, two features which form the foci of this thesis - truth and true beliefs. He asks if we *really* care whether our beliefs are true or not? By the end of Chapter Five, his answer is a resounding, no. He claims that in the first instance the interpretation function - a function that maps beliefs onto truth conditions, propositions, etc., giving them semantic properties - most widely accepted by analytic epistemologist and our commonsense intuition is partial and idiosyncratic; that is, one function among many alternatives. With this conclusion in hand, Stich makes an argument for there being lots of competition for truth and true beliefs. Based on the possibility of alternative interpretation functions and viable competition for truth and true beliefs, Stich claims that anyone who intrinsically values true beliefs is both conservative and unreflective in their epistemic preferences. In the second instance, he argues that there is no obvious instrumental value in having true beliefs. Again, he relies on the notion of competing interpretation functions and the beliefs they produce, but in addition he says that in many cases we already know of instances where the holding of true beliefs would not be the most, or even the occasionally, optimal means to achieving our fundamental goals (e.g., survival). In response, this thesis will argue that Stich has been imprecise in his use and analysis of (alternative) interpretation functions, thereby bringing into question the argument that the function favoured by our commonsense intuition (and analytic epistemologists) is partial and idiosyncratic. In as much as this argument is

weakened, so is Stich's introduction of competing notions of truth and true beliefs. All of this, in turn, exposes the fragility of his main claim in Chapter Five - that there is no obvious intrinsic or instrumental value in true beliefs. This response will employ material found in the existing literature on this subject as well as some original analysis.

PREFACE

A philosopher of this cast is rare, made still rarer by our refusal to venture very far down the road he urges us to travel. Stephen Stich has, in *The Fragmentation of Reason*, asked us to journey with him down an unfamiliar road. During this journey, we are forced to re-examine and re-evalute many aspects of twentieth century philosophy and epistemology which we take for granted. As a record of this journey, this thesis is anything but complete or comprehensive. I hope, however, that what is offered here amounts to at least a thoughtful beginning.

The Outline

This thesis is chiefly concerned with the arguments set out in Chapter Five of Stephen Stich's, *The Fragmentation of Reason*. These are arguments designed with one purpose in mind: To disprove the claim that true beliefs are obviously of intrinsic and/or instrumental value. In examining these arguments the aim will be to clarify and explore the implications of many of Stich's more controversial claims. Often, but not always, the classifications and implications developed in this thesis will weaken, if not neutralize, the conclusions Stich draws.

The organization of the thesis is as follows:

i) Section 1 is an overview of Stich's purpose in writing *The Fragmentation of Reason* and of the arguments he presents in Chapters One through Four. There will also be a brief introduction to the arguments and conclusions found in Chapter Five, which constitutes the main focus of this thesis - particularly, sections 3 and 4.

ii) Section 2 introduces Stich's account of what beliefs are and how it is that they can have the semantic properties of truth and falsity. The account Stich provides here acts as a backdrop for his arguments against the supposed "obvious value of true beliefs".

iii) Section 3 explains and criticizes the first stage of Stich's argument against "the obvious value of true beliefs". The criticisms pertaining to this stage are partly my own, and partly those from the currently available literature.

iv) Section 4 concerns the second stage of Stich's arguments in chapter five. These arguments are directed toward the instrumental and intrinsic

value of having true beliefs. As required, there will be occasional discussion of the version of pragmatism Stich champions in Chapter Six. The sources of criticism here will again be both my own, and those of other authors.

v) Section 5 contains summary remarks.

Section I

OVERVIEW OF THE FRAGMENTATION OF REASON

The Fragmentation of Reason is a book about just that: it is about the fierce intellectual fragmentation which is occurring in much of the conventional philosophical thought regarding cognitive reasoning and certain cognate concepts (e.g., rationality, justification, belief, truth, knowledge, etc.). Stich advances this fragmentation by launching criticism at both analytic philosophy generally and analytic epistemology in particular. All of this philosophical demotion is done so that he might ultimately construct a new foundation upon which to build a better cognitive evaluation strategy - namely, pragmatism.

The first part of this section contains a brief presentation of the first three chapters of Stich's book. The review of these chapters will provide some context for Stich's purpose in writing *The Fragmentation of Reason*; as well as, some of the arguments from the literature which thwart his purpose. The second part of this section is devoted to a more complete exposition of the arguments Stich presents in Chapter Four. The reason for the increased scrutiny of this chapter is that the general argument forms found there will reoccur in Chapter Five, which is the main concern of this thesis. Stich's focus in Chapter Four is on the concepts of

justification and justified belief, and particularly their relation to analytic epistemology.

PART I OF SECTION 1

Stich starts Chapter One by citing the empirical evidence from cognitive psychology studies which purport to show that in many cases, under ordinary circumstances, otherwise intelligent individuals do a very poor job of reasoning (e.g., their inability to draw correct inferences). Some of the experiments cited by Stich include: i) The Selection Task; ii) The Conjunction Fallacy; iii) The Gamblers Fallacy; iv) Pseudodiagnosticity; and v) Belief Perseverance. While the particulars of these experiments will not be examined here, suffice it to say that some philosophers and many psychologists take the results to be indicative of pervasive, consistently poor reasoning involving apparently ordinary cognitive tasks. It is this body of empirical evidence that caused Stich to wonder whether twentieth century philosophical work on such topics as intentionality and rationality might not be inaccurate, if not simply mistaken. These experimental results also led him to another query: On the supposition that we are doing a poor job of reasoning, might not there be some way to ameliorate our cognitive processing strategies and abilities. As Stich himself puts it:

> A very natural reaction to this work is to wonder how we might improve matters. Are there steps we can take that will enable people to do a better job of forming and revising their beliefs? This question suggests two rather different lines of inquiry. One is largely empirical: What sorts of interventions will succeed in changing the way people go about the business of reasoning? The other is normative: What sorts of changes would be *desirable*? What is it that makes one strategy of reasoning *better* than another? These normative questions are centre stage in the second half of *The Fragmentation of Reason*...[1]

In Chapter Two, Stich criticizes philosophy for having largely ignored such possibilities and for having taken, what he calls, normative cognitive pluralism to be an impossibility. Stich defines normative cognitive pluralism as those cognitive processes that people ought to use, as opposed to those that people actually do use (the latter being descriptive cognitive pluralism). Normative cognitive pluralism claims there is no one system of reasoning that people should use, because different cognitive systems may all be equally good or effective.

> ...among philosophers, both historical and contemporary, normative cognitive pluralism is very clearly a minority view. The dominant philosophical view is that there is only one good way to go about the business of reasoning, or at most, a small cluster of similar ways. Good reasoning, philosophers typically maintain, is rational reasoning, and in the view of most philosophers, it is just not the case that there are alternative systems of reasoning differing from one another in important ways, all of which are rational. [2]

This characterization of philosophical thought, favoured by most philosophers, is called cognitive monism, and is strictly opposed to Stich's "floridly pluralistic" and relativistic view of the subject (a view which will not be discussed

here but which he develops in Chapter Six, "A Pragmatic Account of Cognitive Evaluation").

Stich suggests that it is cognitive monism which has, in the hands of philosophers, produced a series of separate arguments which defend what he calls the "irrationality-is-impossible thesis". According to Stich, this thesis makes it impossible to improve or even recognize bad reasoning - a task which forms no small part of his purpose in writing *The Fragmentation of Reason*. Chapter Two is an attempt to weaken, if not completely invalidate, each of the more popular and persuasive arguments in favour of the "irrationality-is-impossible thesis". In Stich's own words:

> ...my concern will be to explore and ultimately to dismiss, a cluster of influential arguments aimed at showing that there are *conceptual constraints* on how badly a person can reason. These arguments maintain that it is simply incoherent to suppose a person's cognitive processing could deviate without limit from the standards of rationality. If the arguments are convincing, there will be limits to the "bleak implications" that might emerge from the empirical exploration of human reasoning. Moreover, if the conceptual constraints are tight and possible deviations from rationality are small, then there is little pressing work for the reform-minded epistemologist to do. If it is impossible for people to deviate very far from ideal standards of good reasoning, then there is not much we can do to improve the way we go about the business of cognition.[3]

The most effective means of presenting the details of this thesis is to briefly examine each of its proponents in the order in which they appear in the book.

Two well known advocates of normative cognitive monism are Donald Davidson and Daniel Dennett. According to Stich, both philosophers maintain that

a high degree of rationality is required in order to provide intentional content to people's beliefs and utterances. By "high degree of rationality" Dennett and Davidson mean that most of our beliefs must be true and that our inference-making ability must produce mainly correct or normatively appropriate inferences. Further, should our cognitive system fail in either of these respects, then our beliefs and verbal utterances will not be assigned any intentional interpretation. Consequently, any inferences that seriously or systematically deviate from this requirement are irrational and therefore devoid of any intentional content.

> ...inference is a process in which beliefs are generated or transformed. But without a high level of rationality and truth there can be no belief, and without belief there can be no inference. Thus it is simply incoherent to suggest that people reason in ways that depart seriously or systematically from what is rational or normatively appropriate. [4]

Stich does not hold much regard for this line of argument. He considers any limits this argument could place on how badly one might reason to be "profoundly uninteresting"; concluding that "neither the conceptual arguments nor the *limits* it imposes need be taken seriously by either empirical psychologists or epistemic reformers".[5] By his lights, the empirical evidence gained from the aforementioned psychology experiments flies in the face of this type of *a priori* argument; supporting, instead, his favoured normative and descriptive cognitive pluralism. In fact, this particular combination of pluralisms produces one of the more radical theses of his book.

> The initial motive for my concern about the Davidson/Dennett thesis that rationality is a prerequisite for cognition was that it threatened to undermine the empirical explorations of irrationality that were producing, and have continued to produce, surprising and unsettling insight into human cognition. A second concern, one that became increasingly important as work on this volume [*The Fragmentation of Reason*] proceeded, was that if the thesis were true, then much of the urgency would be drained from the project of assessing strategies of reasoning and inquiry. The interest and vitality of this branch of epistemological research can be traced..... to the practical worries it addresses: People out there are reasoning badly, and this bad reasoning is giving rise to bad theories, many of which have nasty consequences for people's lives.[6]

The next argument to appear in *The Fragmentation of Reason*, supporting the "irrationality-is-impossible thesis", comes from biological evolution, and is, as Stich describes it, merely "hinted at" in the literature. In fact, Stich has himself had to piece together the best argument he can from his own, and other's, work on the subject. The argument from evolution is unlike the Davidson/Dennett arguments in that the impossibility of systematic irrationality is not grounded in any *a priori* reasons, but rather in the incompatibility of widespread irrationality with the theories and processes of evolution.

> It maintains that *biological evolution* guarantees that all normal cognitive systems will be rational, or nearly so, since organisms whose cognitive systems depart too drastically from the normative standard will run a very high risk of becoming posthumous before they have had a chance to pass on their genes to offspring.[7]

Stich rejects this argument for the impossibility of irrationality saying generally that:

> ...nothing we know about genetics, or evolution, or the acquisition of cognitive systems would even begin to show that descriptive cognitive pluralism is false. It is overwhelmingly plausible that some parts of our inferential system are acquired from the surrounding culture, and it is entirely possible that much or all of the system is a cultural inheritance. Moreover, even if there are parts of our cognitive system that are innate, there is no biological or evolutionary reason to think that those parts do not differ markedly from person to person or from culture to culture. [8]

In concluding his critical analysis of the two previous arguments, which purport to demonstrate that radical or significant departures from rationality are impossible, Stich has this to say:

> The previous two chapters were aimed at opening up a range of possibilities by dismantling the arguments that threatened to foreclose them. One of the possibilities is that there is considerable diversity in human reasoning, diversity that may derive from biological differences, cultural differences, or individual differences in various combinations. The second possibility is that amid this diversity there may be individuals, traditions, or cultures that do a bad job at the business of cognition and that we ourselves may be doing much less well than we might. There is no guarantee, either conceptual or biological, that our own reasoning is good reasoning, or even a close approximation. Thus neither the empirical investigators who are concerned to characterize cognitive shortcomings nor the epistemic reformers who hope to improve cognitive performance need worry that they have embarked on an impossible project. [9]

This concludes Part I of this section. This last quotation sets the stage for what is to follow. Stich places considerable stock in the possibility of radical and significant variability in cognitive strategies, and in the project of evaluating and improving reasoning practices. This stock will produce dividends in the form of

arguments that resonate throughout his book and this thesis. These arguments will now be introduce in Part II of this section.

PART II OF SECTION 1

Before continuing, it is important to note at this juncture that the remainder of this section will be solely concerned with Stich's analysis of certain of our ordinary notions of cognitive evaluation (particularly 'justification') and not his analysis of truth. It is also crucial to point out that, while his critical attention has turned more toward notions of cognitive evaluation, Stich is still partially concerned with the "irrationality-is-impossible thesis". After a review of this portion of the book (Chapter 4), the next section and the body of this thesis, which focuses on *The Fragmentation of Reason*'s most controversial themes, will be presented. It is worth noting, however, that the lines of argument which follow, showing the "undesirability" of justified beliefs, are principally the same as the arguments dealt with in the body of this thesis (Sections 3 and 4), showing the "undesirability" of true beliefs.

To continue then, confident that he has established serious doubt regarding the impossibility of irrationality, Stich makes his final critical advance on

the subject: and, at the same time, begins his discussion on normative cognitive standards and ordinary notions of cognitive evaluation.

> ...I offer arguments against two widely accepted accounts of what it is for one cognitive strategy to be better than another. The first account gives pride of place to the analysis of our ordinary notions of cognitive evaluation - notions like *justification*. The second links cognitive assessment to truth.[10]

Here, his purpose is to hash out what the normative standards of cognitive evaluation - standards which the two previous arguments claimed we could not radically deviate from - actually are. Two authors who take centre stage in this analysis are L. Jonathan Cohen and Nelson Goodman. Cohen questions how it is that many of the subjects in the aforementioned psychology experiments managed to "go on to become leading scientists, jurists, and civil servants", despite having demonstrated rather appalling reasoning skills. "How could they be so successful, Cohen asks, if they do not know how to reason well?"[11] Cohen sees the apparent problem as a failure to recognize the difference between *competence* and *performance*: more accurately, cognitive or inferential competence and performance. The competence of one's reasoning ability involves knowledge of "psycho-logic", namely, "the internalized rules that guide the subject's reasoning and his intuitive judgments about reasoning".[12] A subject's performance, on the other hand, is just that, how well he actually does in the cognitive tasks assigned to him. Cohen then combines this distinction with Goodman's account of what it is for an inferential rule to be justified. For Goodman, an inferential rule or a particular inference are justified

under the following strategy, or, as Stich refers to it, the "reflective equilibrium test": *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected it if violates a rule we are unwilling to amend.* This method for justifying the normative standards of cognition is intended to capture or explicate our commonsense notion of justification - Stich, however, thinks Goodman has desperately failed in this respect. Without detailing Cohen's argument, this combination of commonsense justifactorial practice and cognitive competence/performance evaluation brings him to conclude, as paraphrased by Stich, that:

> ...it is impossible for a person's inferential competence, his underlying psycho-logic, to be anything other than normatively impeccable. [On the other hand], Cohen readily acknowledges that people make inferential errors of many sorts under many circumstances. But he insists that these errors are performance errors, reflecting nothing about the reasoner's underlying, normatively impeccable competence.[13]

In response, Stich argues that neither Goodman's account of commonsense justification (of particular inferences and inference rules) nor Cohen's ingenious cognitive adaptation of the linguistic notions of competence and performance satisfactorily establish the "irrationality-is-impossible thesis". In the first instance, Goodman's account of our commonsense concept of justification does not capture how that concept commonly works.

> On the basis of both controlled studies and anecdotal evidence, we [Nisbett and Stich] argued that patently unacceptable rules of inference would pass the reflective equilibrium test for many people. For example, it appears likely that many people infer in

> accordance with some version of the gambler's fallacy when dealing with games of chance.[14]

In addition he says this:

> It is surely not an *a priori* fact that strange inferential principles will always fail the reflective equilibrium test for all subjects. And if it is granted, as clearly it must be, that the gambler's fallacy...could possibly pass the reflective equilibrium test for some group of subjects, this is enough to cast doubt on the view that reflective equilibrium is constitutive of justification as that notion is ordinarily used.[15]

Stich claims that Goodman's failure is Cohen's failure, in that Cohen's argument is only as good as Goodman's account of our commonsense notion of justification. And according to Stich, even with the addition of some sophisticated "bells and whistles", Goodman's account is lacking.

After having some success in his battle against the "irrationality-is-impossible thesis", Stich proudly dons his reform-minded epistemologist's armor and begins his attack on analytic epistemology and its work on epistemic concepts such as justification and truth. This attack goes to the heart of his concern regarding normative cognitive standards and ordinary notions of cognitive evaluation. As was indicated earlier, the arguments that surface here exercise comparable force in the next chapter of his book (Chapter 5), and, therefore, in the main body of this thesis (Sections 3 and 4).

What Stich thinks of and what he means by "analytic epistemology" is captured in the following:

> I propose to use the term *analytic epistemology* to denote any epistemological project that takes the choice between competing justification rules [those rules that evaluate the justifactory status of beliefs or other cognitive states] or competing criteria of rightness [the set of necessary and sufficient conditions for a set of justification rules to be right] to turn on conceptual or linguistic analysis. There can be little doubt that a very substantial fraction of the epistemological writing published in English in the last quarter century has been analytic epistemology. However, it is my contention that if an analytic epistemological theory is taken to be part of the serious normative inquiry whose goal is to tell people which cognitive processes are good ones or which ones they should use, then for most people it will prove to be an irrelevant failure.[16]

The first issue Stich contends with is analytic epistemology's response to the notion of cognitive diversity. That is, the notion that there are or might be more than one way to reason and evaluate cognitive processes: and that these varied cognitive systems will differ from one another in large and small ways. The import of cognitive diversity to Stich's project is represented by the following passage:

> If we *can* go about the business of cognition differently, and if others actually *do*, it is natural to ask if there is any reason why we should continue to do it our way. Even if we cannot change our cognitive processes once we've acquired them, it is natural to wonder whether those processes are good ones. Moreover, for many people the absence of a convincing affirmative answer can be seriously disquieting. For if we cannot say why our cognitive processes are any better than those prevailing elsewhere, it suggests that it is ultimately no more than an historical accident that we use the cognitive processes we do or that we hold the beliefs that those processes generate....[17]

According to Stich, the analytic epistemologist's answer to which cognitive strategy is the best one, involves analysis of the concepts of epistemic evaluation found in everyday thought and language: that is, our commonsense epistemic notions, such as justification and rationality. Once this is done the analytic epistemologist can then determine which set of justificational rules match these commonsense notions. If our cognitive strategies accord with the right set of justificational rules, then we should continue to use that strategy. If they fail to match, then alternative strategies must be explored. Stich, however, thinks that this approach is mistaken.

> For the analytic epistemologist's effort is designed to determine whether or not our cognitive states and processes accord with our commonsense notion of justification (or some other commonsense concept of epistemic evaluation). ...the analytic epistemologist offers us no reason whatever to think that the notions of evaluation prevailing in our own language and culture are any better than the alternative evaluation notions that might or do prevail in other cultures. But in the absence of any reason to think that the locally prevailing notions of epistemic evaluation are superior to the alternatives, why should we care one whit whether the cognitive processes we use are sanctioned by those evaluative concepts? [18]

Stich then wonders whether people's intrinsic or instrumental values might not operate as evaluative tools, thereby endorsing one cognitive strategy or system over another. In other words, do we *value* having cognitive processes that accord with our commonsense concepts of epistemic evaluation? Stich answers this question in the negative.

> ...it is my contention that when they view the matter clearly, most people will not find it intrinsically valuable to have cognitive states or to invoke cognitive processes that are sanctioned by the evaluative notions embedded in ordinary language. Nor is there any plausible case to be made in favor of the instrumental value of beliefs or cognitive processes that are justified or rational.[19]

For Stich, two considerations work against the intrinsic value of cognitive states sanctioned by commonsense evaluative concepts: i) "the fact that a cognitive process is sanctioned by the venerable standards embedded in our language of epistemic evaluation...is no more reason to value it than the fact that it is sanctioned by the standards of a religious tradition or an ancient text, unless, of course, it can be shown that those standards correlate with something more generally valued or obviously valued"; and ii) "if the most sophisticated recent attempts to analyze our local notions of cognitive evaluation are even roughly on the right track, those notions occupy a small area in a large space of alternative concepts; and there is no obvious virtue that distinguishes our concepts from the alternatives, apart from the fact that we happen to have inherited them".[20]

From these observations, Stich claims that:

> ...once it is seen that the notion we happen to have inherited is but one among many possible alternative notions - most people are not much inclined to say that they find having justified beliefs [justified under our commonsense evaluative concepts] to be *intrinsically* valuable. Since our notion of justification is just one member of a large and varied family of concepts of epistemic evaluation, it strikes most people as simply capricious or perverse to have an intrinsic preference for justified beliefs.[21]

Stich then goes on to assess two lines of argument aimed at establishing the instrumental value of our ordinary, commonsense evaluative notions (i.e., justification and rationality). Presently, however, only an exposition of the first line involving the social or biological evolution of these ordinary notions will be offered, while the second line, pertaining to truth, will be left for the next section of this thesis. Briefly stated the first line of argument claims that because we do in fact operate on "highly evolved intuitive notions of justification and rationality", justified and rational beliefs must clearly be instrumentally valuable; otherwise, we would surely not have evolved, either socially or biologically, to operate on them. Stich is dumfounded that anyone would endorse such a view, saying:

> The obvious conclusion to draw here is that neither biological nor social evolution can be relied upon to produce the best of all possible options, or even one that is close to the best. So the fact (if it is a fact) that our intuitive notions of epistemic evaluation are the product of an extended process of social and/or biological evolution does not make it plausible that they are more conducive to survival or thriving (or anything else) than any of the alternative notions of epistemic evaluation that might be invoked instead. [This is very much the same analysis used to defeat the argument from evolution supporting the "irrationality-is-impossible thesis" mentioned earlier in this section.][22]

IN SUMMARY

This section of the thesis is thus concluded. The function of this section was to reveal Stich's overall purpose for writing *The Fragmentation of Reason*.

From this overview of the book, it should be clear that Stich's project, in general, is epistemological reformation. That is, he feels that the more highly reputed methods of evaluating cognitive processes are mistaken and ineffectual: and that because of this there is little chance that we can improve our cognitive processes (including reasoning, belief formation, inference-making, etc.). In fact, not only are we unable, under the present methodologies, to improve our cognitive processes or strategies in the face of empirical evidence which strongly indicates they need improvement, even if we could we would not be able to tell whether we had yet achieved the best or optimal cognitive system. In other words, these methods of cognitive evaluation, including analytic epistemology, are impotent in terms of their ability to objectively compare and contrast the merits of varied cognitive strategies.

The next section of this thesis critically examines one of the more potent lines of argument found in Stich's book (Chapter 5) - a line of argument which purports to establish that there is no obvious value, either intrinsic or instrumental, in forming and having true beliefs. As was mentioned earlier, this same line of argument was also used in the last half of this section to criticize the effectiveness of our commonsense notion of justification as a cognitive evaluative tool. With respect to our common, everyday evaluative notion of justification, Stich found that it suffered from at least two ailments: i) it was idiosyncratic and ii) it

was partial. Both afflictions were the result of the failure of analytic epistemology and value theory to account for either real or possible cognitive diversity.

These two ailments are also acute in Stich's diagnosis of true beliefs. But before a detailed analysis of Stich's arguments against the value of true belief is offered, some explanation of what Stich takes beliefs to be and how it is that they can have semantic properties is in order.

Section 2

WHAT'S A BELIEF?

The purpose of this section is to lay out Stich's account of what beliefs are and how it is that they can have, what are normally considered, the semantic properties of truth and falsity. This involves an explanation of his view of beliefs as "real psychological states" and not merely behavioral explanatory devices. There will also be discussion of the role of "interpretation functions" in the process of mapping mental states onto entities with semantic properties such as propositions, truth conditions, states of affairs, or possible facts: and what the standards of a good interpretation function are. Tarski's theory of truth, Putnam's and Kripke's causal theory of reference, and functionalism as it is found in philosophy of mind will also be included in the discussion. It is important to recognize that the discussions in 5.2 and 5.3, of *The Fragmentation of Reason*, are to serve as scaffolding from which Stich intends to construct his argument against the claim that true beliefs obviously have either intrinsic or instrumental value.

As was indicated at the outset, Stich takes beliefs to be "real psychological states." In explaining what a belief is, he adopts the "token-identity hypothesis". According to Stich, this hypothesis claims:

> ...that each instance (or token) of a belief is identical with some neurophysiological state or other. though it does not endorse the type-identity hypothesis. which holds that the same belief type in different individuals is always identical with the same neurophysiological state type.[23]

In essence then, Stich concludes that belief-state tokens are brain-state tokens. However, this leaves him with a puzzling predicament: "What is it for a brain-state token - a neurophysiological state or happening - to be true or false?"[24] On the one hand, beliefs are widely recognized as having semantic properties such as being true or false; but, on the other hand, neurophysiological states are not. Stich proposes to solve the puzzle in the following manner:

> One familiar framework in which an answer can be developed posits the existence of a function that maps certain brain-state tokens (including beliefs and perhaps some others) onto entities that are more naturally thought of in semantic terms, entities like propositions, or content sentences, or specifications of truth conditions. A variation on this idea posits a function that maps brain states onto entities like possible facts, states of affairs or subsets of the set of all possible worlds. An account of what it is for a belief token (i.e., a certain brain-state token) to be true can then be given in terms of the entity to which it is mapped: the belief is true if and only if the proposition (or content sentence) to which it is mapped is true; or, if and only if its truth condition obtains; or, if and only if the possible state of affairs to which it is mapped is actual; or, if and only if the actual world is one of the possible worlds to which the belief is mapped.[25]

Now Stich recognizes that, once the mapping is completed, there is the further difficulty of determining whether the propositions and content sentences in question are themselves true or false. He says:

> Unless we have some coherent story about what it is for a content sentence or proposition to be true, the mere fact that we can map beliefs to propositions or content sentences in some well-motivated way will not tell us what it is for beliefs to be true. And without an account of what it is for beliefs to be true, it is all but impossible for us to think clearly about whether we value having true beliefs.[26]

Since Stich's arguments against the value of having true beliefs are concentrated on the particular "interpretation function" which maps neurophysiological states onto propositions, he is willing to assume, for argument's sake, that there is in fact some unproblematic story about what it is for content sentences or propositions to be true. With reference to the particular interpretation function in question, Stich has these general comments:

> [Before] setting out the theory, a word is in order about what standards a theory of interpretation aspires to meet. What are the constraints that govern how the game is played? [27]

> ...in developing a theory of interpretation, we are attempting to explicate and explain a well-entrenched preexisting intuitive concept or capacity. We do, after all, ascribe content to people's psychological states all the time. ...it is crucial that any proposed theory agree, by and large, with the judgments of the man or woman in the street about what content sentences or truth conditions get paired with the ordinary beliefs of ordinary folks. [28]

Additionally, Stich has this to say in response to the challenge that any given theory of interpretation based on intuitive judgments or commonsense practice might have to give way to a more powerful or more elegant theory which is not closely related to commonsense practice:

> I am inclined to think that if an interpretation function does not cleave reasonably closely to commonsense practice, it is hard to see why what the function is characterizing deserves to be considered a *truth* condition. ...my skepticism about the value of true beliefs is restricted to accounts that assign truth conditions largely compatible with commonsense intuition. [29]

In this way, where theory and commonsense practice conflict, Stich sides with commonsense practice irrespective of the theory's "technical merits". A line of criticism which will be developed in the next section argues that Stich has, in criticising our commonsense, intuitive interpretation function, violated or ignored the very standards which he himself sets for all interpretation functions. For now, more needs to be said about this intuitively sanctioned, standard interpretation function and what it is to get the mappings right.

It is here that Tarski's theory of truth, the Putnam-Kripke causal theory of reference, and functionalism, as it is found in philosophy of mind, take center stage. These are the three main components of a theory of interpretation which Stich characterizes as "justifiably popular" and which he adopts as the setting for his arguments against the value of true beliefs. Stich calls this interpretation theory "the causal/functional theory".

Stich describes Tarski's role in the formulation of this interpretation function as follows:

> Tarski's theory of truth shows us how to construct an axiomatic theory about a language that will specify a truth condition for each of the infinitely well-formed sentences, of the language. That is, the theory will entail an infinity of theorems of the form: (T) S is true if and only if p where 'S' is replaced by a structural descriptive name of a sentence in a language and 'p' is replaced by a metalanguage sentence specifying the conditions under which the sentence is true. So, a substantial list of axioms - the so-called base clauses of the recursive truth definition - specifying the semantic properties of the language's noncompound predicates and names must be constructed in order to fully develop a Tarskian truth theory. For instance:

(2) (x) satisfies 'is true' if x is true

...

(3) 'Socrates' denotes Socrates

...[30]

Stich indicates that Tarski's theory of truth has a limitation which is important in the development of the causal/functional theory of interpretation. The limitation is this:

> What generates a problem is that Tarski tells us too little about what it is to get these axioms right. He does not tell us what sort of relationship must obtain between a name and a person if the former is to denote the latter. Nor does he tell us what relationship is to obtain between a predicate and a satisfaction condition if the former is to be satisfied by (and only by) things that fit the latter.[31]

This limitation is overcome by the introduction of the Putnam-Kripke theory of reference, or the causal theory of reference. Stich describes this theory as follows:

> ...the basic idea of the causal theory is that a token of a name denotes an individual if and only if the appropriate causal chain extends from an original use or dubbing to the current production of the name token in question. A broadly similar account can be given for natural kind predicates.[32]

Stich says this with respect to the role the causal theory of reference plays in his argument:

> We need not dwell on the details of the causal story, since they will play little role in the argument set out below. But it is important to note that both the basic argument in favor of the

> causal theory of reference and the detailed working out of the theory rely heavily on commonsense intuition and the commonsense concepts or practices that underlie them.[33]

Further:

> There are...endless varieties of causal chains in the world linking all sorts of events in all sorts of ways. So for my tokening of 'Aristotle' to refer to the great philosopher, it is not enough that there be *some* causal chain linking my utterance to Aristotle's baptism; it has to be the right sort of causal chain. Typically, a theorist will try to show that his account of the relevant sort of causal chain is correct by showing how the implications of the account comport with intuition. If our utterances are linked by the theorist's favoured causal chain to people and objects that intuition insists we are not talking about, it is generally concluded that the theorist's account of the causal chain required for reference is defective.[34]

After having sketched this account of what it is for sentences in a natural language to get their truth conditions, Stich considers how this all applies to the establishment of the same for mental states, or beliefs. The causal/functional theorist makes the connection by, as Stich describes it, simply putting the sentences of the language inside the head.

> The idea is that beliefs are complex psychological states which, like sentences, can be viewed as built up out of simpler components. So by mapping the elements out of which beliefs are constructed to the symbols of some uninterpreted formal language, in a way that preserves well-formedness, we can associate belief tokens with well-formed formulas in that language... To have a belief, then, is to have a token of a well-formed formula stored appropriately in one's brain. The question of how beliefs get their semantic properties can now be rephrased as a question about how we can assign truth conditions to these cerebral inscriptions.[35]

Now that the causal theory of reference has provided us with the necessary means of determining the denotations of extensions of word types such as names and predicates in mental sentence inscriptions, one final piece of the puzzle remains to be fitted. Word types such as quantifiers and connectives need to be accurately identified if we are to fully make sense of the belief inscriptions and then assign truth conditions. Stich says that functionalism has provided us with just such a means of identifying the logical structure of belief inscription:

> What is important is the idea, inspired by functionalism, that mental sentence inscriptions have the logical form they do in virtue of the pattern of causal interaction they manifest with other mental-state inscriptions.[36]

So, if the patterns of causal interaction manifested between the well-formed sentence, 'P * Q', and other mental-state inscriptions approximates the pattern normally associated with the logical rule of conjunction, then * is the symbol for conjunction.

With the details of the interpretation function under discussion having been laid out (an interpretation function which is sanctioned by commonsense intuition), it is now time to move on to Stich's criticisms of this function and the notion of truth associated with it. But before we proceed any further, it should be admitted that there are a considerable number of areas within the details of this interpretation function which could themselves be critically analyzed and explored.

However, for the moment, Stich's presentation of what he calls the "causal/functional interpretation function" will be accepted and direct attention will instead be given to the implications he claims result from the adoption of this function - implications which ultimately purport to undermine any arguments for the value of true belief.

Section 3

STAGE I: PARTIALITY AND IDIOSYNCRASY

With an account of how beliefs can be seen to have semantic properties in place, Stich now takes issue with a specific feature of that account - the causal/functional interpretation function. According to Stitch, the implementation of the causal/functional interpretation function has two obvious implications which make the arguments for the value of true belief considerably more difficult to construct. The first is that this particular interpretation function is limited or partial. The second is that it is idiosyncratic.

PARTIALITY:

An examination of the first consequence of the causal/functional interpretation function will now be undertaken. Stich has this to say generally:

> The first consequence of the causal/functional account of interpretation that I want to draw attention to is that the interpretation function that it favours is a very partial function. The belief-like mental states for which it provides a specification of truth conditions constitute a small subset of the possible belief-like mental states that a human or organism might have. [37]

Stich says that there are two reasons why this interpretation function is partial. The first is the causal notion of reference captured by this function. The second is the account of logical form endorsed by functionalism.

> On the causal side, the point is simply that any plausible specification of the kinds of causal chains required to fix the reference of mental words (or concepts) will entail that these chains are far from ubiquitous. On any account that purports

> even roughly with commonsense intuition, the reference-fixing causal chains are going to cut a relatively narrow swath through the space of empirically possible causal histories of mental words. Thus, there will be all sorts of ways in which a mental word can end up as part of a speaker's mental lexicon, though it is not tied to anything in the world by the special kind of causal rope that the causal theory requires for reference. These mental words may, of course, stand in a variety of other causal relations to a variety of extramental objects or kinds. But the causal account entails that they will not refer to any of these objects or to anything else. And since these mental words have no referents, the mental sentences in which they occur will have no truth conditions assigned by the causal/functional interpretation function.[38]

Some clarification is needed with respect to these last two passages, especially as they relate to the standards Stich sets for a good interpretation function. (This is an issue of clarification which will surface from time to time throughout the analysis of Stich's arguments against both the causal/functional interpretation function and the value of true beliefs.) One way to read these passages (and numerous others like them in the book) is to view the alternative causal chains or interpretation functions referred to - the ones outside the relatively narrow swath of functions that comport with commonsense intuition - as not comporting (maybe even roughly?) with commonsense intuition. If this is true, then one is forced, strictly speaking, to conclude that these alternative functions violate Stich's own standards of how an interpretation function is supposed to operate. In short, they have violated the rules of the game. After all, Stichian standards require that any interpretation function must (or should) cleave reasonably closely to commonsense practice and intuition (see pages 22 and 23 above). This being the case, the alternative functions

alluded to in the two previous passages would be substandard. On numerous occasions throughout Chapter 5, Stich provides ample support for just such a reading, referring to his prized alternative interpretation functions as though they were indeed counterintuitive and in opposition to commonsense practises. However, to construe Stich in this fashion would be unfair.

The more favourable and plausible reading would posit alternative interpretation functions which did in fact comply with commonsense intuition, only not our own, but someone else's - some other individual, or group of individuals, or perhaps some other culture. In this way, the alternative interpretation functions, which are so essential to Stich's arguments in Chapter 5, could be seen as standard, acceptable functions. To read Stich otherwise would force us to conclude that the only reason the causal/functional interpretation function, a function sanctioned by our commonsense intuition, is partial (and idiosyncratic) is that it is one standard interpretation function among a whole raft of alternative, substandard functions. This seems like an untenable line of argument for Stich to take. In fact, there is one passage that lends credence to this more charitable reading.

> These alternative interpretation functions are not the ones sanctioned by our intuitive judgments. They strike us as wrong or inappropriate. But there is no reason to think that we could not retrain our intuitions or bring up our children to have intuitions very different from ours. And having done so, interpretations based on 'the causal theory of' reference would strike us as inappropriate while interpretations based on REFERENCE* 'some alternative theory of reference'... would seem intuitively natural.'[39]

So, when Stich says that his "skepticism about the value of true beliefs is restricted to accounts that assign truth conditions largely compatible with commonsense intuition", he must mean *our, present-day* commonsense intuition and not, as he so often suggests, commonsense intuition in general.

To continue then, the second way in which the causal/functional interpretation function is partial involves the functional account of logical format. Stich says:

> The way in which the functional account of logical form restricts the domain of the causal/functional interpretation function is a bit less obvious, though no less important. The tip of the iceberg was already noted in my brief remarks on the limitations of the Tarskian truth theories, where it was pointed out that there is only a very limited class of constructions for which we know how to give a Tarski-style account of how the truth conditions or referential properties of compounds depend on the referential properties of their components. Once we get beyond the truth functional constructions and standard quantifiers and attend to modal or adverbial or counterfactual constructions, it is not even clear what would count as getting the semantics right. For, as Scott Soames and Robert Stalnaker have noted, we simply do not have for connectors, quantifiers, and other constructions anything like the causal theory of reference for names and kind terms. We have nothing that will tell us whether a proposed account of the recursive rules governing such constructions is correct or incorrect.[40]

Further:

> ...Without some general account of what it is to get the recursive clauses in a truth definition right, the only compound mental sentences in the domain of the causal/functional theorist's interpretation function will be those built from the very limited number of constructions whose projected semantic properties are relatively well understood and for which we already have the requisite recursive clauses.[41]

Stich points out though that even if we could meet this requirement, there would still be many mental sentence constructions for which there are no adequate truth theoretic recursive clauses. Here is why Stich takes this to be true:

> First, recall that the project of providing an account of the interpretation function is an exercise in explication that must ultimately be responsible to our intuitive judgments about content or truth conditions. Next, note that in individuating the "constructions" from which mental sentences are built, the patterns of causal interactions that they manifest play a central role. What makes a mental sentence a conjunction is the fact that it interacts in other sentences in ways that mirror, by and large, what logic permits. Similarly, we could identify modal, counterfactual, and other sorts of constructions in a mental language (whether or not we have a suitable truth definition for such constructions) by noting that the patterns of inference they exhibit largely accords with what is intuitively logically possible. But, and this is the central point, there are indefinitely many possible patterns of formally specifiable causal interactions among mental sentences and thus indefinitely many possible mental sentence constructions, which admit of no intuitively plausible semantic interpretation at all. *Most purely formal, syntactically characterizable patterns of interaction among sentences or well-formed formulas have no intuitively plausible semantics.* ... The space of formally (or syntactically) plausible productions or computations vastly outruns those that our intuitive semantics is prepared to interpret. [42]

IDIOSYNCRASY:

Stich concludes his discussion of the first consequence of the causal/functional theory of interpretation and introduces the function's second consequence - idiosyncrasy.

> What I propose to argue now is that the causal/functional interpretation function is not only limited, it is also highly idiosyncratic. Even in the domain where it specifies interpretations, there are lots of other functions that map mental states to truth conditions (or propositions, or states of affairs,

> etc.), and there is nothing obviously superior or preferable about the one sanctioned by commonsense intuition.[43]

Some discussion of the meaning of the word "idiosyncratic" could well prove useful here, particularly since Stich does not reveal how he intends the word to be understood. In a paper entitled, "The Anastylosis of Reason: Fitting Together Stich's Fragments", David H. Sanford, examines just this issue. Sanford's analysis unfolds as follows.

> An idiosyncrasy is a peculiarity particular to its processor which expresses the strong individuality of its processor. Sleeping with a fountain pen and a pad of writing paper under one's pillow is an idiosyncrasy. 'Eccentricity', unlike 'idiosyncrasy', implies considerable deviation from the norm. Soaking one's feet in Diet Coke for an hour before each logic class is an eccentricity. Having one specific property, a member of a genus that comprises a huge (even infinite) number of alternative, mutually incompatible, specific properties, is not sufficient for idiosyncrasy or for eccentricity. Everyone is unusual and unique in many respects. That does not bestow everyone with an individuality strong enough for an idiosyncrasy to express.[44]

Sanford feels that what is missing in Stich's talk of the idiosyncrasy of justified or true beliefs, and the causal/functional interpretation function in turn, is this "expression of strong individuality". And, according to Sanford, without that telltale expression, idiosyncrasy is not present. Sanford asks:

> Does our cultural tradition have strong individuality? The stock response, 'Compared to what?', underlies the difficulty in applying this concept to cultures, species, or possible worlds to which we belong. If we agree with Stich that justification and true belief [as well as the causal/functional interpretation function] are each just one member of a large family, and we do not assume, without further reason, that justification and true belief [and the causal/functional interpretation function] *do* have obvious advantages over the other members of the family, we still lack grounds for the claims that justification and truth are

> idiosyncratic. Stich offers no additional support for his claims of idiosyncrasy, which he repeats throughout his main argument.[45]

The preceding is offered only as a precautionary note: and, for reasons of expediency, Stich's treatment of the causal/functional interpretation function as an idiosyncratic function will now continue.

In the process of explaining the idiosyncratic nature of the causal/functional account of interpretation Stich takes a moment to examine how causal theorists specify which sort of chain fixes reference.

> Typically, these stories divide into two parts, one part focusing on the process of "grounding" or "reference fixing," whereby a name or predicate is introduced into a language to designate an object or class of objects, the other part focusing on the process of social transmission whereby the name or predicate is passed from one user to another, preserving the reference that was fixed when the term was introduced. In each part, the task for a serious causal theorist is to specify the kinds of events or processes that count as legitimate groundings or legitimate transmissions. And, as ever, an important part of the criterion of legitimacy is how well the resulting story accords with intuition.[46]

More specifically Stich has this to say on reference fixing and transmission:

> When one looks at the sorts of accounts of grounding and transmission that emerge, it appears that in each category the allowable events are a mixed bag having at best only a loosely knit fabric of family resemblances to tie them together. Nor is it at all surprising that things turn out this way. Proper names and nicknames get affixed to all sorts of things - babies, popes, battles ships, breakfast cereals, islands, wars, and tyrants, to name just a few - and the baptismal processes typically involved differ markedly from one sort of object to another. It is hard to believe that they constitute anything like a natural kind. The heterogeneity of intuitively acceptable groundings grows even more extreme when we consider the ways in which predicates come to be paired with their extensions. 'Gold', 'helium',

> 'asteroid', 'electron', 'kangaroo', and 'superconductivity', are, presumably, all natural kind terms, but their groundings are sure to have been very different from each other in all sorts of ways. The processes of reference-preserving transmission are comparably diverse.[47]

Stich assures us that this is not meant as a criticism of the causal theory of reference and that what he ultimately wants to claim is this:

> My point is simply that any plausible elaboration of the causal story will specify *lots* of allowable causal patterns. The causal chains linking my mental tokens of the names of my children to the appropriate young people are very different from the causal chains linking my mental token of 'Socrates' to Socrates. And both of these chains are notably different from the one linking my mental token of 'water' with water and the one linking my token of 'quark' with quarks. What ties all these causal chains together is not any substantive property that they all share. Rather, what ties them together is that commonsense intuition counts them all as reference-fixing chains.[48]

At this point Stich makes his first move in the development of his argument for the idiosyncrasy of the causal functional interpretation function:

> But now if it is indeed the case that commonsense groups together a heterogeneous cluster of causal chains, then obviously there are going to be lots of equally heterogeneous variations on the commonsense theme. These alternatives will depart from the cluster favoured by commonsense, some in minor way and some in major ways. They will link some mental words, or many, to objects or extensions different from those assigned by commonsense intuition. In doing so, they will characterize alternative notions of "reference" - alternative word-world links - which we might call REFERENCE*, REFERENCE**, REFERENCE*** and so on. And the only obvious complaint to lodge against many of these alternative schemes for nailing words onto the world is that they do not happen to be the scheme sanctioned by our commonsense intuitions.[49]

As was mentioned earlier, it is vitally important to remember that whatever these other (interpretation) functions which map mental states to truth

functions are, they must, or at least should, adhere to Stichian standards of a good interpretation function (see pages 22 and 23 above). It seems the only way to achieve this is to assume that Stich means these alternatives depart, in minor and major ways, from the cluster favoured by *our* commonsense intuition and not commonsense, strictly speaking. If it were otherwise, then, as was argued earlier, it is hard to see why we would (or could?) recognize them as viable, alternative standard interpretation functions. It should also be mentioned that, to his credit, we do in fact see Stich being a little more careful in this matter. He claims that the only obvious complaint to raise against these alternative functions is that they are not endorsed by *our* commonsense intuitions. From this, it can be safely inferred that, given his interpretation function standards, Stich is suggesting these alternatives are indeed sanctioned by commonsense intuition, only not our present-day intuition.

In a rare moment, Stich provides some examples to better illustrate his comments. One of these examples uses Jonah from the Bible. According to Stich, causal theorists maintain the term 'Jonah' can still refer to an historical person even if all the *fishy bits* about his life are fictitious. In this way, reference is possible for the causal theorist even where there is widespread mistaken belief about the person or object in question. Stich suggests that one of the "equally heterogeneous variations on this commonsense theme" might involve widening the margin of cases where failed reference can occur. One way to do this, in slight opposition to the causal

theory of reference, would be to consider more instances of mistaken belief or false information as indicative of failed reference. In particular, Stich has in mind a hybrid between the causal and the description-cluster theories of reference. This would produce REFERENCES*, which are different from references, as determined under the causal theorist's formula.

> Let REFERENCE* be a word-world relation just like reference save for the fact that if the majority of the (nontrivial) descriptions a speaker associates with the name actually apply to no one, then the name is empty. Thus, if there was an historical person about whom legend gradually developed, 'Jonah' refers to this person though 'Jonah' REFERS* to no one. Another variation on the commonsense theme - REFERENCE** - might give descriptions a somewhat different role in determining the reference of proper names, so that 'Jonah' might end up REFERRING** to some long-forgotten ancient who actually did survive three days in the belly of an aquatic creature. And REFERENCE*** might be designed so that 'water' includes in its EXTENSION*** not only H_2O but also the famous stuff that looks and tastes just like it, XYZ. ...Alternatives of a slightly different sort can be generated by varying the allowable patterns of transmission which preserve reference as a word is passed from one speaker to another.[50]

After having thoroughly laid out his explanation of alternative theories of reference and their corresponding interpretation functions, Stich makes the following concluding remarks.

> These alternative interpretation functions are not the ones sanctioned by our intuitive judgments. They strike us as wrong or inappropriate. But there is no reason to think that we could not retrain our intuitions or bring up our children to have intuitions very different from ours. ...There is, in short, no reason to think that these alternative interpretation functions might not be the intuitively plausible ones for other people or for our own future selves. And there is no reason, or at least no obvious reason, to think that people whose intuitions diverged from ours in these ways would be any worse off. It is in this sense that the

> causal/functional interpretation function is not only limited but also *idiosyncratic*. It is one interpretation function among many that stands out among its fellows principally because it is the function favoured by local, contemporary commonsense intuition and the largely unknown psychological processes that underlie that intuition.[51]

There is much that needs to be said regarding these final remarks on the partiality and idiosyncrasy of the causal/functional interpretation function and the arguments upon which these remarks rest. Discussion will begin with the central concept at issue here - namely, commonsense intuition - particularly since Stich neither analyses the concept nor provides us with an account of how he uses the concept. Next , Stich's claim that there is "no reason to think that we could not retrain our intuitions or bring up our children to have intuitions very different from ours" will be examined. And finally, Stich's declaration that there is no obvious reason to believe that alternative interpretation functions would leave us any worse off will be scrutinized.

Since much of Stich's critique of the causal/functional interpretation function, along with his critique of truth and true beliefs, is concerned with commonsense intuition it seems prudent to review some deficiencies in the notion of commonsense intuition which Stich seems to be employing. From the canvassing of Stich's arguments thus far and from much of what will follow in the next section, it is possible to read Stich as operating on a somewhat restricted understanding of commonsense intuition. In fact, his entire schema of alternative interpretation

functions seems predicated on a restricted conception of commonsense intuition. It is restricted in the sense that, for Stich, commonsense intuition appears to amount to a static, finite list of intuitions. In this way, any and all referential mappings which are counterintuitive or a variation on the commonsense theme, are immediately associated with alternative functions and considered viable competitors for our commonsense interpretation function. As evidence of this, recall his claim that because commonsense does nothing more than group together heterogeneous clusters of causal chains, then obviously there are going to be lots of equally heterogeneous variations on the commonsense theme: and that these alternatives will depart from the cluster favoured by commonsense, some in minor ways and some in major ways. He claimed that consequently these alternatives will link some mental words, or many, to objects or extensions different from those assigned by commonsense intuition. He then says that these *alternative* functions are not the ones sanctioned by our intuitive judgments - they strike us as wrong or inappropriate. But is this how our commonsense really works? And further, is this an accurate and comprehensive account of how genuine competition for our commonsense interpretation function is generated? Some observations about the nature of commonsense would seem to indicate that the answer on both accounts is, no.

Commonsense is both a part of and the product of a dynamic process, which includes the continuous assimilation and assessment of new information and

experiences and the reassessment of old information and experiences. At minimum, this process includes the judgment or assessment of arguments, evidence, reasoning, and logic - both new and old, as the circumstances dictate. In the specific case of interpretation functions and the theories of reference associated with them, these judgments are employed, along with our commonsense intuitions, in order to decide whether a particular word-world mapping is acceptable or not. In this way, even mappings which are counterintuitive might be accepted by our commonsense interpretation function, on the grounds that the evidence available and/or the reasoning/logic employed (i.e., the arguments presented) are enough to *override* our commonsense intuition about the counterintuitive nature of the mapping. Also in this way then, commonsense intuition is not static, it does not consist of a finite list of unchanging intuitions. As *overriding* considerations demand, certain of our intuitions will have to be amended or dropped while others will be reinforced or reaffirmed. The interesting point, which Stich does not appear to recognize, is that none of this requires the creation of competing interpretation functions, each time our commonsense intuitions are amended or dropped. Stich's assumption that the only way word-world mappings which are counterintuitive (that is, not sanctioned by our intuitive judgments) can be made intuitive is to adopt an alternative interpretation function seems unfounded. Our present-day function regularly manages to assimilate and accommodate counterintuitive mappings or references.

For instance, the numerous developments in scientific theory which often initially presented themselves as highly counterintuitive, but which were nonetheless ultimately endorsed by our commonsense interpretation function - Copernicus' theory of the solar system; Galileo's law for descending objects; Newton's inertia and gravity (field) theories; Darwin's theory of natural selection; : Lavoisier's discovery of oxygen; to name but a few.

You may also recall that Stich suggested we could retrain our intuitions or raise our children to have intuitions other than our own and in this way we could effectively embrace and operate on alternative interpretation functions and the counterintuitive mappings they endorse. However, as has been suggested, the dynamic process of intuition amending and forfeiting does not necessarily require the adoption of alternative interpretation functions in every case - or maybe even in any case - that a counterintuitive mapping occurs. Our interpretation function and the notion of commonsense associated with it is highly adaptive. Our commonsense function regularly engages in the amendment and forfeiture of existing intuitions and the adoption of new and often counterintuitive intuitions. In fact, it would be safe to characterize the nature of commonsense intuition (and the particular intuitions it produces) as highly mutable and open-ended. Therefore, that we should or could *retrain* our intuitions, so to speak, in order to accommodate counterintuitive mappings is nothing new to our commonsense interpretation function, thereby

making, at least in some cases, the adoption of allegedly alternative functions unnecessary.

This is not to say that some counterintuitive mappings simply could not be accommodated by us, and therefore a genuine alternative interpretation function might be the result. In such a case, if we decided that the adoption of such an alternative interpretation function was in our best interests, then we *would* likely have to retrain ourselves or raise of children to have intuitions other than our own. But what would this training and retraining be like? Certainly it seems possible that, with some rigorous retraining, we might find that new and different intuitions (and interpretation functions) could become "well-entrenched" in those of us who operate on our present-day set of intuitions. As to the children, it also seems possible for alternative intuitions and functions to become well-entrenched; although, how *well-entrenched* they would be is somewhat dubious since the effectiveness of a teacher (parent), whose well-entrenched, preexisting intuitions are different from those being taught, would have to be called into question. This is especially true if we accept Stich's characterization of the psychological processes that underlie intuition as, "largely unknown". Depending on how, and if, these psychological processes unfold, training and retraining in this area might prove a rather difficult, if not impossible, task. In the end, one would have thought some argument was in order for Stich's claim that there is no reason to think we could not retrain our intuitions or raise our

children to have intuitions other than our own - that is, counterintuitions that could not be accommodated by our present-day commonsense intuition. On the other hand, it seems plausible that cultures other than ours could harbour well-entrenched, preexisting intuitions different from our own.

The final claim of Stich's which needs to be scrutinized is that there is no obvious reason to think that interpretation functions which diverged from ours would leave people any worse off. As the previous paragraph indicates, no serious objections are to be registered with respect to the possibility of alternative intuitions and interpretation functions. However, this does not then give Stich a free hand to claim there is no obvious reason to think that interpretation functions which diverged from ours would not leave us or our children or other cultures any worse off. He needs to provide an argument for this, because intuitions can be misguided or mistaken, resulting in serious consequences. Some of the ways in which alternative interpretation functions could leave us worse off are brought to center stage in the following criticisms from Fred Dretske and Alvin Goldman.

Fred Dretske, rather pragmatically, is curious how this is all supposed to work in concrete cases. He provides the following example:

> Consider the brain state that, on our local common-sense interpretation function, is assigned the content - the door is closed. What content might this brain state be assigned on some alternative interpretation function? Well, almost anything. It could, I suppose, be given the content: there are giraffes in Africa. ...Or 'getting a little closer to home, causally speaking,' my brain state might be assigned the content: The bedroom window is

> open. Or: my car will not start. ...It is, of course, hard to see why one would want to assign any content other than some proposition about the door, since it is the door I see, the relevant state of my nervous system is being caused by some state of the door (presumably, its being closed) and is, in turn, causally engaged in door-opening behavior. Of what possible value could an interpretation function be that gave me... irrelevant beliefs (ie: about giraffes, windows, and cars) - irrelevant for the purpose of explaining and understanding the causal interactions in which the brain state is engaged? Stich is right about one thing - if we are not careful about the interpretation function, the value of truth is no longer obvious. But this shows not that truth is not a value, but that it is the truth of what we actually believe (on the standard interpretation function) that is the value. ...What we value is the truth of the belief that the door is closed because it is the truth of this belief that helps explain why I open, rather than run into, the door I see.[52]

Dretske provides a further analogy to better demonstrate the obvious reasons why we should value the causal/functional interpretation function and the true beliefs it can produce.

> The interpretation function (for these beliefs) is, I suggest, about as obvious (and this is *very* obvious) as that for simple gauges and instruments - the speedometer in my car, for instance. When the pointer is at *n*, the standard interpretation for the device assigns it the content: "you are going *n* miles an hour." We could, I suppose, assign it a different meaning - that I had, say, *n* gallons of gas left, or that my oil pressure is *n*, or that I am going $1/n$ miles/hour. On these interpretations, truth (actually truth* or truth**) would not I agree be of much value. It would not be of much value because the device is not causally engaged with these conditions (the conditions defining its truth conditions) in such a way as to co-vary with them when things are working right. For such devices, just as for beliefs, we want truth conditions to be conditions the device is capable of *tracking*, conditions about which the belief (under normal conditions) carries information. But that, I suggest, is exactly what the standard interpretation function does for our perceptual beliefs.[53]

Dretske is quick to point out that what he has said about truth and the interpretation function associated with our commonsense intuition has equal force

where the beliefs are other than direct observational beliefs. To demonstrate this point, Dretske employs, the now familiar example of Jerry Fodor being invited to lecture in another state and all parties (Fodor and his hosts) successfully meeting in the airport at the appointed time. He says that under any other interpretation function this "amazing convergence" would not have materialized. He continues:

> All this [is] perfectly explicable, not to mention (barring accidents, etc.) routinely predictable, by appeal to what, on the common-sense interpretation function, these people believe. Is there any reason to think this function, and the idea of truth that goes with it, useful? I should say that the reasons are overwhelming.[54]

Goldman too makes similar observations regarding the efficacy of alternative interpretation functions. Goldman recognizes the conclusion Stich intends to draw from his partiality and idiosyncrasy arguments and paraphrases it as follows: "Once we learn about alternative I.F.s [interpretation functions], why should we value being in brain states that turn out true on the ordinary I.F. (true beliefs) rather than brain states that turn out true [or as Stich says TRUE*...*] on some nonstandard I.F....?"[55]

> ...many of the non-standard I.F.s would not link up with action in helpful ways. Suppose that brain state B* is linked via a non-standard reference to an item X that would be a useful instrument for achieving a certain goal; and suppose that B* is assigned the content, "X would be useful," which is true. But also suppose that B* is so connected with other brain states as to lead the agent to grasp and deploy object Y, not X. (It is not clear that Stich's constraints preclude this.) Then though B* is true on the non-standard I.F.... there is not much utility in it. We have less reason to value it instrumentally than to value a corresponding true belief, since ordinary belief is linked appropriately with action. To be sure, some non-standard I.F.

> might be adequately connected with action. ...We need not restrict ourselves to valuing one category rather than another.[56]

It is important to note that when Goldman says "non-standard interpretation function", he means a function other than the one sanctioned by our commonsense intuition and not a *substandard* function as determined under Stichian interpretation function standards. Goldman is confident that a principle can be found in, what he calls, the notion of "the intimacy of action linkage" (linkage between beliefs about what is useful and our goals), which will allow us to adequately compare interpretation functions. It is interesting to note that Stich offers no response to this line of criticism in the March 1991 Symposium.

It would seem then that there are certain constraints to be placed upon the, as Stich describes it, indefinite number of alternative interpretation functions, of which there is purportedly no obvious reason to think would leave us any worse off. Stich has himself set "standards which any theory of interpretation aspires to meet". Presumably, should a particular interpretation function fail to meet these standards then, it is either inferior or possibly not an interpretation function at all. If we accept what Dretske and Goldman have to say on the matter, then in addition to these standards, an interpretation function must not produce irrelevant beliefs. For Goldman, the function must comply with his, "action linkage principle", where the function in question needs to produce beliefs which are intimately connected with the actions one takes in the pursuit of one's goals. A similar sort of constraint is

offered by Dretske, with the added feature that, in the case of direct observational beliefs, interpretation functions, which fail to explain the causal interactions in which the brain state is engaged (i.e.: direct perception), are irrelevant, and possibly even harmful. For Dretske, a good interpretation function and the beliefs it produces must establish truth conditions which are conditions that can be *tracked*, conditions about which the belief in question (under normal conditions) carries information.

Let's take a moment now to examine one of Stich's own examples of an alternative interpretation function, which is sanctioned by an alternative set of commonsense intuitions, in order to determine whether, as Stich claims, there is no obvious reason to think someone operating on this alternative would be any worse off than ourselves. He says that REFERENCE*** might be the result of a function which gave 'water' the EXTENSION*** H_2O and the famous stuff (why not the virtually unknown stuff?) that looks and tastes just like it, XYZ. Stich says:

> ...the "standard" causal/functional interpretation function would map the belief that I express by saying 'There is no water on the sun' to the proposition that there is no H_2O on the sun, while the interpretation function based on REFERENCE*** would map the belief to the proposition that there is no H_2O or XYZ on the sun. ...So while the interpretation function based on the intuitively sanctioned notion of reference might specify that a certain belief token of mine is true if and only if there is no H_2O on the sun, an interpretation function based on REFERENCE*** would specify that the same belief token is true (or, better, TRUE***) if and only if there is no H_2O and XYZ on the sun. [57]

Some elaboration would seem to be in order here. Generally speaking, the individual who operates on this interpretation function, the one that maps 'water'

onto H_2O and XYZ, is saying either that the natural kind term, 'water', refers to a class of objects of which H_2O and XYZ constitute a single member, or, that the natural kind term, 'water', refers to a class of objects of which H_2O and XYZ constitute two separate and distinct objects. This, it is assumed, is fairly obvious. For the record, Stich should be read as taking H_2O and XYZ to be separate and distinct objects, within a class of objects. In this way, 'water' is a natural kind term which refers to a class of objects in the world, of which H_2O and XYZ are separate and distinct members. When Stich says that 'water' has as its EXTENSION*** *the stuff* that looks and tastes just like H_2O, he should be read as implying that this *stuff* is not, strictly speaking, H_2O, but something that looks and tastes just like it. In this way, H_2O and XYZ are not ontologically identical - one and the same kind or type of object. The difference, perhaps, is at the molecular level - *heavy water* looks and tastes just like H_2O, but it isn't!

At any rate, let's proceed systematically and consider the first case. Would the person operating on the interpretation function which mapped 'water' onto (among other objects) one object which has two referential terms (in this case, H_2O and XYZ) be any worse off than myself, who operates on our commonsense intuition interpretation function? If everyone in the culture uses the same interpretation function, shares the same intuitions, and performs the same mappings (although there is nothing in Stich's position that requires any of this), the answer is

likely, no - this person would not be any worse off. So, when the chemist asks her assistant for 50 ml of water, the assistant can reach for the bottle marked H_2O or the one marked XYZ, it makes no difference. It is important to recognize here that if this is what Stich has in mind, then he has not produced an interpretation function which is in any significant way different from the one we now use. Our commonsense interpretation function involves many instances of objects and classes of objects being referred to by more than one referential term - e.g., Morning Star and Evening Star or dogs and canines. In this particular case, 'water' is simply being used to refer to a class of objects, one member of which is referred to by two referential terms XYZ and H_2O. Alternatively, the term 'stars' would refer to a class of objects, one member of which (Venus) is referred to by two referential terms, Morning Star and Evening Star.

Under the second case, would someone be any worse off, all things being equal (everyone in the culture shares the same interpretation function, intuitions and mappings), if 'water' had as its EXTENSION*** (among other objects) two separate and distinct objects within a class of objects? (Again, this is what Stich should be seen to be proposing.) The answer, it seems, is a little more complicated and the particulars make the difference. The particulars are concerned with matters involving communication, teaching, learning, and practical utility. The complication is compounded by the fact that the two objects or classes of objects

might break down in the following way: both objects (or classes as it were) could exist or one could exist and the other could be nonexistent.

Our commonsense intuitive function makes every effort possible to distinguish between objects and classes of objects which are significantly different. The significance of any difference is often, but not always, a matter of practical utility. (Stich would have us think that it is purely a matter of commonsense intuition when it comes to the causal/functional interpretation function). For instance, does it make any difference, from our practical viewpoint, whether all the white stuff that falls in the Arctic is referred to by the single referential term 'snow' or divided into fifty different referential terms, as the Inuit do? For us, there is no practical need for all of these divisions, which distinguish between the "grades" of Arctic snow. The Inuit, for reasons of practical utility, do require these distinctions. This *mapping* difference between us and the Inuit, however, does not mean that we operate on different interpretation functions which posit the existence of different objects or classes of objects which we consider counterintuitive. In fact, there is good reason to believe that we operate on the same function, but we employ it with greater scrutiny or more exactitude in different areas, as the circumstances demand. In this way, both the Inuit and I agree that snow exists: and I do not deny the existence of the 50 grades of snow which they distinguish - the distinction is not counterintuitive to either of us. This scenario, interestingly enough, is not something that Stich

considers. Surely, one interpretation function in the hands of different people will often produce mappings or make reference to objects which not every individual employing the function would necessarily consider - notice that "concede" was not used here - but which would also not be seen as counterintuitive.

A little closer to Stich's example, what's going on between the Inuit and us is much like what happens when 'water' gets mapped onto things like pure water, salt water, stagnant water, or even heavy water. Notice that there are adjectives placed in front of the word 'water' in each of these cases. The reason for this is that our commonsense interpretation function, as has been mentioned, concerns itself with significant differences and surely there is a significant difference between heavy water and plain old water - while they share many of the same physical and chemical properties (e.g., similar freezing and boiling points), water contains hydrogen atoms and generally can't kill you; heavy water, on the other hand, contains the hydrogen isotope, deuterium, and can prove very hazardous to your health. Imagine I had a glass of heavy water sitting in front of me and my girlfriend asked, "Is that water?", attempting to find out if I was taking a nip of vodka perhaps. What would my answer be? If I were to say yes, then, strictly speaking, I would not be lying. On the other hand, our intuition tells us that I have not exactly told the whole referential story here. If she were to then ask for a drink of the water, a complete and more accurate answer would be immediately forthcoming! Notice that Stich's XYZ is

described as looking and tasting just like H_2O. I am told that heavy water could be similarly described. If, when talking about his alternative interpretation function, 'water' is to be mapped onto H_2O and XYZ, and XYZ is construed along the lines of , say, heavy water or pure water or dirty water, then no serious objections need be registered; other than to point out that under some circumstances the use of the referential term water to *loosely* refer to objects such as heavy water, salt water and the like could have undesirable consequences (e.g., my girlfriend unknowingly asking for a drink of my *heavy water*). In other words, the term 'XYZ' would have to be communally recognized and distinguished by the users as fitting into the same sort of category as stagnant, heavy or sterile water. The reason there are no serious objections to be made here is that Stich has, once again, not actually produced an alternative interpretation function. Our commonsense, intuitive function already performs exactly this task - as illustrated by the Inuit and heavy water examples. So, when the chemist asks for 50ml of water, the assistant knows she means H_2O (provided they share a common language, set of intuitions, etc.), otherwise the chemist would have requested XYZ specifically, if that were what she desired. Once again, this is because 'water' only *loosely* refers to heavy water, stagnant water, salt water, etc.; as it would also only *loosely* refer to XYZ.

There is another possibility here. Water freezes, it turns to slush or steam. Ice, slush and steam are all different referential terms which are employed to

facilitate practical distinctions between the various forms or states in which water can be found. Under our intuitive interpretation function, however, no one maps the word 'water', even *loosely*, onto ice or steam or slush. For instance, no one who asks for scotch and water expects to see ice cubes in their drink when they receive it. On the other hand, no one operating on our intuitive function would suggest that ice, steam and slush are not simply forms of water - after all, ice *is* frozen water. So, Stich could mean by his alternative interpretation function that 'water' is mapped onto H_2O and XYZ, and XYZ is like ice or steam in this respect. Again, no serious objections would be registered. The chemist who asks for 50 ml of water does not expect the assistant to bring her 50 ml of ice or steam - she wants liquid H_2O. And if the chemist and the assistant share the same language, intuition, etc., then the assistant would know to furnish the chemist with liquid H_2O - much as the bartender would know to put water in the scotch, not ice cubes. So once again, there is nothing new being introduced here. Our commonsense interpretation function regularly makes these mapping distinctions. At any rate, it seems unlikely that Stich means for us to construe his XYZ along the same lines as ice or steam; for otherwise, he would not have given it the properties of *looking* and *tasting* just like H_2O.

In the possibilities examined thus far, it would seem that the alleged alternative function, which maps 'water' onto H_2O and XYZ, is, in point of fact, not a distinct alternative at all, but a function which produces mappings our

commonsense interpretation function would happily sanction were it presented with the opportunity. But now, if the alternative interpretation function involved a mapping which saw 'water' referring to H_2O and XYZ, not in the senses talked about above (i.e., heavy water and ice), but where 'water' refers to the separate and distinct objects H_2O and XYZ where these objects were themselves not discriminated in a fashion similar to that found in the case of heavy water and ice, we would certainly run into difficulties. In such a case, objections can be lodged. Under our interpretation function, as has been said, we have many instances where two or more referential terms refer to one object (or classes of objects, as it were) - notice, Morning Star and Evening Star. But we do not have, or at least we try to avoid, cases of one referential term referring, without means of discrimination, to two or more distinct and separate objects or classes of objects (natural kind terms being the exception). The reason for this is clear and has been hinted at already. For instance, if the chemist asks her assistant for 50 ml of water, what is the assistant to do - bring H_2O or XYZ? They are, after all, separate and distinct objects which are similarly referred to by the term 'water', but which themselves are not then distinguished in the language. In other words, the referential term 'water' refers to both H_2O and XYZ as though they were the same thing, as though there was no distinction or difference between them, so the assistant has no way of knowing for sure which substance the chemist wants. But surely the difference here might make all the difference in the

world! In this way, were Stich's alternative interpretation function to apply across the board and map single referential terms or words onto 3, 4, 20 or even 200 separate and distinct objects or classes of objects without means of further distinction between the objects, communication would be made considerably more difficult, if not impossible. Even the telephone company knows the communications problem which would ensue were they to assign one telephone number to more than one separate and distinct residence or business.

Once communication is thwarted, or prevented, the practical utility of this interpretation function and its mappings is made dubious. Imagine instead that this function mapped 'water' onto 6 separate and distinct objects, which themselves were not discriminated in the language. The chemist would presumably (but not necessarily under Stich's account) know what she wanted, but the assistant would only be picking up objects randomly - "No, not that water the *other* water. No! The *other* water!", shouts the chemist. How about teaching someone a language that employed this sort of interpretation function, where a word is mapped onto more than one separate and distinct object or class of objects with no means to distinguish between the referents. You might begin, as the teaching of languages often does, by pointing to one object, say H_2O, and uttering, 'water'. Then you would point to another separate and distinct object, perhaps 7-up soda pop, and say, 'water'. And then another separate and distinct object, for instance white maple syrup, and say

'water': and another and another... Incidentally, there is nothing in Stich's explanation of alternative interpretation functions that precludes these other objects from not "looking and tasting" anything like the first. By this it is meant that another alternative function might map 'water' onto the infamous stuff, ABC, that doesn't look or taste anything like H_2O. In each of these cases, the person learning the language is told that the term 'water' equally refers to all of these substances (or objects) without any further means of distinguishing the substances. At minimum, the word 'water' would cease to have any (useful) meaning for this person and for the community.

The last possibility to consider is that under this alternative interpretation function the word 'water' might refer to two separate and distinct objects, one of which existed, the other of which did not - say, H_2O exists and XYZ is nonexistent. Our intuitive interpretation function allows for referential terms or words to refer to nonexistent objects - 'unicorn', for instance. However, our interpretation function does not allow for the mapping of a word onto both an existent and a nonexistent object or class of objects. Would someone whose interpretation function allowed for this sort of mapping be any worse off than someone who utilized our commonsense function? Well, a unicorn can be described in considerable detail. A representation of one can be tattooed on my arm. Presumably then, the nonexistent object, XYZ, could be similarly described. In fact,

Stich begins the description for us, saying that it is famous and looks and tastes just like H_2O. But if the same chemist were to request 50ml of water, only to be furnished with an empty beaker (allegedly containing XYZ), then it's likely her assistant would finally be replaced.

From this analysis of the H_2O/XYZ example, it should be clear that Stich's undefended claim that there is "no obvious reason to think that people whose intuitions diverged from ours in these ways would be any worse off" is in need of considerable defense and qualification. This is not to say that there are not people with intuitions which are different from ours or that some of those alternative intuitions might not leave them any worse off. It is just to say that it is anything but *obvious*. From the preceding analysis, it should also be clear that these alternative intuitions might not be in strict opposition to our commonsense intuition. In this way, these alternatives might result in mappings we have not actually made, but not in mappings we would be unable to endorse under our present-day commonsense intuition. As such, these alternatives would, so to speak, come under our commonsense intuition umbrella, in that they would ultimately not be construed as counterintuitive. There may also be people whose intuitions and resultant mappings do diverge from ours, and would not be endorsed by our commonsense interpretation function. In these cases, the preceding analysis was designed to demonstrate how

divergent mappings and interpretation functions might in fact leave someone worse off.

This discussion on the nature of our commonsense intuition and Stich's understanding of it should help clarify the parameters of his notion of competing interpretation functions (including the theories of reference and truth associated with them) and the beliefs they produce - a notion that permeates the remainder of this thesis.

By way of introduction to the next section, Stich's closing comments regarding the import of his arguments for the partiality and idiosyncrasy of the causal/functional interpretation function should be addressed.

> One consequence of all this is that when it comes to deciding what we really value in our doxastic states and in the processes that generate them, truth has lots of competition. Any given set of belief tokens that I might have will contain a certain percentage, say n, of true beliefs. The but same 'sic' set will also contain a certain percentage, n^*, of TRUE* beliefs, a certain percentage, n^{**}, of TRUE** beliefs, and so on for indefinitely many variations on the intuitively favored semantic theme. Moreover, in general $n \neq n^* \neq n^{**}$....So it will often be the case that when we increase our percentage of true beliefs, we will decrease our percentage of TRUE* (and/or TRUE**, and/or TRUE***...) beliefs. If we really value true beliefs, presumably we won't much care about giving up TRUTH* or TRUTH**. [58]

But what exactly is the nature of truth's competition? The answer involves discussion of the essence of the distinction between, as Stich says, "plain old truth" and TRUTH**...*, true beliefs and TRUE**...* beliefs, and truth conditions and TRUTH CONDITIONS**...*. It will be argued that this particular method of

distinction (i.e., the use of capitalization and asterisks) and the reasoning upon which it is based causes Stich to falsely conclude that alternative interpretation functions will result in alternative, competing concepts of truth - or as he says, TRUTH**...*. Ultimately, it will be argued that Stich has created the illusion of competing *kinds* or concepts of truth and this calls into question his argumentation against the (instrumental or intrinsic) value of true belief, which is predicated upon there being competing notions or concepts of truth.

Let's start with the simplest sentence - 'That beaker has 50ml of water in it'. Now, strictly speaking, this arrangement of symbols, referred to as a sentence, has no meaning. Once the sentence has been given meaning, however, it becomes a statement or a proposition. And giving such a sentence meaning involves employing concepts such as reference, truth conditions, and logical format or relations - in other words, an interpretation function. In turn, once these concepts are applied to a sentence, we can then sensibly speak of the truth value of the statement. This, of course, sounds very much like Stich's explanation of how a brain state token, by way of an interpretation function, can acquire the semantic properties of meaning and a truth value. Thus, from what has been said so far, it should be clear that interpretation functions and meaning go hand-in-hand. That is, once you apply an interpretation function to a sentence, or brain state token as it were, you give it a particular meaning. It follows from this that where meaning has been provided, so

too have truth conditions and ultimately truth values. To use Stich's example of the term 'water', if there are two distinct interpretation functions being applied to the sentence, 'That beaker has 50ml of water in it', then the statements which are produced will have two different meanings. For instance, interpretation function *A* might give the term 'water' the extension H_2O and interpretation function *B* might have H_2O and XYZ as the extension (or in Stichian terms, the EXTENSION*) of the term 'water'. This, of course, will also result in different truth conditions for various statements which employ the term 'water'. So, the statement, "That beaker has 50ml of water in it," will be true under function *A* if there is only H_2O to be found in the beaker - this could be seen as the product of our causal/function interpretation function. Function *B*, on the other hand, will set TRUTH CONDITIONS* for the statement such that both H_2O and XYZ must be in the beaker for the statement to be TRUE*. Notice the use of caps and asterisks here. Stich emphasizes the distinction between the components of alternative interpretation functions (i.e., the theory of reference used, the truth conditions which are established, etc.) and those of the causal/functional interpretation function by capitalizing and placing asterisks after those components which are generated by the alternatives. In this way, Stich delineates between true beliefs, truth conditions and truth, as affiliated with the causal/functional interpretation function: and TRUE**...* beliefs, TRUTH CONDITIONS**...*, and TRUTH**...*, as

associated with some alternative function which is not sanctioned by our present-day commonsense intuition (e.g., syntactic interpretation functions, causal/description-cluster functions, the function that maps 'water' onto H_2O and XYZ, etc.).

It is clear then that different interpretation functions applied to the same sentence (or as Stich sometimes says, well-formed formula) will produce different statements, in that the statements will have different meanings. So, while statements having different meaning can share the same sentence or symbolic format, they are certainly not the same statements precisely because they have different meanings. And in as much as they have different meanings they also have different truth conditions associated with them. This much, both Stich and I can agree upon. We should now turn our attention to the notion of truth values.

The general approach to determining the truth value of a statement - irrespective of which interpretation function was employed in determining its meaning - involves checking to see if the truth conditions - be they truth conditions or TRUTH CONDITIONS**...* - have been met. Depending on whether they have been met or not, an appropriate truth value is assigned. This is nothing new and nothing peculiar to the casual/functional interpretation function. No matter what the interpretation function, the truth value of a statement is established in this manner. This process, you will notice, has intentionally been described in rather

vague terms, avoiding any reference to words like true or false. This is done to prevent any potential conflict with Stich's alternative *kinds* or concepts of truth - TRUTH**...*. Given that Stich has not openly addressed or contested this feature of truth determination and given that his talk of various truth conditions and theories of reference only makes sense if this understanding of truth determination is at least tacitly endorsed, it is assumed that what has been said thus far is something Stich would have to accept.

So far then, Stich's chain from sentence to TRUE**...* belief and TRUTH**...* goes something like this: An alternative interpretation function is applied to a sentence. The sentence then becomes a statement of belief, having a particular meaning. Employing its alternative theory of REFERENCE**, the interpretation function establishes TRUTH CONDITIONS** for the statement of belief. Should the TRUTH CONDITIONS** be met, then the belief is deemed a TRUE** belief. Stich then concludes that the resultant TRUE** belief and the notion of TRUTH** associated with it are viable competitors for plain old true beliefs and truth, thereby making it anything *but* obvious that the latter are (intrinsically or instrumentally) valuable. [Note: the details of how we determine whether the truth conditions have been met is not of interest to Stich here. Since this issue is not important to his arguments, he concedes that there is some unproblematic way to make this determination. See pp.104-105 of *The Fragmentation of Reason.*]

Allow me to tie these first points together and at the same time encourage their continuous recollection throughout what follows. Thus far, two widely acknowledged philosophical points have been stated: i) A single sentence which is given different meanings would produce statements having different meanings and statements having different meanings will have different truth conditions; and ii) The truth value of a statement is determined by examining the truth conditions of the statement to see if they obtain. Both these points are not in contention. They are considered valid irrespective of the particular interpretation function under discussion.

Because Stich is rather vague about what he takes 'truth' to mean, we need to more closely examine the meaning of two terms found in Stich's lexicon: truth and truth conditions or, as it were, TRUTH**...* and TRUTH CONDITIONS**...*.

Alvin Goldman approaches the meaning of TRUTH**...* and TRUTH CONDITIONS**...* on an ordinary, yet direct, critical path. He wonders whether TRUTH**...* are really rivals of truth. He takes TRUTH**...* to be complexes of interpretation functions and truth values, making them something quite different from truth or falsity, as they appear in truth values.

> Let me put the point slightly differently, focusing on what the ordinary person values in valuing true belief. The ordinary person, who is innocent of philosophy, simply assumes that when she entertains a thought, this picks out a unique proposition, say P. She then has a preference ordering over states of affairs

> involving P, each state consisting of a possible doxastic choice vis-à-vis P and a possible truth value of P. For example, she prefers (believe P and P is true) to (be agnostic about P and P is true) to (believe P and P is false). There is no comparison of believing P and P is true with believing P where P is TRUE*. Since TRUE* isn't a truth value, this wouldn't even make sense.
>
> Stich tends to conflate the question of what I.F. we accept with the question of what truth value we prefer (in our beliefs). In one place, for example, he says that to accept the I.F. that our culture (or biology) bequeaths to us is to let that function determine our basic epistemic value (p.120). But, surely, acceptance of the ordinary I.F. carries no commitment to a preference for truth over error. We might accept the ordinary I.F. but invert the ranking given earlier.
>
> However, I think that the point Stich really wants to make can be reformulated to avoid these objections. Instead of talking about TRUE* and TRUE**, I think he should be talking about BELIEF* and BELIEF**. His real challenge is not why we should value *true* belief, but why we should value true *belief*, as opposed, say to true BELIEF*. Once we learn about alternative I.F.'s, why should we value being in brain states that turn out true on the ordinary I.F. (true beliefs) rather than brain states that turn out true on some nonstandard I.F. (true BELIEFS*)?[59]

These are powerful passages that go to the heart of the terms under discussions. (Incidentally, Stich does not respond to this reading of his work in the March 1991 Symposium on *The Fragmentation of Reason.*) If you will permit some paraphrasing, Goldman seems to be saying that TRUTH**...* are not genuine rivals of truth. Instead, any talk about TRUTH**...* should be replaced with BELIEF**...*, because Stich is actually providing alternatives not for truth but for interpretation functions, which, in turn, are the mechanisms by which mental states get mapped onto propositions or states of affairs. According to Goldman then, there is a difference, which Stich fails to recognize, between the epistemic value of truth

over error and the interpretation function which is employed in the formation of beliefs. In short, to operate on an interpretation function other than our commonsense function is not tantamount to a denial of the epistemic value of truth, just the mappings or beliefs this function would produce. This is, of course, consistent with what has been said thus far. Different interpretation functions produce different statements - or beliefs, as it were - which are assigned appropriate truth values when their truth conditions are satisfied. In other words, as Goldman might say, if there is anything being altered here, it is the statement of belief, not the truth of it.

With Goldman's analysis in mind, we can now examine each link of Stich's chain, from sentence or brain state token to truth conditions (or TRUTH CONDITIONS**...*) and truth (or TRUTH**...*). In so doing, the goal will be to buttress Goldman's analysis of TRUTH**...* and TRUTH CONDITIONS**...* and demonstrate that Stich has not created genuine competition for truth, but merely the appearance of competition.

The first step involves applying two different interpretation functions to the sentence, 'That beaker has 50ml of water in it'. The first interpretation function will be referred to as IF_1 and the second IF_2. IF_1 will represent our causal/functional interpretation function, which operates on a theory of reference that gives 'water' the extension H_2O. IF_2 is an alternative interpretation function

employing an alternative theory of REFERENCE* which includes H_2O and XYZ in the EXTENSION* of 'water'. It should be clear from what has been said thus far that the statement, "That beaker has 50ml of water in it," as determined under IF_1, has a meaning that is different from the statement, "That beaker has 50ml of water in it," as determined under IF_2. It follows then that because they are two different statements they also have different truth conditions. It is here that Stich begins the illusion. He claims that IF_1 will set truth conditions for its statement while IF_2 will instead set TRUTH CONDITIONS*. But surely the use of capitalization and asterisks here is misleading and superfluous. There is nothing unusual and certainly nothing competitive about the truth conditions set under IF_2 that would warrant them being referred to in caps and with asterisks. There is no competition or comparison being made between the truth conditions of the statement "That beaker has 50ml of water in it," as determined under IF_1 and truth conditions of the statement, "That beaker has 50ml of water in it," as determined under IF_2. These are two entirely different statements having two different meanings. We might as well compare, as competitors, the truth conditions of the statement, "Bill Clinton has been re-elected," with those of the statement, "Peru is in South America." And, it should be added, there is nothing to save Stich's distinction (i.e., truth conditions vs. TRUTH CONDITIONS*) in the fact that both statements are derived from the same sentence - sentences have no meaning and can therefore not be expected to

generate competing notions of truth. It would seem thus far then that the root of competition is not to be found in the sentence, the statements, or the truth conditions. Attention will now be given to truth values and truth itself.

To facilitate further discussion, an additional feature must be added to the "That beaker has 50ml of water in it," statements. They should now be seen as statements of belief. In the case of IF_1, the belief statement, "That beaker has 50ml of water in it," would have a truth value of true if the truth conditions for the belief were met - there was indeed 50ml of water in the beaker. Under IF_2, the belief statement, "That beaker has 50ml of water in it," would be, as Stich insists, assigned a truth value of TRUE* if the TRUTH CONDITIONS* for the belief were met - the beaker in fact contained 50ml of H_2O and XYZ. It should be noted that neither asterisks nor capitalization are employed where the term, 'truth values' is used. Stich similarly fails to employ these devices when he uses the term. This is another indication that he endorses the general approach outlined above of how truth values are determined, regardless of the interpretation function involved. By this it is meant that he apparently does not see the need to distinguish the process of assigning truth values by using capitalization and asterisks in cases where different interpretation functions are involved. This having been said, it seems odd that, under IF_2, Stich would then feel the need to refer to beliefs assigned a truth value of true as TRUE* beliefs, rather than simply true beliefs. Surely the statement, "That beaker has 50ml

of water in it," even when determined under IF_2, is nothing more than a true belief, particularly in light of the fact that it is not even the same belief as that which is stated under IF_1. The upshot of this is that there can be no legitimate reason to distinguish, by way of capitalization and asterisks, between true/false beliefs and TRUE/FALSE**...* beliefs; at least, in the sense that they ought to be competitively compared.

The final link in Stich's chain, truth, will now be considered. Admittedly, it does seem natural under Stich's capitalization and asterisk scheme, to assume that IF_2 would involve an alternative, competing *kind* or notion of truth, namely TRUTH* (e.g., the belief 'P' is TRUE* because having it increases one's happiness). It is worth noting, however, that Stich does not devote a single line in his book to theories of truth. His failure to do so leaves one wondering what he takes truth to be - a matter which at least involves distinguishing between truth as a truth value and what truth, as a meta-concept or value, is. Instead, he draws us a picture of a complex array of radii representing alternative interpretation functions which emanate from a single sentence and which, without analysis or explanation, terminate at different theories or notions of truth (i.e., TRUTH**...*). The assumption that each alternative interpretation function will produce different notions of truth, combined with the use of capitalization and asterisks, creates the appearance of competition for truth. However, precisely because the truth value of a statement is

determined in the manner stipulated earlier, regardless of the interpretation function or theory of reference or truth conditions involved, it is never the case that different interpretation functions will produce a choice between competitors for truth. If the truth conditions for a particular meaning of a statement are met then the statement is simply true, and this is so whether the interpretation function involved was causal/functional or some alternative. This is made all the more obvious when we recognize that applying different interpretation functions to a sentence produces different statements or beliefs, having different meanings which, in turn, produces different truth conditions.

This much having been said, it appears that the only area where comparison and competition might be possible is in the reference theories endorsed by varying interpretation functions. If this is true, then the title of Stich's Chapter Five - "Do We Really Care Whether Our Beliefs Are True?" - would be more aptly entitled, "Do We Really Care Whether Our Beliefs Are Based on the Causal Theory of Reference ?". This is a title Goldman might applaud. But perhaps this too would be an illusion. For, as will be argued, if this is where Stich believes competitors for truth are found, then he would be confusing distinct elements: i) the descriptive nature of theories of reference with a prescriptive nature: and ii) truth as a meta-concept or value with true as a truth value.

All theories of reference are designed to describe what it is we do when we refer, not to recommend how we should refer. If this is true, then there is no question whether we *care about* which theory we use. Instead, we simply successfully or unsuccessfully refer, with some of the more important features of this process having been alluded to earlier in this section of the thesis (pp.38-57). The only sensible comparison or competition possible between theories of reference is in how accurately they capture what it is we do when we refer.

As to the notion of a meta-meaning or value of truth, we can now add something to Goldman's remarks. Our ordinary concept of truth involves at least these two related features: i) The beliefs we have are seen to be, ultimately, either true or false - there is no category which says they can be, ultimately, neither true nor false, and ii) We believe the actual state of the world is as it is regardless of which interpretation function we use to decipher or reveal that state - we merely discover the state. The first feature says that our beliefs either do (true) or do not (false) capture the actual state of the word - saying, if you will, something like, "This is how the world actually is." The second feature is about the state of the world, independent of anything we say (or believe) about it. Any concept of truth that does not have these two components, in conjunction with Goldman's notion of preferring truth over error, is significantly different from the concept used in the interpretation functions sanctioned by our commonsense intuition. Talk about truth in this sense

is, however, something quite different from the assignment of truth values, which is where Stich's emphasis appears to lie. In this sense then, truth is something independent of the assignment of truth values or the theory of reference used in determining the meaning and truth conditions of a statement (of belief).

In the end, should Stich be advocating a meta-concept of truth that is different from ours, denying one or more features of our notion of truth, then he owes us an argument to that effect. It is one thing for him to say that there are or might be different ways to go about the business of assigning meaning and truth conditions to sentences thereby creating different statements (of belief) and quite another to say that we ought to drop or reconsider our epistemic priority to think accurately about the state of the world, which includes formulating and maintaining beliefs (or, as Goldman might say, BELIEFS**...*) which capture the actual state of affairs in the world or are true in the meta-sense of the word.

With this analysis of Stich's notion of there being competition for truth and true beliefs in mind, we now proceed to his treatment of the intrinsic and instrumental value of true belief.

Section 4

STAGE 2: DO WE REALLY CARE WHETHER OUR BELIEFS ARE TRUE?

Stich now turns his attention to the issue of whether the value of true beliefs is obvious (whether we do or should value true belief). In particular, he examines the notions of instrumental and intrinsic epistemic value in true beliefs. Stich begins by providing cursory definitions for both:

> ...intrinsic value, the sort of value something has for itself, and instrumental value, the sort that something has in virtue of leading to something else.[60]

INTRINSIC VALUE OF TRUE BELIEFS

Stich contends that, in spite of the nature of intrinsic valuation, a fairly persuasive argument can be formulated which will at least bring into question the intrinsic value of true belief. The argument is as follows:

> ...there is [a] consideration that might be efficacious in persuading someone that she should not, or does not really, accord intrinsic value to the having of true beliefs. Rather than dwelling on the consequence of having true beliefs, we can try to be sure that she sees clearly the real nature of what she values - that she appreciates what having true beliefs comes to. ...if the function pairing beliefs with their truth conditions is the one outlined by the causal/functional theory, then it is both partial and idiosyncratic. And, in rather different ways, both of these facts entail that valuing true beliefs is a profoundly conservative thing to do.[61]

With respect to the partiality of the causal/functional interpretation function and the conservative nature of valuing true beliefs Stich says:

> Consider first the fact that the interpretation function has a very limited domain. What this means is that there is a huge space of possible systems of mental computation and storage whose component states have no truth conditions and thus cannot be true. ...in this huge space there are systems that would vastly increase their user's power or happiness or biological fitness, systems that would lead to substantial reductions in the amount of suffering in the universe, and systems that would significantly reduce the probability that we will bomb ourselves into oblivion along with much of the biosphere. ...But almost all of that space is beyond the reach of the causal/functional interpretation function; it is a domain in which there is neither truth nor falsity. Those who would accord intrinsic value to the holding of true beliefs may well be reluctant to explore that vast space and will resist adopting what may be found, since we know in advance that it contains no true beliefs. But theirs is a profoundly conservative normative stand. For what they value in the end products of cognition must be semantically interpretable, and what is semantically interpretable cannot depart too radically from current patterns of reasoning or from familiar ways of causally tying mental states to extramental reality. To value true belief is to resolve that in matters cognitive, one will not venture very far from where we are now.[62]

Having revealed the conservative nature of intrinsically valuing true beliefs with respect to the partiality of the causal/functional theory, Stich then demonstrates this conservative stance with respect to the causal/functional theory's idiosyncratic vein:

> The conservatism entailed by the idiosyncratic nature of the interpretation function is of a rather different kind. ...if...it is the causal/functional interpretation function that is sanctioned by intuition, then it is not a particularly simple or natural function. Rather, it is something of a hodgepodge, built from a more or less heterogeneous family of strategies for fixing the reference of terms and another family of strategies for transmitting reference from one speaker to another. What distinguishes acceptable groundings and transmissions is not that they share some common natural property but simply that they are found to be acceptable by commonsense intuition.[63]

Stich has one more nail for the conservative coffin of intrinsically valued true belief, namely the origins and character of our commonsense intuition itself:

> Why do we have these particular intuitions rather than those that would sanction REFERENCE*, REFERENCE**, or one of the others? The short answer, of course, is that no one knows in any detail just how these intuitions arise. But it's a good bet that, like other complex systems of intuitions such as those concerning grammaticality or morality or politeness, the intuitions in question are themselves culturally transmitted and acquired by individuals from the surrounding society with little or no explicit instruction... Whatever the explanation, it is clear that our intuitions do not result from a systematic and critical assessment of the many alternative interpretation functions and the various virtues that each may have. One way or another, we have simply inherited our intuitions; we have not made a reflective choice to have them. Those who find intrinsic value in having true beliefs (rather than TRUE* ones, or TRUE** ones....) are accepting unreflectively the interpretation function that our culture (or our biology) has bequeathed to us and letting that function determine their basic epistemic value. In so doing, they are making a profoundly conservative choice: they are letting tradition determine their cognitive values without any attempt at critical evaluation of that tradition.[64]

Stich recognizes that it is one thing to nail the coffin shut and quite another to get it buried. Nothing he has said thus far, Stich concedes, is a "knockdown argument against according intrinsic value to having true beliefs".

> However, there are many people, and I am among them, who are not much inclined to value what is traditional and familiar for its own sake in matters epistemic. And for them the fact that true beliefs must be within the domain of the interpretation function, the fact that the domain of the interpretation function is limited to systems of cognitive storage similar to our own, and the realization that the function is an idiosyncratic hodgepodge bequeathed to us by our cultural and/or biological heritage may well be reason enough to decide that they do not really value true

> beliefs intrinsically. For those whose reflective epistemic values run along these lines, true beliefs may still turn out to be valuable, but their value will be instrumental. They will have to be good for something.[65]

We may now begin by examining what Stich means by: i) "there is a huge space of possible systems of mental computation and storage whose component states have no truth conditions and thus cannot be true...it is a domain in which there is neither truth nor falsity...[and] it [the domain or huge space] contains no true beliefs": and ii) "what they [those who intrinsically value true beliefs] value in the end products of cognition must be semantically interpretable." It can be argued that, as alternative interpretation functions, at least some of these systems, by Stich's own admission, fail to meet his standards of what an interpretation function is supposed to do. This huge space of possible systems of mental computation and storage may also leave anyone actually operating on them, as Stich would say, worse off. These alternative systems will impede or prevent communication, teaching and learning, much as we saw might occur in Stich's example of 'water' having the EXTENSION*** H_20 and XYZ (pp. 47-57 of this thesis).

In a footnote, Stich professes to owe much of his argument for the idiosyncrasy of the causal/functional interpretation function to Peter Godfrey-Smith, and particularly, his article entitled, "Why Semantic Properties Won't Earn Their Keep"(1986). It seems prudent then, to examine the main

conclusions of this paper, in the hope that they will shed some light on how the aforementioned passages are to be read - passages which are used by Stich to establish the epistemic conservatism of intrinsically valuing true beliefs. Godfrey-Smith is concerned with the question of whether commonsense, intuitive notions of intentional content and truth can be maintained in the face of a "more rigorous and advanced science of the mind: whether the semantic notions invoked in folk psychology will play any role in serious cognitive psychology."[66] Ultimately, both Stich and Godfrey-Smith, maintain that syntactic theories likely do a better job than semantic theories of explaining and predicting behavior.

> Contemporary psychology sees cognition as computation - as a formal, rule-governed process. Computation takes place within a computational system, perhaps even an 'inner code'. But to grant all this is not to grant that this system need have semantic properties of any full-blooded (truth-conditional, referential) kind. Indeed, how *could* mazes of extra-cranial causal links - which are what full-blooded semantic properties come down to - be of much interest to the computationally minded psychologist?[67]

It would seem then that the "huge space of possible systems of mental computation and storage" which Stich refers to in the above passages are actually syntactic systems. And these syntactic systems are being offered by Stich as alternatives to our commonsense intuitive interpretation function - an offering which he claims demonstrates the profoundly conservative nature of intrinsically valuing true beliefs. In this way, we can now make some sense of his comments that there would be neither truth nor falsity in the domain of these alternative interpretation

functions. Godfrey-Smith examines several arguments in favor of semantic theories and the question of whether they can be replaced by syntactic theories. One of these arguments involves a woman, Mary, who runs out of a building upon discovering (via inhaling smoke) that it is on fire. The author of this example, Zenon Pylyshyn (1980, p.161), says that a syntactical account of her behavior is inadequate because:

> It does not show how or why this [building-leaving] behavior is related to very similar behavior she would exhibit as a consequence of receiving a phone call in which she heard the utterance 'the building is on fire', as a consequence of her hearing the fire alarm, or...(etc., etc.).

The upshot of this example is that communication (along with teaching and learning) would be very difficult, if not impossible, were it not for semantic interpretation functions: and, in turn, semantic explanations and predictions of behavior. Godfrey-Smith claims this argument does not succeed. He assures us that all the semantic theory really says is this: "Mary, or whoever, *interprets* all the stimuli [e.g., the phone call, the alarm, etc.] in the same way, and what this 'interpretation' comes down to is just the interaction of beliefs, and the fact that the physically different stimuli all lead to the belief that the building is on fire."[68] Both he and Stich claim that an equally plausible and complete account can be provided by a syntactic theory - "all the stimuli lead to the same syntactically individuated state, which interacts with other states to produce the same building-exiting behavior."[69]

Godfrey-Smith and Stich claim communication (teaching and learning) does not necessarily require semantic interpretation, saying:

> Crucially, there is no other interesting property the semantic story can point to in the range of stimuli; anything Mary *does* 'interpret' in a certain way will lead her to leave the building, and anything she *doesn't* interpret in that way won't. ...After all, how could the fine details of the *etiologies* of the stimuli, and of Mary's past links to her environment, be playing an interesting role here? At best there could only be a rough correlation between these semantically relevant external properties of Mary and the stimuli, and the causally interesting properties, which involve what Mary *does* with the inputs.[70]

Stich, in an earlier publication having deconstructionist goals similar to those found in *The Fragmentation of Reason* and Godfrey-Smith's, "Why Semantics Won't Earn Their Keep", argues for a syntactic theory of mind and, in turn, syntactic explanations and predictions of behavior - behavior like Mary's. With respect to syntactic theories replacing semantic theories in the business of behavior prediction and explanation, he argues much the same as Godfrey-Smith, but with one very noteworthy difference.

> ...we should not expect a psychological theory to predict or explain behavior under any and every description countenanced by common sense. To see this clearly, an analogy with chemistry is useful. It may be quite true that boiling a bottle of Chateau Lafitte causes a substantial reduction in its market value. But this is nothing that we expect chemistry alone to explain. Rather, we expect chemistry to explain the effects of boiling described in an appropriately delimited, proprietary chemical vocabulary. Moreover, there is not likely to be any antecedently obvious specification of the range of descriptions appropriate in chemical explanada. Elaborating or delimiting the language in which explanada are to be described is an aspect and often quite a fundamental aspect of theory construction in science. To explain

> why boiling causes a decline in the market value of Chateau Lafitte we will have to supplement the chemical explanation of the effects of boiling with facts about the way chemical changes affect the sensory qualities of a wine and facts about the relation between sensory qualities and the market value of rare Bordeaux wines. The situation is similar in psychology. We cannot expect that a scientific psychology will explain behavioral events under all imaginable descriptions. Rather the psychologist must select or formulate an appropriate descriptive language for his explanada. And the formulation of such a vocabulary will be a fundamental part of psychological theory construction.[71]

As this feature of Stich's argument for syntactic explanations and predictions of behavior plays such a major role in his argument for alternative syntactic interpretation functions, a second analogy he employs should be examined: one in which he characterizes the syntactic language used to describe behavior as, *autonomous behavioral description.* As Stich puts it: "if it applies to an organism in a given setting, then it would also apply to a replica of the organism in that setting." [72] The analogy involves "robot psychology." Stich assumes that theories explaining the behavior of certain complex robots is likely analogous to those which explain the behavior of organisms. The question is then: Would the robot theories be expected to furnish explanations of non-autonomous descriptions of behavior? He says that a robot's behavior can be described in many ways using non-autonomous terms, as opposed to autonomous ones.

> For example, a given robot on the production line at General Motors might, on a certain occasion, successfully perform its millionth weld. Although 'performing its millionth weld' might be a correct description of what the robot does, it is clearly not an autonomous description. If, just prior to performing the weld,

> the robot in question had been replaced by a brand new replica robot, the replica would have performed a qualitatively identical weld. But it would have performed its first weld, not its millionth. In performing a weld, a robot might also be falsifying Professor Hobert's prediction that no robot would ever perform a million welds, and simultaneously fulfilling a provision in the contract between General Motors and the robot's manufacturer. But again, neither of these descriptions of the robot's behavior is autonomous. It seems obvious that if we seek systematic generalizations to explain the robot's behavior, we should not expect our generalizations to explain the robot's behavior under *these* descriptions. The descriptions under which we expect a theory of robot behavior to explain that behavior are autonomous descriptions.[73]

Stich is quick to point out that there is nothing mysterious about non-autonomous descriptions (e.g., the robot's millionth weld). However, he says they are best seen as logical or conceptual hybrids constructed from the autonomous and the non-autonomous descriptions.

> To successfully perform its one millionth weld, a device must successfully perform a weld *and* it must have previously performed 999,999 other welds. The first element in this conjunct describes the behavior autonomously; it is just the sort of fact that we expect a theory of robot behavior to explain. The second element is a historical fact, and it is not at all what we expect a theory of robot behavior to explain.[74]

He concludes:

> The Syntactic theory of Mind...requires purely formal generalizations which ignore those historical and environmental factors that may distinguish an organism from its replica in the eyes of folk psychology. If the argument for the autonomy principle is persuasive, then the STM strategy is the one to be preferred.[75]

Given space constraints and the immediate focus of this thesis, being drawn into a prolonged discussion concerning the merits and demerits of a syntactic theory's ability to explain and predict behavior is to be avoided. However, certain comments must be made with respect to how these passages affect Stich's project in Chapter Five of *The Fragmentation of Reason*, and particularly the comments we have been discussing regarding alternative syntactic interpretation functions. To alleviate any potential confusion surrounding the attention that will be paid to Stich's Syntactic Theory of Mind, it should be said that it is only of interest insofar as it is a syntactic theory of mind and consequently endorses syntactic interpretation functions. We are forced to examine this alternative text - *From Folk Psychology to Cognitive Science* - because Stich is particularly vague in his discussions surrounding the so called, alternative interpretation functions which can be found in the huge space of possible systems of mental computation and storage which contain neither truth nor falsity. Since both Godfrey-Smith and Stich rely on *From Folk Psychology to Cognitive Science* in their respective arguments for syntactic theories of behavioral explanation and prediction, and both provide the cross-references between one another which have presently been discussed, it seems reasonable to conclude that the observations being made here, regarding Stich's Syntactic Theory of Mind (and behavior explanation and prediction), apply with equal force to the alternative syntactic interpretation functions Stich refers to in *The Fragmentation of Reason*.

To begin with, Stich's argument for the possibility of syntactic explanations of behavior is persuasive, in his restricted sense of explanation and the explanada being considered. That is, in the *autonomous* sense of behaviour explanation, having nothing to do with external environmental stimuli or personal histories (e.g., the Chateau Lafitte and robot examples above) and everything to do with the internal structure of the brain. However, the possibility of syntactic predictions of behavior is considerably less persuasive. That is not to say that autonomous or syntactic explanations of behavior are not enormously useful in a modern and complete theory of psychology, particularly in the elimination of dualistic theories of mind. However, something more needs to be said about what exactly it is that syntactic explanations offer, or more accurately, fail to offer. Stich has already indicated one limitation - they will not "predict or explain behavior under any and every description countenanced by common sense." In this way, without reference to the environment - the actual state of the world - the syntactic theory can only offer an infinite list of conditional syntactic formulas indicating how the organism behaves or acts if certain immediate stimuli present themselves to and act upon the organism. As such, syntactic theories are explanations of behavior - in a very restricted, autonomous sense of the words 'behavior' and 'explanation' - after the stimulus and the organism's response have occurred. In other words, syntactic

theories are retroactive and therefore purely and only a form of *explanation* of behavior, not prediction of behavior.

Now it is often, but not always, the case that the theory which explains is also the theory which predicts. Unfortunately, for Stich, that is not the case here. Nothing in the infinite list of conditionals which syntactic theorists employ to explain my behavior (or more accurately, my actions) will allow them to predict what I will do tomorrow at 8:30 in the morning. Or, for instance, one might want to predict (under some circumstances, in a general way, and under others, in a very specific way) what Mary would do if she found herself in a building that was on fire. Well, the list of conditional syntactic formulas would be infinite, including formulas like: i) $F \supset L$ (if the building is on fire, then she leaves): ii) $L \supset D$ (if she wants to leave, then she goes through the door): iii) $T \supset R$ (if there are two differently colored doors, then she chooses the one painted red): iv) $N \supset O$ (if neither door is painted red, then she looks for one that is): etc.. The color of the door, is but one small, possible variable, if you will, in the infinite conditional possibilities available under the syntactic theory of behavior explanation. The infinite conditional syntactic formulae, are supposed to capture all these variables, thereby representing all the possible contingencies in the way the state of the world could actually be and the various ways that Mary might react to those states of the world. I could not, given these infinite possibilities, (either generally or specifically) *predict* Mary's behavior

without knowing what the environment, in which she is operating, is actually like: and this is only possible where reference to the environment (the world) is possible, which is itself only possible through the characteristic features of a semantic theory (i.e., truth, truth-conditions, reference, etc.), and not an autonomous or syntactic description of her behaviour.

But perhaps this position is a bit too strong. It might be argued that syntactic theories can at least predict behavior in the following two ways: First, the syntactic theorist knows that certain brain states are often followed by certain other brain states. So when brain state *A* is detected, the theorist can predict that the organism will at some future time, be in brain state *B*, because brain state *B* often follows brain state *A*. The detection of brain state *A* will then allow for the prediction of the behavior associated with brain state *B*. Another way syntactic prediction of behavior might be possible is if the time frame for prediction is significantly reduced. In this way, syntactic theorists can predict behavior based on brain states provided the brain state detected is *immediately* followed by the behavior being predicted. Now, what does 'immediate' mean here? The answer is, it would have to be a period of time that precluded the significant chance of any outside environmental stimuli from acting on the organism to change its behavior or brain state. Or, so long as the brain state of the organism and all the successive brain states as generated by incoming environmental stimuli up to the time of the behavior to be

predicted are detected by the syntactic theorist, then prediction of the behavior in question is possible. In both of these cases, however, without reference to the outside environment and stimuli, the range of predictable behaviors and the reliability of the predictions would be very limited. It is in this way that autonomous descriptions and syntactic predictions of behaviour display their weaknesses.

A second feature syntactic theories do not offer is any commonsense, semantic notion of truth or falsity. This is why Stich says we know in advance that this huge space of syntactic theories contains no true or false beliefs. Now, strictly speaking, syntactic theories, in as much as they mirror what goes on in formal logic, *do* involve a notion of truth and falsity. However, this notion does not necessarily concern itself with reference to anything in the real world or the external environment. The notion of truth and falsity found in syntactic theories is like that found in the truth tables of logic. For instance, *if* the antecedent were true and *if* the consequent were true, then the logical formula would be true: having no concern whatsoever for the nature of the subject matter or whether the antecedent and consequent are actually true of the real world (or environment). There is nothing being said here about reference to the actual state of the world (our environment), or truth-conditions. Syntactic theories, as Stich and Godfrey-Smith are so keen to point out, also do not concern themselves with "past links to the environment" or the "etiologies of stimuli". Presumably, this rejection of causal histories, the

environment, and etiologies, would include, in Mary's case, things like: who uttered "the building is on fire"; when it was uttered; where it was uttered; the tone in which it was uttered; whether it was uttered in earnest or jest; uttered metaphorically or in hyperbole, or as an act of malicious tantalization, etc..

Along with Stich's syntactic account of behavior explanation, his claim that theories, chemical, psychological, or otherwise, elaborate or delimit the language in which explanada are to be described is certainly acceptable. In this way, they limit the range and types of explanations they can provide. Stich himself readily admits this.

> In evolving hybrid non-autonomous behavioral descriptions, common sense produces behavioral descriptions that are more fine grained than those that would be available if we restricted ourselves to autonomous descriptions. There is nothing unreasonable about this, since often enough our practical concerns demand some more fine grained descriptions of behavior. ...the folk notion of believing that p is an amalgam of historical, contextual, ideological and perhaps other considerations. No doubt this way of slicing the mental pie proved itself to be efficient and useful in the day-to-day business of dealing with other people. ...The thrust of the autonomy principle, however, is that by building historical, contextual, and ideological features into mental state descriptions, folk psychology has taxonomized states too narrowly, drawing distinctions which are unnecessary and cumbersome when we are seeking a systematic causal explanation of behavior. To believe that P is to be in an autonomous functional state *and* to have a certain history, context, and ideological relation to the ascriber. These further factors can surely be studied by various disciplines. But they have no place in a science aimed at explaining behavior.[76]

On the other hand, it is this failure (Stich would likely characterize it as an advantage) of the syntactic interpretation function to adequately or

comprehensively capture non-autonomous or commonsense behavior descriptions which generates grounds for pause. In fact, Stich openly admits exactly this, saying:

> Sometimes there will be a readily available commonsense description of the autonomous component of a non-autonomous act [the complex hybrid]. But this is not always the case.[77]

In this way, Stich seems to concede that syntactic accounts of behavior are, at times, outside the realm of commonsense description. (Stich is silent on exactly how many and what kinds of times are at issue here.) You may recall that Stich had certain standards which any interpretation function was expected to meet (pp. 22 and 23 above). Part of those standards are that it should: i) "explicate and explain a well-entrenched preexisting intuitive concept or capacity;" and ii) "cleave reasonably closely to commonsense practice." It would seem that at least some of Stich's alternative syntactic interpretation functions might actually violate the very standards he set for an interpretation function. There is again, of course, the possibility that the commonsense intuitions of the man or woman on the street could be changed with retraining or our children could be raised to have commonsense intuitions other than our own. However, there is no talk of that here. In fact, Stich is quite clear. Commonsense descriptions of behavior such as Mary's are often, unlike syntactic descriptions, "richer and more complex," including factors like "the history of the individual in question, the history of the terms he uses, the linguistic, social, legal, and ritual practices that obtain in the society of which he is a part." [78]

He is also clear that commonsense descriptions, or for our present purposes, commonsense interpretation functions, are not the concern of syntactic theories: just as the chemical composition of a boiled Lafitte is not our concern when we want to know why its market value has dropped. Whatever the commonsense intuition involved and whoever possesses the intuitions, it seems that Stich is separating the syntactic theories of interpretation from the *full* realm of commonsense intuitive theories. The question now is, how large is this separation? Does the separation result in a failure to agree *by and large* with the judgments of the man or woman on the street or to cleave *reasonably closely* to commonsense practice? And finally, in as much as they deviate from commonsense, do syntactic theories of interpretation leave the person operating on them any worse off?

To answer these questions, we need to turn our attention to how Stich deals with claims like Pylyshyn's (p. 77 of this thesis), that syntactic theories inadequately explain Mary's building-leaving behavior. As has been mentioned, Stich, like Godfrey-Smith, argues that nothing would be lost were syntactic theories to replace semantic. Stich's syntactic answer to 'Why did Mary come running from the building?' is complex and will be quoted at length here.

> First, she had a long-standing D-state 'Desire-state' whose syntactic form was that of a conditional, viz. F ⊃ L 'corresponding to the desire to leave the building if it is on fire', where F and L are themselves syntactically well-formed strings. F and L should be thought of as syntactically complex, and thus Mary is likely to have many further B-states 'Belief-states' and D-states which involve various parts of F and L, compounded

> with other symbols into well-formed formulas (wffs). But none of these "background" B-states and D-states will function in the explanation of her behavior, save of course for those mentioned below. At this point the story branches into several versions.
> In version 1, Mary began to inhale smoke. This caused her to have the B-state F. The causal connection here is indirect and complex. The direct consequence of inhaling the smoke was that Mary came to have a B-state I (corresponding to the belief state that she was inhaling smoke). From this B-state and the long-standing B-state I ⊃ N (corresponding to the belief that if one is inhaling smoke then there is a fire near by) she inferred (i.e., was caused to add to her belief store) a token of N. And from N and the long-standing B-state N ⊃ F she inferred the B-state F. This last B-state, F, interacted with the conditional D-state F ⊃ L to produce a D-state L. The D-state L in turn interacted with the B-state R ⊃ L (corresponding to the belief that if one runs out the door, then one will leave the building) to produce the D-state R. And the D-state R, finally, led Mary to run out the door.
> In version 2, when Mary picked up the phone an utterance of "the building is on fire!" was heard. This caused her to have the B-state H (corresponding to the belief that she was hearing an utterance of "the building is on fire!"). That B-state led her...to have the B-state F. From here the story rejoins version 1, and leads again to Mary running out of the door.[79]

As was conceded earlier, Godfrey-Smith's and Stich's syntactic explanation of what goes on when Mary comes into contact with stimuli will be granted. In saying this, however, only the autonomous explanation of Mary's behaviour has been admitted. That is, Mary interprets the stimuli in a way that sees her either leave the building or not and all of this is explainable (in a delimited, autonomous sense of the word) without any need to go into semantic notions of reference, truth, etc.. This much is granted, recognizing, of course, the two qualifications which have already been alluded to. The first qualification was provided by Stich himself. That is, syntactic explanations will not always agree with

commonsense explanations, and it is not the job of syntactic theories to provide accounts of the commonsense explanations. The second qualification is that syntactic theories do not provide a satisfactory means of predicting behavior.

But, from a syntactic point of view, what does the notion of Mary *interpreting* stimuli amount to? Godfrey-Smith regularly uses italics and scare-quotes when he refers to any interpretation that Mary might be performing. Consequently, a closer examination of what lay behind the notion of interpretation seems in order here. The more so, because it is intimately connected to Stich's syntactic answer to the question, "Why did Mary come running from the building?" Let's suppose the phone call indicating "the building is on fire!" was the ninth one Mary had received that week, and the eight prior to it had been prank calls. Under these circumstances, when Mary interprets the phone call stimulus, what exactly is she doing? Presumably, she will be skeptical, given the eight prior prank calls, that this is a genuine fire alert. If she is skeptical, then she will have to make a decision - is this message genuine or a prank? Now all sorts of information would be useful in the making of this decision. For instance: is it the same young voice she heard on the other occasions when the call was a prank; is it an older voice that sounds very much like her landlord's; does the utterance sound earnest or insincere; does it seem as though the person has got the right building; etc.. All of this information, while likely processed in mere seconds, would be assessed by Mary. Ultimately, she has to

decide whether what has been uttered is true or false. This is so because, despite Godfrey-Smith's claim that reference and truth-conditional content are causally impotent in the production of behavior, in one very ordinary sense, whether Mary leaves the building or not will depend on the truth or falsity determination she makes with respect to the phone call. If she decides the utterance is false (that is, a prank), then she will not leave; if she decides the utterance is true (that is, the call is genuine), then she will leave the building. Without truth or falsity in this case, how is Mary to decide? Syntactic theories are devoid of truth and falsity, of this kind (the "full-blooded" referential, truth-condition kind), so how could they help Mary make her decision or explain her behavior? It is clear from this, then, that effective *interpretation* relies heavily on truth and falsity; which means interpretation of stimuli does too.

Stich might respond by saying that all the relevant information could be represented in formal symbols such as, $G \supset F$ (corresponding to the belief that if the utterance is genuine then the utterance should be believed), or something similar. The story would then proceed as Stich syntactically indicates. And, he might add, this is all a syntactic theory of behavior explanation and prediction is supposed to do (remember the chemist and the Lafitte's drop in market value). But in order for Mary to know whether the antecedent (G) is *actually* true, doesn't she have to reflect on or examine her environment, her personal history, her intuitions about genuine

and sincere tones of voice, whether the prank calls occurred in any pattern and whether this call matches that pattern, etc.? Otherwise, how is the truth of the antecedent to be established? The thing is, Mary is not concerned with whether the antecedent is true in the syntactic sense of true - that is, in any truth-table, formal logic sense of true. She wants to know whether the phone call is referring to the actual present state of the world. In other words, she wants to know whether the antecedent *is* true of the actual state of the world, not *if* it were true, then what would happen. But if this is so, then surely Stich's syntactic explanation is highly unsatisfactory to Mary. Mary isn't interested in infinite conditional possibilities and that is all the syntactic theory has to offer her. It is these essential semantic features which Godfrey-Smith and Stich fail to capture, when they say Mary *interprets* the stimuli. Let's examine this from another angle.

Stich starts the story saying: "Mary began to inhale smoke."[80] Then he says: "Mary had a B-state I (corresponding to the belief that she was inhaling smoke)."[81] The concern here is much the same as in the previous paragraph. How does Stich, syntactic theories in hand, get from a state of the world to Mary's belief about that state of the world? In short, why does Mary believe she has inhaled smoke? What is the syntactic explanation of *this* behavior? This question is asked because surely the formation of beliefs can be characterized as a human behavior, which the syntactic theory could be expected to explain. Our commonsense,

semantic theory has a ready answer, but it is chock-full of semantic notions like truth, truth-conditions, reference, etc.. The syntactic theory, on the other hand, does not allow for any reference and truth-conditions. If Mary were operating on a syntactic interpretation function, how would she form beliefs about her environment, let alone accurate beliefs about the actual state of affairs in the world? Mary simply cannot interpret the stimuli without these semantic tools; and if she cannot interpret the stimuli, then she cannot form beliefs.

Stich describes the causal connection between Mary inhaling smoke and her belief (or B-state) that the building is on fire (or F) as indirect and complex. The complexity is captured, in Stichian terms, by the following direct causal series: I: $I \supset N$; $N \supset F$; $F \supset L$; $L \supset R$ (see Stich's lengthy explanation above, pp.88 and 89). In plain old English it reads as follows: Mary inhales smoke. If Mary believes she has inhaled smoke, then Mary believes there is a fire nearby. If Mary believes there is a fire nearby, then she believes the building is on fire. If Mary believes the building is on fire, then she wants to leave. If Mary wants to leave, then she wants to run out the door.

But merely from the fact that Mary has inhaled smoke, we cannot necessarily conclude, on any purely syntactic reading, that she believes she has inhaled smoke. For instance, if she had never experienced the inhaling of (or smell of) smoke before or misinterpreted the smoke for some other phenomenon, then perhaps she

would not believe she had inhaled smoke. And if this is true, then we cannot necessarily conclude that she will want to leave and, therefore, will run out the door. Without some notion of reference to and truth-conditions about the state of affairs in the world (both past and present), Mary is unable to form beliefs about the world.

This observation about the intimate connection between belief formation and the external environment (including the semantic tools we use to talk about it such as reference, truth conditions, etc.) adds further credence to the earlier observation that syntactic theories of interpretation would not prove to be useful predictors of behaviour. The reason this is so can be demonstrated in the following alteration of the Mary example: Mary inhales smoke. If Mary believes she has inhaled smoke, then she *does not* believe that there is a fire near by. This puts the rest of the causal series in question. But how is it possible to say, despite her having believed she inhaled smoke, that Mary does not believe there is a fire nearby? What allows this alteration in the example is that the smoke Mary inhaled was actually from toast burning or popcorn burning, or incenses, or soldering, etc.: and Mary believed one of these to be the source of the smoke (something, incidentally, she could not do under a purely syntactic interpretation function). In other words, our commonsense, semantic explanation of Mary's behavior demands more be known about Mary's belief than merely that she inhaled smoke. And it demands things which the syntactic theories are unable to offer, such as: what the smoke smelled like:

what were its likely origins: had she smelled a similar smell in the past: was it coming from the building or from outside somewhere, etc. With these answers in hand, the commonsense, semantic interpretation theory could not only facilitate belief formation for Mary, where the syntactic could not, but it could also better predict whether Mary was going to leave the building or not.

Now, Stich might respond by saying that the particular symbolic representations he provided (i.e., $I \supset N$; $N \supset F$; etc.) are simply not sophisticated enough to capture all these factors, but that he could have used, say, quantified predicate logic to do the symbolizing. This more sophisticated symbolic system could capture what commonsense demands. This might be true: and the claim will not be contested here. The goal here was only to thrash out a more complete understanding of what Godfrey-Smith and Stich meant by Mary's *interpreting* stimuli. If they do not at least have in mind interpretation along the lines that have been suggested, that is, referentially and truth-conditionally dependent interpretation, then their intended meaning is lost. Perhaps they do have another meaning in mind and maybe that's why Godfrey-Smith uses italics and scare-quotes when he speaks of interpretation. If this is so, then surely a detailed explanation of this meaning is in order.

At the outset of this discussion, the objective was to establish two things: i) alternative syntactic interpretation functions, in many cases, fail to meet

Stich's standards of what a good interpretation function is: and ii) alternative syntactic interpretation functions will likely leave anyone operating on them worse off (than were they to operate on our commonsense semantic function). In the first instance, Stich, himself, admits that syntactic theories of behavioral explanation and predication, including syntactic interpretation functions, do not always lend themselves to commonsense interpretation or description, which puts at least some of those functions found in the "huge space" of alternatives outside the realm of commonsense, and therefore outside the Stichian standards of a good interpretation function. Stich says this is how it is *and* how it should be. The second matter was whether these alternative syntactic functions would leave someone operating on them worse off. It was concluded that were Mary operating on such a function she would likely be unable to form beliefs about the actual state of the word. Nor could she determine whether statements other people make about the state of the world are accurate or true (or not), thereby making it impossible for her to effectively rely upon others' statements. Surely this would significantly reduce Mary's ability to meaningfully communicate with others and her environment. In turn, were we operating on such a function, we would be unable to effectively predict Mary's behavior. And without effective prediction of behavior, there can be no effective communication, teaching or learning. After all, how can we rely upon and make inferences from utterances, body-language, and written language which do not allow

us to even reasonably predict what the communicator will say, do or write in the future? The reason for this is that the effective prediction of behavior is facilitated, in no small part, by our knowledge of the external environment (both past and present), and without semantic tools like reference and truth-conditions, construction of this knowledge would be desperately impeded, if not completely halted.

Stich claimed that anyone who intrinsically valued true beliefs (established by our commonsense, semantically-based interpretation function), was being "profoundly conservative", especially in the face of the "huge space" of alternative functions available. He said that, given the idiosyncratic and partial nature of our commonsense interpretation function, anyone who intrinsically valued true beliefs, without investigating the available, in this case, syntactic alternatives, has "not made a reflective choice". In light of the preceding analysis, it seems that we should nevertheless value our commonsense interpretation function and the true beliefs it can produce.

INSTRUMENTAL VALUE OF TRUE BELIEFS

Having critically examined the possibility that true beliefs could be of intrinsic value and deciding in the negative on this matter, Stich redirects his critical eye toward the argument for the instrumental value of true beliefs.

> To explore the instrumental value of true beliefs is to ask whether having true beliefs will lead to something else that is valued, where

> the *something* else may itself be valued either instrumentally or intrinsically. Since people probably value many things intrinsically, and surely value many things instrumentally, I'll make no attempt to argue that having true beliefs *could not be* instrumentally valuable. For to demonstrate this would require showing that true beliefs don't facilitate *anything* that people might sensibly value. I have no idea how one might argue for so sweeping a conclusion, and my goal is much more modest. It is widely believed that the instrumental value of true belief is obvious - that having true beliefs is clearly good for lots of things. However, it is my contention that this doctrine is anything but obvious. It requires some serious argument of a sort that, so far as I know, no one has even begun to provide. In support of this contention I'll make three points, the first aimed at clarifying what is at issue, the second aimed at short-circuiting one argument that seems to tempt a fair number of people, and the third sketching a general difficulty that any argument for the instrumental value of truth must overcome.[82]

Before the three points mentioned above are considered, a clarification seems in order. In the above passage, Stich's argument against the instrumental value of true beliefs is qualified in that he claims he will "make no attempt to argue that having true beliefs *could not be* instrumentally valuable." He says to argue along these lines would involve proving that "true beliefs don't facilitate *anything* that people might sensibly value;" and this is a conclusion for which he can provide no argument. At other times in the presentation of his argument against the value of true beliefs, or more accurately, the *obvious* value of them, he takes a different tack. He starts out Chapter Five saying: "What I propose to argue in this chapter is that once we have a clear view of the matter, most of us will not find *any* value, either intrinsic or instrumental, in having true beliefs."[83] (emphasis is mine) Near the end of Chapter Five his goal is considerably more modest: "It is widely believed that the instrumental

value of true beliefs is obvious - that having true beliefs is clearly good for lots of things. However, it is my contention that this doctrine is anything but obvious." [84] The most charitable way to read Stich, in the face of these disparate goals, is to take him to be arguing, not that most of us will not find *any* value, either intrinsic or instrumental, in having true beliefs, but that there isn't any *obvious* (intrinsic or instrumental) value in having true beliefs.

Something Fred Dretske says about Stich's goal, as it is now to be read, helps put the task at hand into perspective.

> Since, as we all know, philosophers can find almost everything, and certainly everything with which they disagree, "far from obvious," this is not much to get one's teeth into. It puts anyone...who finds the value of truth pretty obvious in the awkward position of arguing that the instrumental value of truth is not *that* far (however far that is) from being obvious.[85]

We now begin the presentation of Stich's three points, which, when combined, lead to the conclusion that the instrumental value of true beliefs is far from obvious. With reference to the first point - namely, clarification of the issue - Stich has this to say:

> It might be thought that to ask whether truth is instrumentally valuable is to ask whether having true beliefs would increase the likelihood of some other valued goal being attained. But posing the question in this way is seriously misleading, for it does not specify what the instrumental value of true beliefs is to be compared with. In the absence of such a specification, it would be easy to suppose that the relevant comparison was between true beliefs and *false* ones and that our question was whether having true beliefs is more instrumentally valuable than having false beliefs. But showing that the answer to this question is 'yes', though hardly trivial, is not nearly enough. What really needs to be shown is not just that true beliefs are more conducive to some

> independently desirable goal than false beliefs but also that true beliefs serve us better than TRUE* ones or TRUE** ones, or any of the other categories of belief picked out by interpretation functions that don't happen to be favoured by intuition and tradition... Moreover, it will not always be the case that TRUE**...* beliefs which aren't true will be false. For some of the mental states to which TRUTH**...* CONDITIONS are assigned may have no truth conditions at all. So there will be TRUE**...* beliefs that are neither true nor false. Thus showing that true beliefs are better at achieving some goal than false ones does not come close to establishing that true beliefs are more instrumentally valuable in pursuit of that goal than TRUE**...* ones.[86]

It should be clear from what has been said thus far that Stich's labeling of TRUE**...* beliefs is misleading. The assigning of a truth value to a (statement of) belief is achieved by determining whether the belief's truth conditions have been met or not, thereby simply making the belief either true or false. This process, Stich would acknowledge, has nothing to do with how the truth conditions were derived or whether they are presented as TRUTH CONDITIONS**...*. Stich's talk of TRUE**...* beliefs tends to create the illusion of other categories of truth and true beliefs, which he then sees as competing against one another. In fact, using different interpretation functions or theories of reference on sentences only produces statements with different meanings and truth conditions. That we could constructively compare in some competitive way - with the hope of increasing the likelihood of achieving our goals - statements having entirely different meanings now seems irrational. This being the case, Stich's comment that not all TRUE**...* beliefs which are not true will be false is incorrect. Where different statements

which contain words having different meanings are concerned, the characterization of those statements as TRUE**...* or FALSE**...* is misleading - they are either true or false. So, when those beliefs which Stich refers to as TRUE**...* are not true, they will be false.

This much said, the notion of interpretation functions that produce TRUTH CONDITIONS**...* which are not truth conditions and TRUE**...* beliefs which are neither true nor false certainly seems mistaken. It is possible that Stich is referring here to those syntactic interpretation functions which have been discussed at length above. But given these earlier arguments, the idea that syntactic interpretation functions would produce anything like truth conditions or true beliefs seems inappropriate. Not only that, if these syntactic interpretation functions are to be compared with commonsense semantic interpretation functions, then, given the earlier analysis of such functions, it seems that there can be no comparison. Commonsense semantic interpretation functions are evidently superior to their syntactic cousins - or put in Stichian terms, they would clearly leave us better off.

The second point Stich makes involves a line of argument in favour of the instrumental value of true beliefs which, as he says, many people find tempting. It is an argument with which he contended earlier, in Chapter Three of his book - the argument from evolution. One is inclined to agree with Stich's criticisms of this line of argument, although agreement here in no way effects the analysis presented in this

thesis thus far. In short, Stich is right to claim that there is no reason to think that biological or sociological evolution will produce an optimal interpretation function. In light of this, Stich's third and final point in support of his argument against the instrumental value of true beliefs will now be considered. Stich characterizes this point as a general difficulty that any argument for the instrumental value of truth must overcome.

> The final point I want to make under the heading of the instrumental value of true belief is that in many cases, we already know that having true beliefs would not be the *best* way to achieve our more fundamental goals. Consider survival. Is true belief *always* more conducive to survival than false? Clearly the answer is no.[87]

Stich goes on to provide an example in which the having of a true belief did not facilitate the achievement of a more fundamental goal - namely, survival. This example, which is dubbed here, the "Traveling Harry" example, is designed to show that "in many cases we already know that having true beliefs would not be the *best* way to achieve our more fundamental goals." The example unfolds like this: Harry has a flight to catch. He believes the plane leaves at 7:45 a.m. and makes all the necessary arrangements including ordering a cab the night before and asking his wife to wake him at 6:30 a.m.. Harry's belief about the plane's departure time was true and he was airborne by 7:50 a.m.. Unfortunately, the plane crashed, and Harry died. Stich draws the following philosophical moral from the story:

> Had Harry falsely believed that the flight left at 8:45, he would have missed the flight and survived. So true belief is sometimes

> less conducive to survival than false belief. Now it might be protested that this is an illusion, since Harry had some other false beliefs that contributed to his death. On arriving at the airport, he no doubt believed, falsely, that the plane would not crash. If he had a true belief on this matter, he would never had boarded the plane. This protest misses the point, however. For the question at hand is not whether omniscience would foster survival but whether more true beliefs are always better than fewer. In the case of Harry it is clear that if he had had one more false belief and one fewer true one, and if everything else in his cognitive life had remained as much the same as possible, his life would have been longer.[88]

Clearly, Stich is right. True beliefs are not *always* more conducive to our fundamental goals than false beliefs, or for that matter, TRUE**...* beliefs or FALSE**...* beliefs. But it is not clear what is to be made of this observation. It doesn't show that true beliefs or truth (in the meta-sense of the word) are not of obvious instrumental value. If this were so, then every interpretation function and the beliefs or (as Goldman would insist) BELIEFS**...* they produce (TRUE**...* or otherwise) would suffer from the same character flaw. The scenario which Stich depicts in the "Traveling Harry" example is applicable with equal force to all interpretation functions and the beliefs or BELIEFS**...* they produce. In light of this, all functions are on equal ground in this respect, making selective criticism on this basis inappropriate. Perhaps Stich has in mind an alternative interpretation function which does foster omniscience: but this seem unlikely. Instead, he has in mind something a little more pragmatic. In Chapter Six of his book, he advocates that we should choose those beliefs (and interpretation functions) which will leave us better off, in the sense that they will most likely allow us to achieve our fundamental

(intrinsic and/or instrumental) goals. Unfortunately, the phrase, 'most likely', has considerable probabilistic overtones, and probabilities, by their very nature, do not provide guarantees in any omniscient sense. So, choosing the highest probability function or belief will not guard against a chaotic and often unpredictable world *proving*, along lines of the "Traveling Harry" example, that this pragmatic formula is not *always* conducive to the achievement of our more fundamental goals.

In his concluding remarks, Stich has this to say about the efficacy of his argument against the obvious instrumental value of true beliefs:

> ...the instrumental value of having true belief is far from obvious. It is surely not the case that having true beliefs is *always* the best doxastic stance in pursuing our goals, and it would be no easy matter to show that believing the truth is *generally* or even occasionally! instrumentally optimal.[89]

But perhaps we can demonstrate that believing the truth is occasionally instrumentally optimal. Let's consider Stich's "Traveling Harry" example again. Remember: Harry's belief about the plane's departure time was true and his fundamental goal of survival was intact. Only one feature of this story need be changed to demonstrate the occasional instrumental value of Harry's true belief - the plane does not crash! Suffice it to say that, in this case, it *is* instrumentally optimal to believe the truth.

But what about the possibility that truth is *in general* or *in the long run* instrumentally optimal? Stich claims that, as far as he knows, "no one has any inkling of how that argument might go".[90] It is hoped, however, that what has been

said throughout this thesis is at least a beginning to such an argument. It is a beginning that shows how truth and true beliefs do not have the sort of competition Stich would like us to believe they do; and that even if many of these alleged alternatives did turn out to be genuine competitors, they would fail to measure up. In fact, if this is true, then the question of the *optimality* of true beliefs seems moot.

In conclusion, it would seem that the three points mentioned above, which Stich employs in his argument against the obvious instrumental value of true beliefs, have lost most of their force in the face of the earlier analysis found in this thesis. This, combined with the observations made here, work to lend credence to the claim that there is indeed obvious instrumental value in true beliefs.

Section 5

SUMMARY REMARKS

The purpose of writing this thesis was to explore, in some detail, the nature of Stich's arguments against the (intrinsic and instrumental) value of truth and true beliefs. In the process of this exploration an attempt was made to expose some of the more fine-grained features and implications of Stich's most controversial and ambiguous claims. As a piece of philosophy, *The Fragmentation of Reason*, has performed its job admirably. It has forced us to re-examine and re-evaluate certain deeply entrenched tenets of twentieth century epistemology and philosophy. The aim here was not necessarily to defend those tenets, but rather to critically examine the force and merit of those arguments which Stich employs in his deconstruction of them.

PARTIALITY AND IDIOSYNCRASY

The goal here was simple. Three crucial aspects of Stich's argument for the partiality and idiosyncrasy of the causal/functional interpretation function were examined: i) the nature of commonsense intuition: ii) the claim that there is no obvious reason to believe that alternative interpretation functions would leave us any worse off (than would the function sanctioned by our present-day commonsense

intuition); and iii) the nature of the alleged competition for truth, true beliefs and the causal/functional interpretation function.

With respect to commonsense itself, Stich clearly did not have in mind certain features of our commonsense intuition when he formulated his argument for the possibility of alternative, competing interpretation functions. He stated that counterintuitive references and word-world mappings that were a variation on the commonsense theme would automatically produce interpretation functions that were competitors for our commonsense interpretation function. Review of the concept of commonsense intuition revealed, however, that, in many ways, our commonsense intuition is receptive to and able to accommodate both counterintuitive and variable word-world mappings, without having to change theories of reference or interpretation functions. Stich did not consider the open-ended, changing nature of commonsense intuition. Instead he saw it as a static, finite list of intuitions which would require the adoption of different theories of reference and interpretation functions, as well as radical retraining and training (in the case of our children), in order to accommodate counterintuitive and/or varying word-world mappings.

With the help of Fred Dretske and Alvin Goldman, it was established that certain qualifications ought to be placed on Stich's claim that there is no obvious reason to think alternative interpretation functions would leave us any worse off. Aside from the constraints set by Stich's own standards for a good interpretation

function, Dretske and Goldman pointed out that a function must not produce irrelevant beliefs. More specifically, Goldman said that interpretation functions must comply with his "action linkage principle", where the function in question must produce beliefs which are intimately connected to the actions one takes in the pursuit of one's goals. Dretske demanded that interpretation functions produce beliefs with truth conditions that can be tracked - truth conditions about which the belief in question (under normal circumstances) carries information - and that, in the case of direct observational beliefs, the function must accurately explain the causal interactions in which the brain is engaged (i.e., direct perception). Analysis of Stich's own example of an alternative interpretation function - the one that mapped 'water' onto H_2O and XYZ - also revealed some interesting constraints for alternative interpretation functions. Among them are: i) the requirement that a single referential term not be mapped onto two or more separate and distinct objects or classes of objects without some further means of distinguishing the objects or classes of objects and ii) the requirement that a single referential term not be mapped onto two separate and distinct objects or classes of objects one of which is nonexistent while the other does exist. Should an alternative function violate either of these constraints, then, it is argued, communication, teaching, and learning would be thwarted. In these ways, any function which did not comply with these constraints would, in fact, leave someone worse off.

The final aspect of Stich's partiality and idiosyncrasy argument is concerned with competition for truth and true beliefs. In conjunction with the first two aspects (the nature of commonsense intuition and the question of whether alternative functions would leave us any worse off), it was argued that Stich's alternative interpretation functions did not produce competition for true beliefs or truth. Relying on the accepted, general method by which truth values are determined and the recognized distinction between sentences and statements, it was decided that the truth conditions of alternative interpretation functions (as Stich says, TRUTH CONDITIONS**...*) would not necessarily result in the adoption of a different notion of truth. This was concluded on the grounds that alternative interpretation functions would not produce TRUE**...* beliefs rather than true beliefs, given the accepted, general method of determining truth values. In both instances, Stich's use of capitalization and asterisks merely created the illusion of competing notions of truth and true beliefs.

Ultimately, the analysis of Stich's partiality and idiosyncrasy argument demonstrated that the notion of competing, alternative interpretation functions and their resultant competing concepts of truth and true beliefs (i.e., TRUTH**...* and TRUE**...* beliefs) is not as clear as Stich would have us believe. In fact, the notion of competition itself in this matter was drawn into question.

INTRINSIC AND INSTRUMENTAL VALUE OF TRUE BELIEFS

The observations made here are, in part, dependent upon those made in the preceding section (Section 3 - on the partiality and idiosyncrasy of the causal/functional interpretation function). Stich attempted to show that the intrinsic valuing of true beliefs over TRUE**...* ones was highly conservative, given the vast number of available alternatives and that there was no obvious reason to prefer our commonsense interpretation function over these available alternatives. He suggested that many of these alternative interpretation functions would have neither truth conditions nor would they produce beliefs that were either true or false. These alternative interpretation functions, it was decided, were syntactic in nature.

Aside from the arguments that were constructed in the previous section (Section 3) - which purported to establish that the notion of alternatives which are in competition with truth and true beliefs was, as a whole, ill-founded - this section (Section 4) argued that, in many cases, the syntactic interpretation functions Stich offers as alternatives violate his standards of a good interpretation function, thereby making them substandard. It was also argued that syntactic interpretation functions were in fact inferior to semantic, in that they either impeded or prevented communication, teaching, and learning. This deficiency was the result of their operating on autonomous descriptions of behaviour (which delimited their language to strictly those events which happened in the brain, to the exclusion of any reference

to external environmental stimuli), which, in turn, effectively precluded their ability to predict behaviour in any useful way. It was decided that in light of this, there was good reason to prefer semantic interpretation theories to syntactic.

As far as the instrumental value of true beliefs was concerned, Stich worked to establish three points which, when combined, purportedly demonstrated that the instrumental value of true beliefs was anything but obvious. Those three points were: i) A move to show that merely proving true beliefs were more valuable than false ones was not enough to secure their instrumental value. Instead, one also had to show that true beliefs were more valuable than Stich's alternatives (i.e., TRUE**...* beliefs); ii) This point involved the defusing of an argument from evolution which claimed that true beliefs were the result of evolutionary processes and therefore they had to be the most optimal. (There was agreement with Stich's critique here, although this in no way affected the rest of the analysis found in this thesis.); and iii) There was an attempt to show that in many cases we already know of instances where valuing true beliefs would not be conducive to our more fundamental goals (e.g., survival) and therefore they are not always of instrumental value or, alternatively, even occasionally of instrumental value.

The force of the response to the first point lay in the discussion which preceded it. From Section 3 of this thesis, through to the part in Section 4 on the intrinsic value of true beliefs, it was argued that Stich's claim that there was

alternative competition for truth and true beliefs was merely an illusion. His use of capitalization and asterisks was considered misleading in light of the general technique for determining truth values and the recognized distinction between sentences and statements. There was also no competition to be found for truth and true beliefs in Stich's syntactic interpretation theories since, on the one hand, the idea that they could produce anything like truth (or TRUTH**...*) and true beliefs (or TRUE**...* beliefs) seemed inappropriate and, on the other, as competitors, they proved inferior (i.e., inability to usefully predict behaviour and they impeded or prevented communication, teaching and learning).

The third point (recalling that his second point was conceded) was answered by revealing how all interpretation functions, including those of Stich's alternatives, suffered from the flaw which he relied upon in the "Traveling Harry" example - a flaw which said that the having of true beliefs is not *always* conducive to the attainment of our more fundamental goals. But surely, given the unpredictable and changing nature of our world, any interpretation function, short of an omniscient one, would suffer from this same infirmity, making isolated criticism of any one of these functions on this basis inappropriate and uninstructive. Stich also claimed it would be difficult to show that true beliefs are *generally* or even occasionally instrumentally optimal. But by altering the "Traveling Harry" example so that Harry's plane did not crash, it was demonstrated how true beliefs could in fact

occasionally be instrumentally optimal. As to their general optimality, it was acknowledged that no full-fledged argument was offered here, but it was hoped that the arguments presented in this thesis were at least a good beginning to such an argument.

ENDNOTES

[1] Stephen Stich, "The Fragmentation of Reason: Precis of Two Chapters" (1991) 51:1 Philosophy and Phenomenological Research 179 at p. 179
[2] Stephen Stich, *The Fragmentation of Reason*, (Cambridge, Massachusetts: The Massachusetts Institute of Technology Press, 1993) at p.14
[3] *Ibid.* at p.29
[4] *Ibid.* at p.11
[5] *Ibid.* at p.54
[6] *Ibid.* at p.15
[7] *Ibid.* at p.16
[8] *Ibid.* at p.74
[9] *Ibid.* at p.75
[10] *supra* note 1 at p.179
[11] *supra* note 2 at p.17
[12] *Ibid.* at p.81
[13] *Ibid.* at p.82
[14] *Ibid.* at p.83
[15] *Ibid.* at p.84
[16] *Ibid.* at p.91
[17] *Ibid.*
[18] *Ibid.* at p.92
[19] *Ibid.* at p.93
[20] *Ibid.* at p.94
[21] *Ibid.* at p.95
[22] *Ibid.* at p.97
[23] *Ibid.* at p.103
[24] *Ibid.* at p.104
[25] *Ibid.*
[26] *Ibid.*
[27] *Ibid.* at p.105
[28] *Ibid.*
[29] *Ibid.* at 106
[30] *Ibid.* at p.107-108
[31] *Ibid.* at p.108
[32] *Ibid.*
[33] *Ibid.*
[34] *Ibid.* at p.109
[35] *Ibid.*
[36] *Ibid.* at p.110
[37] *Ibid.*
[38] *Ibid.* at p.111
[39] *Ibid.* at p.116

[40] *Ibid.* at p.111
[41] *Ibid.* at p.112
[42] *Ibid.* at p.113
[43] *Ibid.* at p.114
[44] David H. Sanford, "The Anastylosis of Reason: Fitting Together Stich's Fragments" (1991) 35 Inquiry 113 at p.130
[45] *Ibid.* at p.130
[46] *supra*, note 2 at p.114
[47] *Ibid.*
[48] *Ibid.* at p.115
[49] *Ibid.*
[50] *Ibid.* at p.116
[51] *Ibid.*
[52] Fred Dretske, "The Fragility of Reason" (1992) 31:2 Dialogue 311 at pp.317-318
[53] *Ibid.* at p.318
[54] *Ibid.* at p.318-319
[55] Alvin Goldman, "Stephen P. Stich: The Fragmentation of Reason" (1991) 51:1 Philosophy and Phenomenological Research 189 at 191
[56] *Ibid.*
[57] *supra*, note 2 at pp.116-117
[58] *Ibid.* at p.117
[59] *supra*, note 55 at pp.190-191
[60] *Ibid.* at p.118
[61] *Ibid.*
[62] *Ibid.*
[63] *Ibid.* at p.119
[64] *Ibid.* at p.120
[65] *Ibid.*
[66] Peter Godfrey-Smith, "Why Semantic Properties Won't Earn Their Keep" (1986) 50 Philosophical Studies 223 at p.223
[67] *Ibid.* at p.225
[68] *Ibid.* at p.228
[69] *Ibid.* at p.229
[70] *Ibid.*
[71] Stephen Stich, *From Folk Psychology to Cognitive Science*, 1st ed. (Cambridge, Massachusetts: The MIT Press, A Bradford Book, 1983) at p.167
[72] *Ibid.*
[73] *Ibid.* at p.168
[74] *Ibid.*
[75] *Ibid.* at p.170
[76] *Ibid.* at p.169
[77] *Ibid.*
[78] *Ibid.*
[79] *Ibid.* at pp.174-175

[80] *Ibid.* at p.175
[81] *Ibid.*
[82] *supra*, note 2 at p.121
[83] *Ibid.* at p.101
[84] *Ibid.* at p.121
[85] *supra*, note 52 at p.316
[86] *supra*, note 2 at p.121
[87] *Ibid.* at p.122
[88] *supra*, note 2 at p.123
[89] *Ibid.* at p.124
[90] *Ibid.* at p.123

BIBLIOGRAPHY

Alan, Barry G., *Truth in Philosophy*, 1st ed. (Harvard University Press, Cambridge, 1993)

Biro, John & Kirk Ludwig, "Are There More than Minimal *A Priori* Limits on Rationality?" (Mr 1994) 72:1 *Australian Journal of Philosophy* 89-102

Cohen, Jonathan L., "Stephen P. Stich, *The Fragmentation of Reason*" (Mr 1991) 51 *Philosophy and Phenomenological Research* 185-188

Davidson, Donald, "Truth and Meaning" (1967) 17 *Synthese* 304-323

Devitt, Michael, "Donnellan's Distinction" (1981) 6 *Midwestern Studies in Philosophy* 511-524

Devitt, Michael, "Thoughts and Their Ascription" (1984) 9 *Midwest Studies in Philosophy* 385-420

Devitt, Michael, "Against Direct Reference" (1989) 14 *Midwest Studies in Philosophy* 206-240

Donnellan, Keith S., "Reference and Definite Descriptions" (1966) 77 *Philosophical Review* 203-215

Dretske, Fred, "The Fragility of Reason" (Spr. 1992) 31:2 Dialogue 311-320

Field, Hartry H., "Theory Change and The Indeterminacy of Reference" (1973) 70 *Journal of Philosophy* 462-481

Field, Hartry, H., "Mental Representation" (1978) 13 *Erkenntnis* 9-61

Godfrey-Smith, Peter, "Why Semantic Properties Won't Earn Their Keep" (1986) 50 *Philosophical Studies* 223-236

Goldman, Alvin I., " Stephen P. Stich, *The Fragmentation of Reason*" (Mr 1991) 51 *Philosophy and Phenomenological Research* 189-193

Harman, Gilbert, "Justification, Truth, Goals, and Pragmatism: Comments on Stich's *The Fragmentation of Reason*" (Mr 1991) 51 *Philosophy and Phenomenological Research* 195-199

Hetherington, Stephen Cade, "Furthering Stich's Fragmentation" (Ja 1993) 53:1 *Analysis* 42-44

James, William, *Pragmatism*, Edited by Bruce Kuklick (Hackett Publishing Company, Indianapolis, Cambridge 1981)

Joachim, H.H., *The Nature of Truth*, 1st ed. (Oxford University Press, 1906)

Konvitz, Milton R. and Gail Kennedy, eds., *The American Pragmatists*, 5th ed. (Meridian Books, The world Publishing Company, Cleveland, New York 1967)

Kripke, Saul, "Speaker's Reference and Semantic Reference" (1977) 2 *Midwest Studies in Philosophy* 255-276

Kripke, Saul, *Naming and Necessity*, 2nd ed. (Harvard University Press, Cambridge, Mass., 1980)

Lycan, William G., "Kripke and The Materialists" (O. 1974) 71 *Journal of Philosophy* 677-689

Lycan, William G., "Why We Should Care Whether Our Beliefs Are True" (Mr. 1991) 51 *Philosophy and Phenomenological Research* 195-199

McGinn, Colin, "Rigid Designation and Semantic Value" (Ap 1982) 32:127 *The Philosophical Quarterly* 97-115

Pierce, Charles S. "What Pragmatism Is" (Ap 1905) 15 *The Monist* 161-181

Putnam, Hilary, "Meaning and Reference" (N 1973) 70 *Journal of Philosophy* 699-711

Putnam, Hilary, *Mind, Language and Reality*, 1st ed. (Cambridge: Cambridge University Press, 1975)

Pylyshyn, Zenon, "Cognitive Representation and the Process-Architecture Distinction" (1980) 3:1 *Behavioral and Brain Sciences*

Rescher, Nicholas, *Methodological Pragmatism*, 1st ed. (Oxford: Basil Blackwell, 1977)

Sanford, David H., "The Anastylosis of Reason: Fitting Together Stich's Fragments" (1992) 35 *Inquiry*

Stich, Stephen P., "Dissonant Notes on the Theory of Reference" (N 1970) 4 *Nous* 385-397

Stich, Stephen P., "Davidson's Semantic Program" (Jn 1976) 11:2 *Canadian Journal of Philosophy* 201-227

Stich, Stephen P. & Richard E. Nisbett, "Justification and the Psychology of Human Reasoning" (1980) 47 *Philosophy of Science* 188-202

Stich, Stephen P., *From Folk Psychology to Cognitive Science: The Case Against Belief*, 1st ed. (Cambridge MIT Press, 1983)

Stich, Stephen P., "Could Man be an Irrational Animal: Some Notes on the Epistemology of Rationality" (Jl 1985) 64 *Synthese* 115-134

Stich, Stephen P., "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity" (Mr 1988) 74 *Synthese* 391-413

Stich, Stephen P. , "Do True Believers Exist?" (1991) 65 *Proceedings of the Aristotelian Society* 229-244

Stich, Stephen P., "Evaluating Cognitive Strategies: A Reply to Cohen, Goldman, Harman, and Lycan" (Mr 1991) 51 *Philosophy and Phenomenological Research* 207-213

Stich, Stephen P., "The Fragmentation of Reason: Precis of Two Chapters" (Mr 1991) 51 *Philosophy and Phenomenological Research* 179-183

Unger, Peter, "The Causal Theory of Reference" (1983) 43 *Philosophical Studies* 1-45