

UNIVERSITY OF TARTU  
INSTITUTE OF PHILOSOPHY AND SEMIOTICS

Sander Vaus

**The Normativity of Truth in Cognitive Evaluation**

Bachelor's thesis in philosophy

Supervisor: Professor Dr. Daniel Cohnitz

Tartu 2014

## Contents

Introduction .....	3
1. Cognitive Pluralism.....	5
1.1 Is bad reasoning conceptually possible? .....	5
1.2 Is bad reasoning biologically possible?.....	8
2. Evaluative-Concept Pluralism and its Bearing on “Analytic Epistemology” .....	12
2.1 <i>Conceptual</i> analysis.....	12
2.2 Epistemic intuitions.....	13
2.3 The task of epistemology .....	14
3. The Normative Function of Truth in Epistemology.....	16
3.1 On the value of truth.....	16
3.1.1 Causal/functional theory of the semantic properties of mental states .....	17
3.1.2 The limits and idiosyncrasies of the interpretation function .....	18
3.1.3 The value of true beliefs .....	21
3.2 Is truth a pre-eminent epistemic value?.....	24
3.2.1 Cognitive evaluation vs. epistemic evaluation .....	25
3.2.2 How a more truth-conducive cognitive system is worse for attaining the things that an agent values .....	28
3.2.3 The pre-eminence of truth in second-order beliefs .....	30
3.2.4 How Kornblith fails to overcome Stich’s attack on the value of truth.....	31
Conclusion.....	33
References .....	35
Tõe normatiivne roll kognitiivsete süsteemide hindamisel.....	37

## Introduction

In this essay I will defend Stephen Stich's pragmatic theory of cognitive evaluation. The most striking feature of his view is that it doesn't place truth center-stage as a criterion of a good cognitive system. In fact, he argues that having true beliefs is of little value, both intrinsically and instrumentally, and what we should really want is to have beliefs that allow us to attain the things that we *actually* value. This view is contested by Hilary Kornblith, who argues that even if truth is not something we value intrinsically, then it is *always* of instrumental value for attaining the things that we actually want to attain. For him, truth plays a pre-eminent role in cognitive evaluation, and placing value on truth is thus *pragmatically preferable*.

Interestingly, both accounts share a lot of common ground, stemming from their critical views on the classical analytic approach to epistemology. The method of conceptual analysis that relies on pre-theoretic epistemic intuitions for developing its theories is for both unreliable. Kornblith is critical of epistemic intuitions because they are messy and lacking in the kind of theoretical rigor that an investigation of knowledge, understood by him as a natural kind, would require. Stich's problem with pre-theoretic intuitions is that they are learned and culturally derived, and building up epistemological theories from the basis of the idiosyncratic intuitions of male Western philosophers, amounts to either epistemic chauvinism or some form of ethnographic epistemology. Both accounts also appreciate the normative dimension of epistemology, but it is in the details of how this normativity should manifest that they disagree on.

In the first chapter I will present the empirical thesis of cognitive pluralism – that there are significant and systematic differences in the ways different people go about cognition. While there is a lot of empirical evidence for this thesis, it will also have to be shown to be tenable on conceptual and biological grounds. Establishing cognitive diversity is necessary for further considerations of cognitive evaluation, for to properly evaluate different cognitive mechanisms, such mechanisms must be shown to actually diverge in some significant sense.

In the second chapter I will turn to criticize the tradition of analytic epistemology. Relevant to this is another empirical thesis – evaluative-concept pluralism, which maintains that there are significant and systematic differences in the epistemic concepts and judgments that people employ in evaluating cognition. This undermines the intuition-driven methodology of analytic epistemology, as the intuitions employed in building up such epistemological theories are shown to be non-uniform among the population. There is also empirical evidence for

## Vaus, The Normativity of Truth in Cognitive Evaluation

evaluative-concept pluralism, in the face of which, holding onto the analytic method will turn out to constitute something that Stich calls ethno-epistemology.

In the third chapter I will turn to the issue of the normativity of truth in cognitive evaluation. I will first present Stich's argument for why true beliefs, understood for what they are, turn out to have little value in our cognition. Then I will turn to Kornblith's criticism of Stich's view. I will show it to misread Stich on several accounts, and to fail at establishing the pre-eminent value of truth that he insists upon.

## 1. Cognitive Pluralism

In his famous “Epistemology Naturalized”, Quine presents his view of epistemology as being nothing more than a chapter of psychology, i.e. it falls into the scope of the natural sciences (Quine 1969: 82). A common objection to this view is that it robs epistemology of its normative force. This becomes a problem when one of the tasks of epistemology is taken to be the *evaluation* of our cognitive mechanisms. Understood this way, an important part of the epistemological project is that of improving the strategies of reasoning that we employ in our everyday doings. The importance of such a project is partly due to the “bleak” results from a number of studies by experimental psychologists that show humans to often reason very badly in predictable ways (cf. Nisbett and Ross 1980). This naturally raises the question of how we are to determine that an instance of reasoning is a bad one, and if it is shown to be so, then how are we to go about fixing it. To answer these questions, we will first have to know if the divergence in people’s reasoning is genuine, and this is where cognitive pluralism comes to play.

Pointing to empirical evidence of cognitive pluralism isn’t enough to make a good case for it. While quite a lot of such evidence has surfaced in the works of Richard Nisbett and his colleagues (cf. Nisbett and Peng 1999; Nisbett *et al.* 2001; Nisbett 2003), any divergence from the norm could still be chalked up to conceptual confusion, linguistic divergence, suspect methodology and so forth. This is of great importance when considering views that imply the impossibility of significant cognitive diversity. Stich (1990) criticizes two lines of arguments for such a view. First of them argues against the possibility of significant cognitive divergence from a conceptual angle, and the second one turns on biology. Both arguments are framed in terms of “bad reasoning”, namely that humans *cannot* genuinely reason badly.

### 1.1 Is bad reasoning conceptually possible?

The assertion that it is simply incoherent to maintain that a person’s cognitive processing could deviate without limit from the standards of rationality can be implicitly found in Quine’s “Words and Object”, where he maintains that implications of seemingly bad reasoning are best understood as a problem of translation (Stich 1990: 29). So if certain natives of another language are said to sincerely accept as true certain sentences that are by our standards clearly wrong (‘p and not p’ for example), then it’s more likely a case of bad translation or linguistic divergence, rather than the respective natives’ “silliness” (Quine 1960: 58-9). Employing the widely shared *language of thought* paradigm which maintains

that mental processes are best viewed as manipulations and transformations of internalized, sentence-like representations, Stich stresses a strong parallelism between the project of translating a speaker's sincere assertions and the project of interpreting or intentionally characterizing a person's mental states. Viewed in this light, Quine's argument isn't a mere contribution to the theory of translation, but an attempt to set out some conditions constraining the intentional characterization of a speaker's beliefs. Namely that, beyond a certain point, silliness of belief is exceedingly unlikely. (Stich 1990: 33-4) According to Stich, Quine's view is also properly understood as an *a priori* conceptual truth (pp. 36-7). Taken *a posteriori*, it would be simply question begging: in order to determine whether someone is reasoning badly, we would have to interpret their words using some translation manual, while any acceptable translation manual is written by already eschewing the ascription of silly belief. However, as a conceptual truth, Quine's view seems to be an assertion that intentional describability requires some degree of rationality. This naturally raises the question of how much rationality is required and how it should manifest.

Stich criticizes three answers to this question. One defended by Daniel Dennett, which calls for "perfect rationality". A weaker view defended by Martin Hollis, which maintains that there are certain rational "bridgehead" inferences shared by all humans. And the weakest view of "minimal rationality", which only requires that an intentionally describable agent manifests some reasonable subset of possible rational inferences. (p. 39-43) The details of each account and their shortcomings are currently of little consequence, but what's important is that they all share a common implicit assumption. That in order for mental states to be characterizable in intentional terms, they must interact with each other and the environment in ways that are similar to our own (p. 48).

This requirement of similarity stems, according to Stich, from Quine's "principle of charity", which requires that when you are translating some speaker's language, you should make sure that most of their sincere assertions turn out to be true and most of their inferences turn out to be rational (p. 44). A more feasible reformulation of that principle, Grandy's "principle of humanity", maintains that we should opt for translations on which the pattern of relations among beliefs, desires, and the world be as similar to our own as possible. (pp. 44-5) Intentional description thus presumes that the cognition of the one we're describing can be categorized into states that play belief-like and desire-like roles, and in providing an intentional description of an agent in a sentence like 'S believes that p' we are in fact attributing to them a belief state similar to the one which would underlie our own assertion of

‘p’ were we just then to have sincerely uttered ‘p’. The relevant aspects and the requisite degree of similarity are themselves largely determined by context. (p. 49) Understood this way, the minimal rationality condition is thus a byproduct of the principle of humanity, which itself obtains because in providing intentional descriptions we characterize other people’s cognitive states by their similarity to our own (p. 50).

Stich uses an example to show that intentional description thus explicated is seriously limited (pp. 52-3). Taking mental processes to be certain manipulations and transformations of internalized, sentence-like representations, we can imagine a hypothetical scenario where we have a sequence of people, each of whom has brains that exploit the same class of formal structures, thus having syntactically identical “languages of thought”, and each manipulates these structures according to exactly the same rules. Then we assume those people to differ only in one respect, namely that for each person down the sequence they will have exactly one belief-sentence different from the previous person. Viewed holistically, each sequent pair of persons will be very similar to each other and any attempt to divide those people into two classes (those describable in intentional terms, and those not) would be ill-motivated. When adopting the perspective of the first person in the sequence, however, a division will be necessary. Those close to us will have intentionally characterizable states, and those very far away will not. The same picture could also be altered to keep the belief-sentences fixed, while incrementally changing the principles that govern how the belief-sentence inscriptions interact with one another. Those close to us will again be intentionally characterizable, and those far away will not.

The conceptual arguments against cognitive diversity revolve around fixing the limits of how badly people can reason to their intentional describability. This in turn rests on how similar those people are to us cognitively – in the belief tokens they manifest and the principles that govern how those tokens interact with one another. It distinguishes “real” beliefs from “belief-like” states and “real” inferences from other “inference-like” processes, all on the basis of their degree of similarity to our own. Such a stance is seen by Stich as following the “capricious contours of intentional describability”, one that will never be sufficiently clear-cut, and one that could be easily overcome by adopting a less chauvinistic epistemic vocabulary (p. 53). Given all the empirical evidence showing that people *do* in fact reason in different ways, this seems like just the thing to do.

## 1.2 Is bad reasoning biologically possible?

Another line of argument against cognitive pluralism is that systematic bad reasoning is simply biologically impossible. This is a view that rests on two premises: (1) evolution produces organisms with good approximations to optimally well-designed characteristics or systems, and (2) an optimally well-designed cognitive system is a rational cognitive system (Stich 1990: 56). For the first premise, the notion of a well-designed system will itself have to be explicated in terms of biological fitness, so an organism with such a system will be more likely to survive and reproduce successfully. From that, natural selection is supposed to choose the most fitness-enhancing genetically coded characteristic or system from all those available in the gene pool. Over a long period of time, we'll be left with a system that is about as well designed as it is possible to be, and the cognitive system we're left with will be very near optimal in the business of enhancing our biological fitness. (p. 57) From this and the second premise, we are to conclude that an optimally fitness enhancing cognitive system is a rational cognitive system (*ibid.*).

For anyone familiar with the complexities of evolution and the problems with classical Darwinism, some difficulties with such a will already become apparent. The most feasible way to defend such an evolutionary story would be to claim that a rational cognitive system can reliably produce true beliefs, whereas true beliefs themselves are fitness enhancing, and thus favored by natural selection. This view is more recently defended by Hilary Kornblith, according to whom 'knowledge' is properly understood as a natural kind – a capacity shared by all animals that embodies reliably produced true beliefs, which are instrumental in producing fitness enhancing behavior, and is brought about by the interplay of the informational requirements of the environment and Darwinian natural selection (Kornblith 2002: 61-2). A question to ask at this point is whether nature actually cares about truth, which doesn't seem to be the case. However, in order not to be dismissive, a more careful look into the matter is called for.

To see whether reliable cognitive systems are fitness enhancing, it is useful to make a heuristic distinction between *external* and *internal* fitness. External fitness pertains to how a gene leads a system's behavior in different environments and circumstances. Internal fitness concerns what goes on within, where a good genetic program will achieve its effects without making excessive demands on the memory, energy, and other resources of the organism. (Stich 1990: 60-1) How a less truth-reliable genetically coded cognitive system can still turn out to be more fitness enhancing, both internally and externally, and thus favored by natural



selection, can be shown by an example. Imagine two systems G1 and G2 so that G1 is the one more reliable in producing truths and avoiding falsehoods. If the higher reliability of G1 comes at the cost of greater resource requirement or time consumption, the less reliable system G2 may easily turn out to be favored by natural selection when considering the reducing marginal utility of information (p. 61). The link between internal fitness and reliable cognitive systems is thus easily severed. For external fitness, it will be useful to make a further distinction between *false positives* and *false negatives*. In the example of poisonous foods, a cognitive system arriving at a false positive would categorize a non-poisonous food as poisonous. For false negatives, it's the other way around – a poisonous food is inferred as non-poisonous. In terms of biological fitness, it's the false negatives that are more costly – in an environment with other possible food sources, unnecessarily avoiding a falsely-inferred poisonous food has little consequence, whereas failing to avoid a falsely-inferred non-poisonous food may mean illness or death. Thus a more fitness-enhancing system would be the one that's more risk averse, despite arriving at less truths and more falsehoods. (pp. 60-3)

Another hurdle for the evolutionary story is the question of whether evolution actually produces close approximations to optimally well-designed systems. This doesn't seem to be the case for two reasons: (1) the varied causes for evolution, and (2) natural selection's limited options (pp. 63-6).

The first problem is underlined by the fact that evolution happens for a variety of other reasons than natural selection, namely mutation (natural selection doesn't act upon mutations that are neither beneficial nor harmful), gene flow (i.e. migration), genetic drift (mainly in small populations) and non-random mating. As for natural selection's options, it is highly implausible to assume that mutations coding for optimal design will generally be available. Stich names three considerations that make this apparent: (1) the fact that technologically improved organisms (those employing prosthetic limbs, for example) will far surpass those not-improved in terms of fitness, (2) the persistence of pleiotropic linking where one gene has a positive effect on one system and a negative one on another shows that optimal mutations that would either replace the gene for its negative effect or introduce another gene that would suppress the harmful effect, aren't always readily available, and (3) the persistence of less-fit homozygous phenotypes in the population point to the fact that mutations offering an optimal phenotype in both heterozygous and homozygous forms aren't available either. (pp. 63-6)

Furthermore, even if genes producing optimal phenotypes were to be readily available, it is far from clear that this phenotype would spread throughout the population. In the case of

“cheating” genes, some genes in meiosis have the capacity to end up significantly overrepresented in sperm or eggs, and in a number of cases, cheating genes have been shown to produce harmful effects. (p. 66) Furthermore, Philip Kitcher has shown how natural selection can cause an optimally fit homozygous phenotype to disappear from the population (cf. Kitcher 1985: 215). The idea that natural selection must result in close approximations to optimally well-designed systems is therefore a highly dubious one.

The very claim that inferential strategies are the result of evolution is itself also questionable. For the evolutionary story to be viable it must be shown that (1) populations exhibit some variation in inferential strategies of a sort that affects the reproductive success of organisms in a systematic way and (2) the variance must be either directly or indirectly under genetic control. Showing that these conditions actually obtain isn't as trivial as some of the more naïve takes on biological evolution would imply. Consider, for example, a possibly parallel case of language. It is not obvious whether the capacity to speak some language is more fitness-enhancing than the capacity to speak another. Such capacities depend on the location of a person's birth and their general life-choices, and furthermore, if some language is indeed more fitness-enhancing than another, then it does not follow that this language will become a dominant one, as languages are known to rise and decline due to factors other than biological. In the same line, some inferential systems may come to be the dominant ones due to extra-biological reasons. To show that a dominant inferential system is optimal in promoting survival and reproductive success, we would need to know how such a system came to be the dominant one, to which end we don't have nearly enough evidence. (Stich 1990: 67-70)

In Kornblith's view, the complexities of evolution don't raise a serious problem for his naturalistic account of an animal's cognitive capacities – that the capacity to reliably attain true beliefs has developed by natural selection. While he is certainly aware of the issues just raised, his account is simple and robust enough to not be seriously endangered by them. He presents two considerations for that. Firstly, cognitive capacities as explained by Kornblith are no different from other complex organs of an animal: just as lungs are best explained to have been selected for their role in introducing oxygen to the blood, so too are cognitive capacities selected for their role in meeting the informational demands of the environment (Kornblith 2002: 58-9). Secondly, that natural selection may easily favor a less truth-conducive cognitive capacity, if it is more conducive to fitness, is a claim that proves far too much: a carnivore's teeth are selected for their function of ripping flesh, and the shape of a panda's thumb is selected for its ability to strip bamboo leaves from the stalk, but despite that,

the practice of carrying out those activities may at times conflict with the goal of survival (*ibid.*). It is worth noting here that Kornblith's account isn't specific to human knowledge: the evolutionarily developed capacity he talks about is common to all animals and doesn't require human rationality, social practices, or reflection. The capacity he draws out functions quite well as a robust natural category, or a natural kind. However, the crutch of his view seems to be its appeal to truth – that the naturally selected and fitness-enhancing cognitive systems are also truth-conducive. This is not something to simply overlook, as Kornblith is quite concerned about the common criticism towards naturalized epistemology – that it strips epistemology of its normative dimension (p. 137). For him, the normativity of his take on epistemology is saved by the fact that when it comes to evaluating various cognitive systems, their truth-conduciveness *always* plays a pre-eminent role, so that epistemic norms function as universal hypothetical imperatives (pp. 157-9).

What to take out of this chapter is that “bad reasoning” is possible both conceptually and biologically, and an instance of “bad reasoning” is typically understood as one that diverges significantly from our normal reasoning. This is where the normative dimension of epistemology becomes central. It must be shown why our normatively sanctioned strategies of reasoning are preferable to others. A common answer is that they are more conducive to truth. However, before turning to that issue, I will set out some general problems with analytic epistemology. These are problems that Kornblith also takes issue with, although it can be shown that the normative side of his theory fails to overcome them.

## **2. Evaluative-Concept Pluralism and its Bearing on “Analytic Epistemology”**

I borrow the term ‘analytic epistemology’ from Stich and use it to refer to any epistemological project which is guided by the method of conceptual analysis, such that it develops its theories by analyzing and explicating our epistemic concepts by comparing them with our relevant commonsense intuitions. Now, evaluative-concept pluralism is the empirically supported thesis that different people invoke different epistemic concepts and judgments when evaluating cognition. If these differences are significant and systematic, then there is reason to believe that the project of analytic epistemology is methodologically flawed, because if different groups of people intuitively employ concepts and judgments that diverge from those of an idiosyncratic group of professional academic philosophers, then it is not so obvious that we should favor the latter ones. The twofold nature of conceptual analysis reveals a twofold problem with analytic epistemology. First, it is focused on the *concepts* of various epistemic phenomena, rather than what underlies said concepts. Second, it draws upon pre-theoretic epistemic intuitions which aren’t necessarily trustworthy or uniform throughout the population.

### **2.1 Conceptual analysis**

The first charge is one that can be found in various projects of naturalized epistemology. Kornblith’s account of knowledge as a natural kind is based on the idea that the subject matter of epistemology should be knowledge itself, not our concept of knowledge (Kornblith 2002: 1). How one is to disentangle the concept from its underlying phenomenon is a matter more complicated, but in the naturalist view, the general idea is to endeavor for a more scientific approach. However, whether epistemology should constitute a subchapter of psychology (e.g., Quine 1969), or merely live up to the empirical standards of special sciences (e.g., Kitcher 1992), the concern raised is that our epistemic concepts as conceived may be inadequate in capturing the actual workings of cognition. An objection raised in favor of conceptual analysis might be that the epistemic concepts evoked in traditional epistemology reflect the *intuitive* categories and judgments we have pertaining to matters epistemic – whether ‘knowledge’, ‘justification’ or other concepts can ever find grounding in an extra-epistemological world is beside the point, for these are simply the very categories we use in carving our world through an epistemic lens. This brings to the second concern with analytic epistemology – its reliance on idiosyncratic epistemic intuitions.

## 2.2 Epistemic intuitions

What exactly differentiates intuition from other epistemic faculties is a question not easily answered. Some replies include the notions that intuitions are simply a certain kind of belief (e.g., Lewis 1983: p.x), that they are dispositions to believe (e.g., Inwagen 1997: 309), that they are *sui generis* occurrent propositional attitudes wholly separate from beliefs (e.g., Bealer 1998), or those that deny intuitions of having propositional content at all (Pust 2012 considers this possibility, without adhering to it himself). Despite this colorful picture, intuitions are routinely used as a methodological device in analytic epistemology, where the project of comparing certain epistemic concepts with one's intuitions is seen as a wholly uncontroversial one. However, the findings of evaluative concept pluralism point to the fact that the intuitions commonly employed in epistemology are largely idiosyncratic.

I won't argue that intuitions have no value, or that they haven't played a substantial role in building up some of the most influential theories in epistemology. This could be accounted for by the fact that our intuitions provide a relatively reliable way to work out the rough early details of a theory that tries to capture a part of our immediate world. In this case, our cognition in the way it is presented to us here in our culture. However, following Kornblith, I would question the relevance of such intuitions when faced with more complex contemporary theories (cf. Kornblith 2002: 8-20). I would argue that one should take a critical look on what those intuitions might actually be – not a well of unchangeable evidential data, but something much more malleable and shaped by a person's life-experience (or by their genetic buildup, which would be considerably harder to defend). This is suggested by empirical findings that show people to invoke different epistemic concepts and judgments when making evaluations about the various hypothetical scenarios routinely employed by epistemologists (Gettier cases, Knobe-effect cases, etc.). Divergence of intuitive judgments have been recorded between Westerners and East Asians (Weinberg *et al.* 2001: 443), people of low and high socioeconomic status (*ibid.*, 447-8), students who have taken no philosophy courses and those who have taken two or more (Nichols *et al.* 2003: 232), men and women (Buckwalter and Stich 2010), philosophers and non-philosophers (Knobe 2003a).

The last mentioned point of divergence is also implicit in routine philosophical practice – when professional philosophers solicit epistemic intuitions from the beginners in their courses, they are quick to dismiss the intuitions that don't fit the accepted central principles of current epistemology (Nagel 2007: 805). So when a beginner contends that in some cases knowledge can be false, they are invoking the *wrong* concept of knowledge. For Kornblith,

this is an indication that the various folk-psychological conceptions of phenomena like ‘knowledge’ don’t exhibit a necessary-and-sufficient conditions unity that knowledge itself (as a natural kind) might require (*ibid.*). That philosophers favor their own educated intuitions is, however, indicative of a kind of theory-contamination among philosophers already committed to some epistemological theory (Goldman and Pust 1993: 183). For this, Goldman and Pust call for favoring folk intuitions, while Kornblith suggests writing off intuitions as a historical chapter in the early days of a working theory. Goldman and Pust must then deal directly with the problems stemming from the evidence of evaluative-concept pluralism, whereas a naturalist view like Kornblith’s must come up with an all-encompassing theory of knowledge wholly independent of both *our* intuitions, and those that diverge, while overcoming the hurdles brought up in Chapter 1.

### **2.3 The task of epistemology**

One common objection to the kind of criticism just raised is that for an epistemological theory to be correct, it must leave our epistemic situation largely unchanged, so a theory which accurately captures our reflective epistemic judgments is better than one that does not (Bishop 2009: 118). In other words, the fact that analytic epistemologists work with the idiosyncratic intuitions that they share among themselves, is not a problem, as whatever is picked out by their intuitive concepts is just that which they are interested in. The fact that our epistemic vocabulary picks out different concepts in different cultures (or in other groups not properly trained in our terminology), is unproblematic as these other concepts aren’t part of our epistemological agenda. This naturally raises the question of what *is* the task of epistemology, properly understood.

A prominent defender of analytic epistemology, Ernest Sosa, questions the findings of evaluative-concept pluralism on four accounts: (1) it is unclear what the subjects questioned *actually* disagree on, (2) the choices offered to the subjects might be too limited, (3) the concepts of ‘knowledge’ and ‘justification’ and so forth might simply refer to different commodities in different cultures, none of them exclusively valuable, and (4) epistemology is not properly understood in lines of it making normative verdicts (Sosa 2009: 107-10).

The first two points, while possibly valid, call for further empirical research that doesn’t already assume cognitive monism. The latter two are specific to Sosa’s understanding of what epistemology should be about, a view not shared by all, and not all that easy to defend. For Stich, the idea that divergence in epistemic intuitions can be accounted for by the fact that the

epistemic concepts used in different cultures simply refer to different commodities, is not satisfying when one is in the business of evaluating and improving human cognition. When faced with universal claims of the kind ‘if S’s belief that  $p$  is an instance of knowledge, then, *ceteris paribus*, S ought to believe that  $p$ ’, then why should we favor the notion of ‘knowledge’ of high SES white Western males over any conflicting ones (Stich 2009: 234)? For Sosa, this is a nonstarter, as valuing one notion does not devalue the rest – one can value different senses of ‘knowledge’ just as one may value owning both money banks and river banks (Sosa 2009: 109). For Sosa then, there is no conflict. However, this kind of pluralism is difficult to maintain. Consider a parallel: if a Yanomamö intuitively judges that it is *morally permissible* to kill the men of another tribe and rape their women, and for me it is not, then we are invoking different concepts of moral permissibility, and it would be no easy task to show how the “commodities” invoked by our diverging notions are equally valuable (Stich 2009: 235). If we are interested in improving human cognition, then the cognitive categories that constitute the cognitive world of people in other cultures also call for careful studying. Also, there may be cases where a person values both the commodity of knowledge and some other commodity that’s picked out by ‘knowledge’ in another culture, but not being able to have both, must choose which is more valuable. Stich’s contention is that the kind of epistemology favored by Sosa is that of *ethno-epistemology*, where the epistemologist is concerned with *describing* a singular idiosyncratic account of knowledge, and at that, it fails to live up to the empirical standards of ethnography (p. 236). If, however, one wishes to come up with a theory of knowledge that is more than the explication of how they are used to employing their *concept* of ‘knowledge’, they should either do away with intuitions at this stage (e.g., Kornblith 2002) or show why the intuitions as favored by their community cut closer to what constitutes knowledge, which would be no easy task.

Kornblith’s naturalized account of epistemology seeks to overcome these limitations of conceptual analysis and analytic epistemology, and seems to mostly succeed. Study of concepts is exchanged for the study of *actual* knowledge. Pre-theoretic intuitions are understood as something of value only in the early stages of an investigation, which should at this stage, when we are studying knowledge as a proper natural kind, not be looked back to for verification. However, Kornblith’s robust take on epistemology seems to fall apart in its proposed normative dimension, where truth is placed center-stage in all cognitive evaluation. To see why this is the case, a careful look into the proposed value of truth in our cognition is necessary.

### 3. The Normative Function of Truth in Epistemology

In assessing whether a cognitive mechanism is a good one, the central criterion is generally thought to be its truth-conduciveness. Other important aspects may concern its effectiveness in terms of time and energy, accessibility to the agent, internal consistency, and so on, but only insofar as they pertain to the central aim of attaining true beliefs. Why we should look at a cognitive mechanism's truth-conduciveness in order to properly evaluate it, is explained by the assertion that true beliefs are simply valuable to have. Even if they might lack intrinsic value, then they are at least *instrumentally* valuable. This is an assertion that Stich explicitly argues against – in his view, true beliefs are neither intrinsically *nor* instrumentally all that valuable to have. In the following I will first sketch Stich's argument against the value of truth when it comes to the assessment of various cognitive mechanisms. Then I will look at Kornblith's arguments against Stich's view, and show them to be unsuccessful in establishing the normativity of truth in cognition.

#### 3.1 On the value of truth

For Stich, one of the underlying causes for the *prima facie* understanding of truth as something valuable to have, is a cluster of philosophical metaphors that have long ago become part of the fabric of commonsense wisdom. These metaphors liken beliefs to pictures or maps by which we steer in our world, and it is the veridical pictures and the accurate maps, i.e. those standing for true beliefs, that allow us to steer well. That these are simply metaphors is clear when one considers a belief like 'I believe that there are exactly four prime numbers between ten and twenty'. First, it is not clear how one is to picture such a belief in any way that corresponds to a picture or a map. Second, if one is a materialist, then they'd be hard-pressed to find the belief-pictures corresponding to various beliefs in the brain, which would be necessary to show the metaphor to be more than just that. (Stich 1990: 101-3)

A second issue concerns the nature of beliefs and what it is for them to be true. This problem is considerably more complicated, for both of these questions are greatly disputed and subject to radically different answers. Stich's approach is to first tackle one of the more plausible accounts – what he calls the causal/functional theory, before generalizing his criticism to other feasible views.



### 3.1.1 Causal/functional theory of the semantic properties of mental states

The causal/functional theory that Stich uses as a backdrop for his arguments assumes that (1) beliefs are real psychological states, and that (2) each instance, or token, of a belief is identical with some neurophysiological state, whereas the same belief type in different individuals isn't necessarily identical with the same neurophysiological state type. A peculiar property of beliefs is that unlike most other brain states, they are either true or false. This is explained by introducing some interpretation function which maps certain brain-state tokens onto semantic entities like propositions, content sentences, or specifications of truth conditions. A belief is thus true if and only if the proposition it is mapped to is true, or if and only if its truth conditions are obtained, and so on. (Stich 1990: 103-5) This requires that we define some interpretation function which maps mental states to things with semantic properties.

The first theoretical background for the causal/functional theory is Tarski's theory of truth. In Tarski's view, an axiomatic theory of language must specify a truth condition for each of the infinitely many well-formed sentences of the language. This is done by forming a meta-language sentence which specifies the conditions under which an object-language sentence of a certain structure is true. This approach has two limitations. First, there lacks a well motivated general account of what it is for a specification of truth conditions to be adequate or acceptable. Second, Tarski doesn't tell us enough about how to properly form the list of axioms, or base clauses of a recursive truth definition, which would specify the semantic properties of the language's non-compound predicates and names. (pp. 106-8)

This is where the Putnam-Kripke causal theory of reference comes into play. The causal theory provides us with a general account of what it is for an arbitrary name or predicate in an arbitrary language to refer to a particular object or a particular class of objects. In general, a name denotes an individual if and only if an appropriate causal-historical chain extends from the first use, or baptism, of the name, to its current use. The same story also applies to natural kind predicates. Since there are endless varieties of causal chains in the world linking all kinds of events in all kinds of ways, the *relevant* kind of causal chains are taken to be those that are compatible with our intuitions. If some proposed causal chain links our utterances to people or objects that intuition insists we are not talking about, then such a causal story is seen as defective. Furthermore, that the causal theory cuts closer to our intuitive understanding of reference is also generally seen as an advantage of such a theory over descriptive theories of reference. (pp. 108-9)

The last piece of the causal/functional theory must take it from an account of how sentences in a natural language get their truth conditions, to a semantic theory of mental states. For this, beliefs are seen as complex psychological states which, akin to sentences, can be viewed as built up out of simpler components. This allows us to associate belief tokens with well-formed formulas in some uninterpreted formal language, or as neurally-encoded inscriptions of the relevant well-formed formulas. Truth conditions for these inscriptions are specified by identifying certain of the words they are built up from as names or predicates, and the rest as connectives or quantifiers open to a Tarski-style theory of truth. Names and predicates are then paired with appropriate denotation or extension by tracing their causal ancestry, while connectives and quantifiers get their pairing from the pattern of interactions between names and predicates that they manifest. (pp. 109-10)

### **3.1.2 The limits and idiosyncrasies of the interpretation function**

The problem of the value of true beliefs is properly understood as something that turns on the interpretation function by which certain brain-states are mapped onto propositions, conditions of the world, or other such concepts. One feasible account on what such an interpretation function should look like can be found in the previously sketched causal/functional theory of reference. However, by drawing out the problems with such an account, the purported value of true beliefs can be shown to be not as obvious as commonly thought of. These problems are the limitedness and the idiosyncrasy of any such interpretation function.

That the interpretation function favored by the causal/functional theory is limited in its domain is due to two reasons. The first reason relates to the causal theory of reference, while the second one turns on the functional story about logical form. On the causal side, any account that even roughly fits commonsense intuition will only specify a limited domain of reference-fixing chains out of the empirically possible causal histories of mental words. Because of this, there will be many mental words that can end up in a speaker's mental lexicon that aren't tied to the world in the special way that the causal theory of reference requires. (Stich 1990: 110-11) For example, if the word 'Aristotle' enters my mental lexicon by me falsely overhearing a conversation about turtles, then by the causal account, my uttering of 'Aristotle' won't refer to anything, because, by common sense, this is not a proper way to be introduced to the name 'Aristotle'. Any mental sentences in which mental words like this occur in won't be paired with any appropriate truth conditions. As for limitedness stemming from the functional account of logical form, the issue lies with the fact that there are endlessly many syntactic structures for which there are no adequate truth theoretic

recursive clauses. Since an account of the required interpretation function is an explication of our intuitive judgments about contents or truth conditions, any structure that doesn't have an intuitively natural semantic interpretation will be left out of its domain. For example, a mental sentence is properly interpreted as a conjunction, or a counterfactual or any other logical construction when it interacts with other mental sentences in ways that mirror what logic intuitively permits. However, one can formally characterize an indefinite number of patterns of interaction among sentences or well-formed formulas that have *no* intuitively plausible semantics. (pp. 111-13)

The causal/functional interpretation function is also idiosyncratic. Even within the domain where it specifies interpretations, there can be lots of other functions that map mental states to the world in ways that aren't sanctioned by commonsense intuition. Let's look at the causal theory of reference. On the one hand, it specifies how a new name or a predicate is introduced into a language by the process of fixing its reference, or grounding it, to an object or class of objects. On the other, it specifies the process of social transmission whereby the name or predicate is passed from one speaker to another, while maintaining its original reference. In both instances, the legitimate groundings and transmissions are once again those that accord with intuition. However, both names and predicates get their groundings in markedly different ways: a baptismal process of a baby is quite unlike that of naming a war, and how the predicate 'gold' came to be paired with its extension is probably very different from how it was with 'kangaroo'. Similar diversity manifests in the processes of reference-preserving-transmissions. What ties all the acceptable causal chains together from grounding to transmission isn't some substantive property they all share, but rather that commonsense intuition counts them all as reference-fixing chains. The causal/functional theory is thus well suited for explicating the pre-theoretic views we have on how the words in a mental language are related to what they designate. However, it fails to capture the endless multitude of alternative views, some of which depart from commonsense in minor ways, and some in major. (pp. 114-15)

These alternative causal chains will link some or many mental words to objects or extensions different from those ascribed by commonsense intuition. Thus, they will characterize alternative notions of "reference", which we may call REFERENCE\*, REFERENCE\*\*, and so on. Now let's say that the name 'Jonah' in the Bible refers to a real, historical person about whom mythical tales were told as historical facts were forgotten. If we were to modify the intuitively sanctioned notion of reference by forming a kind of a hybrid between the causal

theory and descriptive theories, we could end up with the notion of REFERENCE\*, which is a word-world relation just like reference, except that it also requires that if the majority of the nontrivial descriptions a speaker associates with the name actually apply to no one, then the name is empty. In such a case, if there actually was a historical person about whom grand tales of surviving in the belly of a whale gradually developed, then 'Jonah' will refer to that person, whereas it will REFER\* to no one. Furthermore, by giving descriptions a somewhat different role in determining the reference of proper names, 'Jonah' might end up REFERRING\*\* to some long-forgotten ancient who actually *did* survive three days in the belly of a whale. Also, a notion of REFERENCE\*\*\* could be designed so that 'water' includes in its extension not just H<sub>2</sub>O, but also the watery stuff XYZ. While these alternatives turn on the grounding of a word, a different kind of alternatives can also be produced by varying the allowable reference-preserving patterns of social transmission. (pp. 115-16)

Each of the alternative word-world relations provides an alternative set of base clauses on which we can build alternative interpretation functions not favored by commonsense intuition. While according to the intuitively sanctioned account a belief 'Jonah was a Moabite' is properly mapped to a proposition that is true if and only if the historical person behind the legends was a Moabite. However, an interpretation function based on the REFERENCE\*\* relation would map the same belief to a proposition that is true if and only if a certain long-forgotten ancient who survived in the belly of the whale for three days was a Moabite. An intuitive interpretation would map the belief 'there is no water on the sun' to the proposition that there is no H<sub>2</sub>O on the sun, while one based on REFERENCE\*\*\* would map it to the proposition that there is no H<sub>2</sub>O or XYZ on the sun. (p. 116)

Now there is no reason to believe that the intuitive judgments underlying the standard readings of reference are in any way innate. The many findings of evaluative-concept pluralism point to the opposite – if there are significant and systematic differences in the intuitive concepts and judgments that people invoke in evaluating cognition then there is reason to believe that intuitions also diverge on other accounts. Perhaps the most poignant example here is the divergence in intuitively acceptable accounts of reference between Westerners and East Asians, recorded by Machery and colleagues, which shows that in Kripke's Gödel case (where Gödel actually stole the famous incompleteness theorem from his friend Schmidt) Westerners were more likely to favor the causal theory for grounding the reference of 'Gödel', while East Asians were more in favor of the descriptive theory (Machery *et al.* 2004). Interestingly, the same result wasn't found in the Jonah case, but this

could easily be due to reasons not pertaining directly to the intuitive judgments of the causal or descriptive models of reference – that East Asians did not favor the descriptive story of reference in fixing reference for ‘Jonah’ could be because according to the descriptive account, ‘Jonah’ will end up referring to no one, which could be seen as something to avoid for pragmatic reasons (*ibid.*).

For natural kind terms, the picture is even more colorful. Consider, for example, the New Guinean language Kâte. In this language there’s a word ‘*bilin*’, which denotes a certain kind of grass with strong roots which are said to hold the earth together during earthquakes. Now, when nails were first introduced to the native speakers of this language, the natives applied the same word to them – as well as to wire, iron rods, and everything else that served the purpose of holding things together. (Cassirer 1953: 41) Also consider the findings which show that while Westerners are likely to rely on rules of similarity relations in reasoning and categorization, the Chinese are more likely to group together objects for functional (e.g., pencil-notebook) or contextual (e.g., sky-sunshine) considerations, and Russians have been found to have a strong tendency to group together objects for their practical function (Henrich *et al.* 2010: 72). It is reasonable to assume that the idiosyncratic reliance on similarity relations for reasoning and categorization among Westerners has had a strong bearing on the development of what we call natural kind terms, some of which may not be as intuitively acceptable to people from other cultures. Furthermore, it is very likely that the intuitions sanctioning our acceptable causal chains of reference could be changed, and at that they could be changed for the better, as there is no reason to maintain that any alternative accounts of REFERENCE\*, or REFERENCE\*\*, and so on are in any way seriously flawed or less preferable. The regular, commonsense interpretation function stands out from the alternatives only by virtue of it being favored by local, contemporary intuition, while what actually constitutes that intuition, as well as what underlying psychological mechanisms give rise to it, is something we know very little about.

### **3.1.3 The value of true beliefs**

When it comes to deciding what we really value in our doxastic states, truth has a lot of competition. As noted, while an intuitively sanctioned account of the interpretation function might specify that a certain belief token is true if and only if there is no H<sub>2</sub>O on the sun, an alternative account based on REFERENCE\*\*\* would specify that the same belief token is true (or TRUE\*\*\*) if and only if there is no H<sub>2</sub>O or XYZ on the sun. Any given set of belief tokens one might have will then contain a certain percentage *n* of true beliefs. It will also

contain a certain percentage  $n^*$  of TRUE\* beliefs, a percentage  $n^{**}$  of TRUE\*\* beliefs, and so on. It will often be the case that by increasing our percentage of true beliefs, we will decrease our percentage of TRUE\* and TRUE\*\* beliefs. A question that presents itself at this point is that by valuing true beliefs, are we ready to give up TRUTH\* or TRUTH\*\*. (Stich 1990: 117-18)

True beliefs can be taken to be valuable either intrinsically or instrumentally. An intrinsically valuable commodity is good to have for its own sake, while an instrumentally valuable commodity is useful in order to achieve some other ends – which may themselves be of either intrinsic or instrumental value. Thus, if a person genuinely holds true beliefs to be intrinsically valuable, then little can be done (or should be done) to dissuade them from that. The best one can do is to show what those true beliefs actually are and what intrinsically valuing them entails. As for the instrumental value of true beliefs, it can be shown that they are not all that valuable when it comes to achieving other desired ends – alternatives could well be a lot more conducive in that regard.

Intrinsically valuing true beliefs can be shown to be a profoundly conservative thing to do for two reasons. Namely, that the intuitively sanctioned interpretation function which pairs beliefs with their truth conditions is both partial and idiosyncratic. The limited domain of the interpretation function entails that there will be a vast set of possible cognitive systems which admit to no semantic evaluation. In that domain there is neither truth nor falsity, and one placing intrinsic value on true beliefs may thus resist even considering, not to mention adopting, such systems, because they know in advance that these systems can't arrive at true beliefs. For them, the end products of cognition must be semantically interpretable, and to be semantically interpretable is to not depart too radically from current patterns of reasoning or from familiar ways of causally tying mental states to the world. However, something to consider here is that while likely much of the semantically uninterpretable domain is useless, or chaotic, it may very well contain systems that are highly conducive to happiness, biological fitness, or other desirable ends. (pp. 118-19)

Conservatism stemming from the idiosyncratic nature of the interpretation function is due to it relying on commonsense intuition. The causal/functional interpretation function singles out certain truth conditions from TRUTH\* CONDITIONS, TRUTH\*\* CONDITIONS, and other variations by looking at what the intuitively sanctioned causal chains of reference allow. Such an interpretation function is in no way a simple one – the vast variety of groundings and reference-preserving transmissions aren't unified by sharing some substantive property, but

simply by the fact that they accord with commonsense intuitions. However, those very intuitions are subject to change – they are likely culturally transmitted and acquired from the surrounding society. By intrinsically valuing true beliefs one is simply accepting the interpretation function that our culture has bequeathed to us and letting that function determine their basic epistemic value. (pp. 119-20) Furthermore, even *if* our intuitions were somehow genetically determined, then intrinsically valuing true beliefs would amount to simply accepting that which biology has ordained – and it is not clear why we should allow that.

To say that true beliefs are instrumentally valuable is to say that having them can lead to something else that is valued. Since the scope of all things that people may value is likely an endless one, showing that true beliefs have no instrumental value would be no easy task. However, what can be show is that the instrumental value of true beliefs isn't as obvious as generally thought. In assessing whether something is instrumentally valuable, it should be specified what it is evaluated against. If two different commodities are both conducive to attaining some other end, then the one more conducive to it, all things considered, will be the one favored. Therefore, to maintain that true beliefs are of great instrumental value, it isn't enough to simply show that they are more conducive to attaining desired ends than, say, false beliefs. While it may be that in many cases true beliefs are much more preferable to false ones (which is itself hardly a trivial claim), this says nothing for TRUE\*, or TRUE\*\* beliefs, or for any other categories of belief that are simply picked out by an interpretation function not sanctioned by intuition and tradition. It won't always be the case that TRUE\*\*...\* beliefs which aren't true will be false. Some mental states to which TRUTH\*\*...\* conditions are assigned may have no truth conditions at all – they are neither true nor false, while possibly being more conducive to attaining a desired goal than regular old true beliefs. (pp. 121-2)

At this point it could be argued that our current intuitive theory of interpretation is the result of a long evolutionary process through which many alternative mappings from mental states to propositions have been rejected, and the one we've ended up with presumably does a very good job of fostering survival and success. However, as noted in **1.1.2**, 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. Furthermore, even if using the intuitively sanctioned interpretation function were to be especially conducive to survival, this wouldn't entail that having true beliefs is more instrumentally valuable than having TRUE\*\*...\* beliefs. It would also have to

be shown that the intuitively sanctioned interpretation function is conducive to survival *because* it fosters believing the truth. (pp. 122)

A final consideration against the great instrumental value of true beliefs is the fact that in many cases we already know that having true beliefs is *not* the best way to achieve our more fundamental goals. (pp. 122-3) If my true belief about the time of departure of my plane leads me to get on the plane on time and it happens to crash, then having had a false belief in its stead would have been more conducive to my survival. The same is also true for a multitude of cases where having false beliefs is more optimal in the pursuit of happiness, or pleasure, or self-fulfillment, or other commodities I might value. For all these cases, alternative belief-to-world mappings that foster me with TRUE\*\*...\* beliefs could well be favorable to those that get at regular, intuitively-arrived-at truths.

While Stich assumed that the causal/functional account of the interpretation function is on the right track, his criticism of the value of truth easily carries over to other plausible accounts. Whatever interpretation function one comes up with, showing that true beliefs are of great instrumental value will still require that they'd be shown to be more conducive to our goals than TRUE\*\*...\* beliefs. And given any plausible account of the interpretation function, there will be instances of alternative cognitive mechanisms that this function does not cover, and favoring it will still end up amounting to unnecessary conservatism. Also, if we were presented with a more theoretically elegant and less messy account of the interpretation function than the causal/functional one, there's nothing to say that the people adhering to the messier account won't lead happier, healthier and more rewarding lives than those whose cognitive systems tend to produce true beliefs. (pp. 124-7)

### **3.2 Is truth a pre-eminent epistemic value?**

By showing that the value of true beliefs is far from obvious, Stich proposes that in order to evaluate one system of cognitive mechanisms as preferable to another, we should consider if by using it we are more likely to achieve those things that we intrinsically value (Stich 1990: 24). More pointedly, the system to be preferred is the one that would be most likely to achieve those things that are intrinsically valued *by the person* whose interests are relevant to the purposes of the evaluation. The person relevant will most of the time be the one who might be using the system. (pp. 131-2) Interestingly, this pragmatic account of cognitive evaluation is challenged by Kornblith on the grounds that it fails precisely where a pragmatic account should be the strongest – in allowing us to act so as to serve whatever interests we may care



about (Kornblith: 2002: 156). For Kornblith, truth *always* plays a pre-eminent role in cognitive evaluation, regardless of what the people employing those systems are trying to achieve. In his view, whatever cognitive system one is to adopt, it is central that the chosen system will actually, in truth, get them to whatever it is they are trying to achieve. In a way, valuing the truth-conduciveness of our cognitive mechanisms is, for him, pragmatically preferable to denying this pre-eminent role played by truth.

### 3.2.1 Cognitive evaluation vs. epistemic evaluation

Kornblith contends that Stich's view seems to lead to a strange situation where a person's judgment of a belief as being unjustified provides a conclusive reason for rejecting it. This is because on Stich's view epistemic evaluation already takes account of everything an agent values, so a judgment that a belief is epistemically unacceptable is no different from the judgment that all things considered, it is unacceptable. (Kornblith 2002: 151) This reading misses the point for two reasons. First, it assumes that to deem a belief as unjustified is something of a trivial task. In Stich's scheme, it is clearly not. Stich's account is highly relativistic, taking in account the agent of cognition, her context, and what she intrinsically values. To take in one glance all these relevant details is beyond our cognitive capacities, so no conclusive verdict on the justification of a given belief is likely even possible. Kornblith takes Stich's relativistic and context-sensitive strategy of cognitive evaluation and turns it on its head by assuming that the evaluator has actual access to the whole story. However, let's recall that Stich's account is a *pragmatic* one, not one that sets constraints on the cognitive evaluations of omniscient beings. Moreover, Stich is in fact very careful with assigning judgments to instances of reasoning that may seem bad, or unjustified (cf. Stich 1990: 149-58). A second problem with Kornblith's contention is that, for Stich, there are no intrinsic epistemic virtues and thus no special cognitive or epistemic values, but just *values* as such (Bishop 2009: 120). That a belief is unjustified, on any account, might then mean nothing to the believer in case they simply place no value on justification.

On the same theme, Kornblith criticizes Stich's view for it raising epistemic evaluation to an all-encompassing status. If having a certain belief is epistemically ill-advised, then one might still have good reason, all things considered, to come to that belief. In Kornblith's understanding, this is not sanctioned by Stich's view. (Kornblith 2002: 151-2) This is clearly not the case: Kornblith assumes wrongly that cognitive evaluation for Stich means evaluation done in purely epistemic terms. What Stich is saying is quite the opposite: if having a certain belief is ill-advised for purely epistemic reasons as normally understood, like its lack of

justification, or it leading to falsehoods, while being conducive to that which the agent *actually* values, then for Stich, one should by all means retain that belief. In other words, Stich does *not* identify epistemic propriety with any proposed epistemic concepts. His account of cognitive evaluation easily sanctions evaluations done on moral, aesthetic, or any other grounds.

Kornblith also raises the question of what exactly makes Stich's account of epistemic evaluation epistemic (p. 152). Herein seems to lay the misconception Kornblith has about Stich's account: properly understood it is in fact *not* an account of *epistemic* evaluation. It seeks to offer a strategy that's good for evaluating cognitive states and processes in general. If such processes are at times best evaluated in terms other than purely epistemic, then evaluations from other perspectives are completely acceptable and even encouraged. Granted, this is a point of concern for Kornblith, as he points out that he is interested in finding out what room can be made for genuine *epistemic* evaluation. However, considering his contention that knowledge is a natural kind, this seems far too modest of him. If knowledge is truly a natural capacity shared by all animals that has developed by natural selection, and which produces fitness-enhancing behavior, then it is clearly not just the purely epistemic considerations of cognitive evaluation that we should be concerned with.

First, we should consider what constitutes fitness-enhancing cognitive behavior. For Kornblith, the required fitness-enhancing behavior is that which stems from the animal attaining *true beliefs* about their environment, and a fitness enhancing cognitive mechanism is thus one that reliably produces true beliefs (given that it is capable of getting at such beliefs at a reasonably low cost of time and resources). As shown in **1.1.2**, true beliefs are not always fitness-enhancing. Kornblith surely understands this, but in his mind they do a relatively good job at that, and in the end, truth-reliable cognitive mechanisms will still turn out to be more fitness-enhancing than those that get at substantially more false beliefs (pp. 63-7). However, as noted in **3.1.3**, this is not enough to show that the naturally selected truth-reliable system is in any way the most fitness-enhancing of all possible systems. This is evident from the fact that natural selection cannot produce optimally (or even near optimally) well-designed systems. In order to establish that any naturally selected truth-reliable cognitive system is optimal to an animal's fitness, it would also have to be shown that it is more fitness-enhancing than some alternative TRUTH\*-reliable system. More fitness-enhancing alternatives to the regular truth-reliable systems can easily be constructed. Furthermore, since optimal phenotypes aren't always readily available in the gene pool, the truth-reliable cognitive

system that we've ended up with is very likely *not* the most optimal to biological fitness, even if we suppose the regular notion of truth to be wholly unproblematic – given different circumstances we could have ended up with a system that more reliably gets at true (not TRUE\*) beliefs. If it is the *normativity* of epistemology we are concerned with, i.e. if we are in the business of evaluating cognition and making prescriptions about which cognitive systems an agent should employ, then such alternatives (both the truth-reliable and TRUTH\*-reliable ones) shouldn't be ignored. The strong link between truth and biological fitness that Kornblith argues for seems to be very fickle. This result is not surprising because, for one, nature doesn't seem to care about truth, and furthermore, even if it were the case that truth-reliable cognitive systems are generally fitness-enhancing, it's far from clear that they are fitness-enhancing *because* they are truth-reliable. However, if the normative dimension of epistemology were to follow naturally from its descriptive dimension, which Kornblith thinks is the case (pp. 160-1), then the story about natural selection and biological fitness shouldn't be simply set aside in favor of truth either, and if (one of the) relevant criteria for evaluating cognition is taken to be its conduciveness to biological fitness (which is not neatly linkable to truth-conduciveness), then such evaluation is clearly not subject to considerations purely epistemic. None of this goes to say that Kornblith's account is, without a doubt, deeply off-track, but it does suggest that he cannot so easily dismiss Stich's view.

The preceding assumes that biological fitness, or for that matter, natural selection, is a normatively relevant concern when it comes to cognitive evaluation. This is not quite obvious: even *if* natural selection could be shown to produce optimally fitness-enhancing cognitive systems, and even *if* such systems were fitness-enhancing *because* they are truth-conducive, then this by its own will still fail to establish that such systems are normatively preferable. Kornblith himself notes that knowledge as conceived by him is conducive to fitness, and as such it is instrumental in producing the kind of behavior that satisfies a creature's biological needs (p. 160). Now it is clear that we have needs other than biological, and it is also clear that for many of us a number of our biological needs are things that we would rather not fulfill. So even if truth-conduciveness were to be preeminent in evaluating whether a cognitive system is fitness-enhancing, there is no cause to extend this to all cognitive evaluation. In case the agent in question simply places no inherent value to biological fitness, it would still not be clear that all cognitive evaluation should turn on purely epistemic notions.

### 3.2.2 How a more truth-conducive cognitive system is worse for attaining the things that an agent values

Kornblith also notes a deeper flaw in Stich's account: namely that it fails when it comes to any actual cost-benefit calculation of one's cognitive systems. A simple example of that would be the case where a person is deciding between two toasters in a store, where the toaster that has the highest expected value in terms of that which the person is trying to attain is the one that the person should buy. In doing this, the person makes use of some cognitive system. Now for Kornblith, such a system will *always* have to get at the truth of the matter. Imagine that all the person in question cares about is their happiness, and thus the best choice of a toaster would be the one that will ultimately make them happier. To know which of the toasters will lead to that, the cognitive system they employ will have to inform them about the *actual* consequences of choosing either of the toasters. In other words, the appropriate cognitive system will have to produce true beliefs about the consequences of either decision. (Kornblith 2002: 153-5) In Kornblith's view, the system that produces more happiness-conducive beliefs and less true ones will then fail to actually be conducive to a person's happiness, as it will only tell the person what they would be happiest to believe that the consequences of their decisions are, and therefore fail to tell the person which decision would *actually* make them the happiest. (p. 155) A more successful system would be the one that reliably produces *true* beliefs about the happiness-conduciveness of their behavior, or rather, of their already held beliefs.

Kornblith's move from first-order beliefs to second-order, though not explicit, is quite noteworthy: it places the normative dimension of his view on beliefs *about* beliefs. While one may easily favor a cognitive system that leads them to attain the variety of things they value, they will also want to know if such a system will *in truth* get them there. This is why, for Kornblith, truth is pre-eminent in all cognitive evaluation. However, as a criticism of Stich, this is quite unfair. Before turning to the question of why this is so, I will first show how truth-conduciveness is not a necessary component of a successful cognitive system. For this, I will use Kornblith's example of deciding between buying two toasters, but first I will introduce a pair of relevant psychological phenomena: confirmation bias and belief perseverance.

Confirmation bias is a psychological phenomenon that causes a person to favor any information that confirms their beliefs or hypotheses (Plous 1993:233). This causes a person to both select their information in a way that reinforces their beliefs, and to interpret any

information pertaining to their beliefs in a biased way. Another phenomenon closely tied to it is belief perseverance, which is a tendency to cling to one's initial beliefs even after receiving contradicting or disconfirming evidence against them (Anderson 2007: 109). Interestingly, this also seems to persist even if the disconfirming evidence is by all accounts completely acceptable to the person, and thus not filtered out by confirmation bias (cf. Ross *et al.* 1975: 880-92). An example of how the named psychological phenomena may affect people's decisions can be seen in how the serial positioning of objects affects the judgments of consumers. Nisbett and Wilson devised two experiments where they asked shopgoers to rate the quality of different articles of clothing (four nightgowns in the first experiment, and four identical nylon pantyhose in the other). In both experiments, the clothes were lined up from left to right, and in both experiments there was a pronounced position effect in people's evaluations, such that the right-most garments were heavily preferred to the left-most garments. Upon questioning whether the positioning of the clothes had any effect on their choice, nearly all subjects denied and even vehemently rejected such an influence. (Nisbett and Ross 1980: 207)

Now let's consider how these psychological tendencies might bear on my decision of which toaster to buy, and how a happiness-conducive cognitive system will do a better job of leading me to happiness than a truth-conducive (or a happiness-and-truth-conducive) one. Suppose that I have some pretty specific criteria for the technical details that I want my toaster to have. Suppose further that I am on a limited budget, while this is a sad fact that I would rather not be reminded of. Suppose also that all I really value is my happiness, but that I am also a critically-minded person and easily swayed by scientific truths about human psychology. For the last assumption, let's say that I have read about the serial positioning effect. I am presented with two toasters and the one on the left matches closer with my technical criteria. However, the one on the right is considerably cheaper, and due to the positioning effect, there is something inexplicably charming about it. Since the cognitive system I'm employing is not very truth-conducive, I will fail to realize that I am under the influence of this effect. Being also on a limited budget, I will subsequently distort and reinterpret the technical data of each and happily buy the cheaper one. Once home, I will also reinforce my belief that it was the right decision by reading favorable reviews of my new toaster (while ignoring unfavorable ones), and since I bought the cheaper one, I will have money left to buy quality bread to use with it. Had I employed a more truth-conducive cognitive system, I would have acknowledged that I am influenced by the positioning effect, and I would have stopped myself

from committing to confirmation bias. I wouldn't have been as happy with my cheaper toaster, because I'd have been painfully aware that it didn't match my technical criteria. Neither would I have been as happy with the more expensive one – I would be saddened by the fact that I couldn't find a perfect toaster that *also* meets my budget. In both cases I would have had to face the sad truth that I simply can't comfortably afford a good toaster. If all I value is my happiness, it's difficult to see why the more truth-conducive cognitive system would be preferable to me.

### **3.2.3 The pre-eminence of truth in second-order beliefs**

The pre-eminent normative role of truth in cognitive evaluation, according to Kornblith's account, seems to manifest in second-order beliefs. That is, beliefs about beliefs. For any cognitive system designed to produce in an agent beliefs that are conducive to attaining their various goals, we want to know whether the beliefs therein produced are *actually* conducive to those goals. Therefore, the cognitive system we employ to evaluate various cognitive systems, will itself have to produce *true* beliefs about the systems under evaluation. A way to explain this better is to adopt the perspective of a third-person evaluator looking at my cognition in the toaster example. From this perspective, it will be clear that the less truth-conducive cognitive system was indeed preferable to me for attaining happiness. However, this third-person evaluation is done in terms of truth. In evaluating what kind of a cognitive strategy best suits my desires, the evaluator will have to attain true beliefs about my cognition, about what I value, and the situation. If the third-person evaluator was herself using some non-truth-conducive cognitive system that doesn't truthfully capture all the details relevant to the story, then the cognitive system she would recommend to me, might not actually lead me to happiness.

If this is indeed Kornblith's criticism of Stich's view, which it seems to be, then it is quite unfair. At no point has Stich commended, that we, as epistemologists, should forgo our pursuit of knowledge on matters epistemic and collectively employ some alternative cognitive systems which will instead make us feel happy about whatever we are doing. Stich's account of cognitive evaluation is very clear about its contingent character, and very careful about understanding the agent, their intrinsic values, and the context of evaluation. In every step of this, it is necessary to draw valid inferences about the relevant data, and in doing that, we of course appeal to some notion of truth. However, this does *not* establish that truth is a pre-eminent normative value in all cognition. Cognitive systems are still properly evaluated in virtue of that which the agent employing the system is trying to attain, or what they value.

Stich's criticism of the value of truth concerns strictly the idea that truth is of central normative importance in all cognition – it is not an all-encompassing attack on epistemologists' endeavor to develop a theory that is, in some sense, the one that matches the truth of the matter.

Furthermore, this is not necessarily the case. An epistemologist not concerned with the truthfulness of their theory may put no stock in truth even in their own evaluations of any cognitive systems. Such an epistemologist may once again only value their own happiness. In evaluating any cognitive system, they wouldn't have to concern themselves with whether the cognitive system under evaluation is *actually* conducive to producing the things that the agent in question values. Such an epistemologist would be content with recommending any cognitive system that they are *happy to believe* will cause an agent to attain the things that the agent values. If their own happiness is the only relevant concern, then there is nothing wrong with this. There is also nothing wrong with an epistemologist trying to develop a theory with some satisfactory degree of truth-likeness, to value truth in their work. However, truth-conduciveness is not a pre-eminent epistemic value that makes or brakes any cognitive system.

### **3.2.4 How Kornblith fails to overcome Stich's attack on the value of truth**

A final point to make is that my use of the word 'truth' in the preceding was a cautious one, and it may just as easily stand for some anti-realist account of truth, or for TRUTH\*, or any other alternative. Kornblith seems to favor some form of scientific realism (cf. Kornblith 2002: 157), but in his criticism of Stich, he doesn't get any more specific, and neither does he consider that whatever account of truth he favors may itself be problematic for any of the reasons that Stich mentions. Never does he consider *why* true beliefs might not be as valuable as usually thought of. It is not because we have simply other things we value besides truth, but because the very notion of truth itself, when applied to our beliefs via some interpretation function, is partial and idiosyncratic. That is because the acceptable linking of our mental states to something with semantic properties is derived from our commonsense intuitions, which is a practice that Kornblith himself is also critical of.

It is entirely in line with Kornblith's view to say that the pre-theoretic intuitions underlying the commonsense understanding of truth were instrumental in the earlier stages of epistemology. They offered a quick and mostly reliable way to distinguish true beliefs from false ones. However, Kornblith doesn't take the next step. He completely disregards the

possibility that the commonsense truth-conduciveness to which he gives a pre-eminent place in all cognitive evaluation, might not hold up as a proper theoretical category. Yet this is exactly the kind of criticism he mounts against the more traditional projects of epistemology. That his understanding of truth-conduciveness might not exhibit a good theoretical unity, can be seen by once again considering cognitive pluralism and evaluative-concept pluralism. The way people go about their cognition is significantly diverse, and as we've seen, this diversity cannot be easily chalked up to linguistic divergence or other similar considerations. If we are to do normative work in epistemology, we must find some way to properly evaluate these diverging cognitive strategies. Stich and Kornblith agree that the agent's goals should play some role here, but for Kornblith, truth will always be pre-eminent. From evaluative-concept pluralism, we know that people's epistemic intuitions also diverge significantly. Any intuitively sanctioned interpretation function mapping brain-states to the world is therefore going to be partial and idiosyncratic. However, epistemic norms according to Kornblith are *universal* hypothetical imperatives. Why should cognitive systems that produce more true beliefs, rather than TRUE\*\*...\* beliefs, then be evaluated higher? Kornblith might say that relevant to the 'knowledge' and 'cognitive systems' he is talking about, it is the regular true beliefs that matter. But aren't we then back to Sosa-style conceptual analysis? In addition to being at odds with Kornblith's own criticism of such an approach, this would be difficult to defend as a normative project of cognitive evaluation, as the very thing under evaluation – cognition – doesn't seem to manifest uniformly throughout the population.



## Conclusion

The empirical findings of cognitive and evaluative-concept pluralism show respectively that people go about the business of cognition in significantly varied ways and that people employ different epistemic concepts and judgments when evaluating cognition. Cognitive pluralism thus raises the question of how we are to evaluate different cognitive systems, and evaluative-concept pluralism undermines one of the common answers to this question: conceptual analysis.

For the issue of cognitive evaluation to have any real substance, it must be shown that there actually is some significant and systematic diversity in the ways that people go about their cognition. Conceptual arguments against the possibility of cognitive pluralism, properly understood, follow the capricious contours of intentional describability, which itself turns on the described agent's similarity to us. This amounts to unmotivated epistemic chauvinism. Evolutionary arguments on the same theme, however, greatly overestimate the power of natural selection and its control over the development of cognition. These considerations, coupled with relevant empirical evidence, make a good case for cognitive pluralism.

Faced with cognitive pluralism, any epistemological theory hoping to retain its normativity will have to offer a good account of how these different cognitive strategies should be evaluated. The answer to this question from the field of analytic epistemology insists that we should look at how well they fare as explications of our pre-theoretic intuitions. Evaluative-concept pluralism shows that the intuitions commonly adhered to in epistemology, are highly idiosyncratic, and this raises the question of why we should favor the theories built up from the basis of some certain idiosyncratic intuitions in the face of all the alternatives. If we are interested in the normative project of evaluating cognition in general, we shouldn't simply restrict the domain of our theories to the concepts evoked in a philosophy class.

Hilary Kornblith's naturalized account of epistemology seems to overcome these problems, while avoiding the charge often mounted against naturalized epistemology – that it strips epistemology of its normative character. In Kornblith's view, normativity in epistemology is tied to the truth-conduciveness of our cognitive systems. Stephen Stich shows this to be unmotivated: when viewed clearly, true beliefs fail to have nearly as much value as usually thought of. A more viable theory of cognitive evaluation is thus a pragmatic one, which places center-stage the things that an agent actually intrinsically values.

Kornblith's criticism of Stich's view shows Kornblith to place the normativity of his view on second-order beliefs. While his contention that a truth-reliable cognitive system is always conducive to attaining whatever the agent values is easily shown to not hold, the value of truth can be somewhat saved when considering second-order beliefs about the beliefs of the agent under evaluation. However, that too is only so if truthfulness is something that the evaluator wishes to attain. In all this, Kornblith seems to ignore the bulk of Stich's attack on the value of true beliefs, which makes his take on the normativity of truth in epistemology all the more difficult to save in the face of any alternative TRUTH\*\*...\*-linked accounts.

## References

- Anderson, C. A. (2007) "Belief perseverance," in Baumeister and Vohs (eds.), *Encyclopedia of Social Psychology*, Thousand Oaks: Sage. pp. 109-10.
- Bealer, G. (1998) "Intuition and the Autonomy of Philosophy," in DePaul and Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Oxford: Rowman & Littlefield. pp. 201-39.
- Bishop, M. (2009) "Reflections on Cognitive and Epistemic Diversity: Can a Stich in Time Save Quine," in D. Murphy and M. Bishop (eds.), *Stich and His Critics*, Chichester: Wiley-Blackwell. pp. 113-36.
- Buckwalter, W. and Stich, S. (2010) *Gender and Philosophical Intuition*. Available at URL (used May 2014) [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1683066](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1683066)
- Cassirer, E. (1953) *Language and Myth*. New York: Dover.
- Devitt, M. (2012) "Whither Experimental Semantics?" *Theoria*, **73**: 5-36.
- Goldman, A. and Pust, J. (1998) "Philosophical Theory and Intuitional Evidence," in DePaul and Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role In Philosophical Inquiry*, Oxford: Rowman & Littlefield. pp. 179-97.
- Henrich, J., Heine, S. J., Norenzayan, A. (2010) "The weirdest people in the world?" *Behavioral and Brain Sciences*, **33**: 61-135.
- Kitcher, P. (1985) *Vaulting Ambition*, Cambridge, Massachusetts: The MIT Press, Bradford.
- Kitcher, P. (1992) "The Naturalists Return," *Philosophical Review*, **101**: 53-114.
- Knobe, J. (2003a) "Intentional Action and Side Effects in Ordinary Language." *Analysis*, **63**: 190-3.
- Kornblith, H. (2002) *Knowledge and Its Place in Nature*, Oxford: Clarendon Press.
- Lewis, D. (1983) "Truth in Fiction." *Philosophical Papers: Volume I*. New York: Oxford University Press.
- Machery, E., Mallon, R., Nichols, S., Stich, S. P. (2004) "Semantics, cross-cultural style." *Cognition*, **92**: B1-B12.
- Margolis, E. [2011] *Concepts*. URL (used May 2014) <http://plato.stanford.edu/entries/concepts/>
- Nagel, J. (2007) "Epistemic Intuitions." *Philosophy Compass*, **2**: 792-819.
- Nichols, S., Stich, S. and Weinberg, J. (2003) "Meta-Skepticism: Meditations on Ethno-Epistemology," in S. Luper (ed.), *The Sceptics: Contemporary Essays*, Aldershot: Ashgate Publishing. pp. 227-47.
- Nisbett, R. (2003) *The Geography of Thought: How Asians and Westerners Think Differently . . . And Why*, New York: The Free Press.
- Nisbett, R. and Peng, I. (1999) "Culture, Dialectics, and Reasoning about Contradiction." *American Psychologist*, **54**: 741-54.

## Vaus, The Normativity of Truth in Cognitive Evaluation

- Nisbett, R., Peng, I., Choi, I., and Norenzayan, A. (2001) "Culture and Systems of Thought: Holistic vs. Analytic Cognition." *Psychological Review*, **108**: 291-310.
- Nisbett, R. and Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*, New Jersey: Prentice-Hall.
- Plous, S. (1993) *The Psychology of Judgment and Decision Making*, New York: McGraw-Hill.
- Pust, J. [2012] *Intuition*. URL (used May 2014) <http://plato.stanford.edu/entries/intuition/>
- Quine, W. V. O. (1960) *Word and Object*, Cambridge, Massachusetts: The MIT Press.
- Quine, W. V. O. (1969) "Epistemology Naturalized," in *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- Ross, L., Lepper, M. and Hubbard, M. (1975) "Perseverance in Self Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." *Journal of Personality and Social Psychology*, **32**: 880-92.
- Sosa, E. (2009) "A Defense of the Use of Intuitions in Philosophy," in D. Murphy and M. Bishop (eds.), *Stich and His Critics*, Chichester: Wiley-Blackwell. pp. 101-12.
- Stich, S. P. (1990) *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Massachusetts: Bradford Books.
- Stich, S. P. (2009) "Replies: Reply to Sosa," in D. Murphy and M. Bishop (eds.), *Stich and His Critics*, Chichester: Wiley-Blackwell. pp. 228-36.
- Van Inwagen, P. (1997) Materialism and the Psychological-Continuity Account of Personal Identity. *Philosophical Perspectives*, **11**: 305-19.
- Weinberg, J., Nichols, S. and Stich, S. (2001). "Normativity and Epistemic Intuitions." *Philosophical Topics*, **29**: 429-60.

## **Tõe normatiivne roll kognitiivsete süsteemide hindamisel**

Kognitiivne paljusus on tees, mille kohaselt leidub inimeste mõtlemisviiside vahel olulisi ja süstemaatilisi erinevusi. See tees on ühtlasi saanud ka palju empiirilist kinnitust. Vastukaaluks on sellele aga esitatud argumente, mille kohaselt on kognitiivne paljusus kas kontseptuaalselt või bioloogiliselt võimatu.

Kontseptuaalsed argumendid eeldavad tugevat sidet agendi ratsionaalsuse ja tema intentsionaalse kirjeldatavuse vahel, ning Stephen Stich on näidanud, kuidas see viimane on sellistes argumentides mõtestatud läbi agendi sarnasuse iseendale. See osutub aga põhjendamatuks ning analüütiliselt väärtusetuks määratluseks, mida on kerge ületada võttes kasutusele lihtsalt vähem šovinistlik sõnavara. Kui kognitiivne paljusus on aga vastuvõetav, siis tõstatab see meile olulise küsimuse sellest, kuidas neid erinevaid mõtlemisviise või kognitiivseid süsteeme hinnata. Olukord on veel raskendatud tänu tõigale, et empiirilist kinnitust on leidnud ka hannanguliste-kontseptsioonide paljusus, mis ütleb, et kognitiivsete süsteemide hindamisel rakendavad erinevad inimesed oluliselt ja süstemaatiliselt erinevaid episteemilisi kontseptsioone ja hinnanguid. See seab kahtluse alla analüütilise epistemoloogia metodoloogilise võtte kontrollida enda teooriaid läbi selle, kuivõrd hästi nad vastavad meie eelteoreetilistele intuitsioonidele.

Kuna inimeste episteemilised intuitsioonid ei ole ühtsed, siis seab nende järjepidav kasutamine teatud epistemoloogilistes projektides kahtluse alla nende projektide normatiivse külje. Stich nimetab selliseid projekte etno-epistemoloogiaks, ning ka Hilary Kornblith on epistemoloogia kui *kontseptuaalse* analüüsi suhtes kriitiline – tema meelest peaks uurimus teadmises uurima teadmist ennast, mitte selle kontseptsiooni. Ka Kornblithi jaoks on epistemoloogia normatiivsus tähtis küsimus, ning tema jaoks mängib siinpuhul keskset rolli tõde.

Seisukoht, et kognitiivsete süsteemide hindamisel on esmaseks kriteeriumiks nende tõesus, on vägagi levinud, ning üldjuhul ei eitada tõe väärtust meie kognitsioonis. Stich vaidleb aga, et tõesed uskumused ei ole miski, mida oleks eriliselt väärtuslik omada. Seejuures pole nad väärtuslikud ei iseenesest ega ka instrumentaalselt. Seda seetõttu, et meie tavamõistlik arusaam tõesest, rakendatuna meie uskumustele, toetub kultuuritundlikele eelteoreetilistele intuitsioonidele. Omistamiseks meie mentaalsetele seisunditele (nt uskumustele) semantilisi omadusi, peame me nad kaardistama vastavatele semantiliselt kirjeldatavatele propositsioonidele, tõekriteeriumitele jms. Igasugune tõlgendusfunktsioon, mis vastab

tavamõistuslikele intuitsioonidele, on aga paratamatult piiratud ja idiosünkraatiline. Tavapärased tõesed uskumused, mida me väärtustame, on seega erilise väärtusega vaid seetõttu, et nad vastavad tavamõistusele ja kinnitavad traditsiooni. Erinevate intuitsioonide või üldse teiste aluste baasilt loodud tõlgendusfunktsioonid võiksid meile aga pakkuda pragmaatiliselt väärtuslikemaid kognitiivseid süsteeme. Kuna ei ole lihtne näidata, miks tõde on väärtuslikum kui mingi alternatiivne TÕDE\*, siis peaks kognitiivsete süsteemide hindamisel keskenduma sellele, mida agent iseenesest väärtustab ja kuivõrd erinevad kognitiivsed süsteemid teda nende asjadeni viia võiksid.

Kornblith-i jaoks jääb tõde aga endiselt esmaseks. Ta nõustub küll Stich-iga, et meie kognitiivsed süsteemid peaksid täitma meie soovitud eesmärke, kuid seejuures jääb Kornblith-i arusaama järgi keskseks endiselt küsimus sellest, kas meie vaadeldavad süsteemida ka *tõepoolest* agendi oma soovitud eesmärkideni viivad. Seejuures paigutab Kornblith enda teooria normatiivse külje teise tasandi uskumuste valda, kus esmase tähtsusega igasuguse kognitiivse süsteemi hindamisel on see, et meil kui hindajatel oleks tõesed uskumused nii vaadeldava süsteemi, selle kasutaja kui ka olukorra kohta. Selline vaade ei ole aga vastuolus Stich-iga, kelle jaoks on täpne olukorra hindamine niisama tähtsal kohal. Seda aga vaid siis, kui see täpsus on miski, mida hindaja ka väärtustab. On aga selge, et iga kognitiivse süsteemi puhul ei ole vajalik, et see toodaks tõeseid uskumusi. Näiteid, kus rohkem tõeseid uskumusi tootvad süsteemid agendi enda soovitud eesmärkidest eemale viivad, on lihtne konstrueerida. Peale selle ei pööra Kornblith suurt tähelepanu kogu Stich-i tõe-kriitika sisulisele küljele. Näitamata, kuidas 'tõde' tema arusaama järgi on väärtuslikum kui TÕDE\*, paistab tema vaade jäävat teatavale intuitsioonide tasandile. See on aga miski, mille suhtes on Kornblith ka ise kriitikat üles näidanud.

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina Sander Vaus

(sünnikuupäev: 30.11.1991)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „The Normativity of Truth in Cognitive Evaluation“

mille juhendaja on Daniel Cohnitz

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 05.05.2014