

THE EVOLUTIONARY AND SOCIAL PREFERENCE  
FOR KNOWLEDGE: HOW TO SOLVE MENO'S PROBLEM  
WITHIN RELIABILISM\*

Markus WERNING  
Heinrich-Heine-Universität Düsseldorf

*Summary*

This paper addresses various solutions to Meno's Problem: Why is it that knowledge is more valuable than merely true belief? Given both a pragmatist as well as a veritist understanding of epistemic value, it is argued that a reliabilist analysis of knowledge, in general, promises a hopeful strategy to explain the extra value of knowledge. It is, however, shown that two recent attempts to solve Meno's Problem within reliabilism are severely flawed: Olsson's conditional probability solution and Goldman's value autonomization solution. The paper proceeds with a discussion of the purpose of having a higher value of knowledge as opposed to merely true belief, both in evolutionary and social terms. It claims that under a reliabilist analysis of knowledge it can be explained how knowers could evolve rather than just truthful believers. Subsequently, the paper develops an account of how we can manipulate our testimonial environment in an epistemically beneficial way by valuing reliably produced true belief more than just true belief and so gives an indirect justification of the extra value of knowledge.

Even though every instance of knowledge is an instance of true belief, knowledge—at least in most contexts—is regarded as more valuable than a merely true belief with the same content. When a person believes something true on the basis of, say, a lucky guess, reading tea leaves, or wishful thinking, that is, without knowing it, most of us would say that she is in a less valuable state than if she had knowledge. The doctrine of the extra value of knowledge (see Goldman & Olsson 2009, henceforth “G&O”)

---

\* The main argument of the paper is based on a semi-published master thesis (Werning, 1997) I wrote more than 10 years ago. Special thanks go to Dirk Koppelberg who first raised my interest in reliabilism. I am very grateful to Alvin Goldman, Erik J. Olsson, Ludwig Fahrbach, Gerhard Schurz, and Leopold Stubenberg for helpful comments on earlier drafts of the paper.

is as old as epistemology itself and was first introduced by Plato. For him the doctrine gave rise to a problem that he proposes in his dialogue *Meno* and which is now known to epistemologists as Meno's Problem (Kvanvig 1998, Koppelberg 2005). Plato puts forward the problem as one of rational choice. Assume our rational agent have the desire to go to Larissa. He has to choose between a guide who knows how to get there and a guide who truthfully believes how to get there, but does not know. Since the probability of the agent's desire to be fulfilled, everything else being equal, depends solely on the truth values of the guides' beliefs, it is as rational to choose the second guide as it is to choose the first one. For, the fact that the first guide in addition to having a true belief also knows the way does not increase the probability of success. Plato uses the Greek adjective *ophelimos* 'profitable, useful' to express that true/correct belief (*orthē doxa*) is not less useful than knowledge (*epistēmē*) (Platon 1968, Meno 97c). The conclusion of his critical reasoning can thus be summarized as the claim: True belief has just the same utility as knowledge. The question for us is: Why is it still rational to value knowledge higher than merely true belief?

I would like to stress that Meno's Problem in its original version is phrased in terms of practical rationality and attaches mere instrumental value to truth. The truth of a belief is valuable—so Plato apparently implies—solely because it increases the probability of one's desires to be fulfilled. Meno's Problem in its original pragmatic version thus consists of the following three propositions, which apparently form an inconsistent set:

- MP1. *Extra value of knowledge.* A person's knowledge is more valuable than a person's merely true belief with the same content.
- MP2. *Rational belief evaluation.* A person's belief is the more valuable, the more probable it makes successful action.
- MP3. *No pragmatic difference.* A person's knowledge makes successful action more probable only insofar as the person's merely true belief with the same content would make successful action more probable.

In the paper I will argue that a version of reliabilism provides a solution to the problem and, as far as I can see, the only viable solution. I do however think that it does so for other reasons than G&O have proposed in their article. I will begin with some clarifications regarding Meno's Problem and continue with a discussion of its relation to the so-called Swamping Problem and the value of truth. I will then discuss two ideas by G&O:

the conditional probability solution and the theory of value autonomization. After a criticism of their proposal, I will turn to a problem which is structurally analogous to Meno's Problem, but regards the evolution of knowers. In the final section I will explore how a relatively straightforward solution to the Evolutionary Problem can be transferred to the human social case. The main idea is that valuing instances of knowledge in others (and ourselves) more than instances of merely true belief is itself a means to make our own beliefs more likely to be true, given the conditions under which we may influence our testimonial environments. The underlying psychologically well-founded assumption is that valued practices—in our case: grounding one's beliefs on reliable processes—are more likely to be repeated in the future than unvalued ones. The proposed solution thus is fully coherent with the general epistemological attitude that G&O label *psychological naturalism*.

### *Clarifications*

Before I will develop my argument, the three propositions deserve some additional comments. Clarifications are in need also because there are a number of dissolutions of this problem that easily come to one's mind, but in my eyes are insufficient. First, one might argue that the doctrine of the extra value of knowledge, MP1, is simply false. Kutschera (1981), e.g., argues that knowledge fully consists in a subjective and an objective component. The subjective component is credence or subjective probability. The objective component is truth or objective probability. Both components are maximal in true (firm) belief. According to Kutschera, knowledge, hence, consists in nothing but true belief (see also Beckermann 1997).

Whereas Kutschera generally equates knowledge and true belief, contextualists (e.g. DeRose 1995) have argued that *in certain contexts* the epistemic standards are so low that true belief alone amounts to knowledge. The player in a quiz show who has the choice between two possible answers might, e.g., be said to have known the answer simply on the basis that he gave a true response without further deliberation.

Goldman (1999) advocates the view that the word "know" is polysemous in that it has both a weak and a strong sense. The objections against MP1 would certainly deserve more discussion than is possible in this paper. In accordance with the epistemological mainstream, I will here

simply assume that MP1 is true for an appropriate use of the verb “to know”.<sup>1</sup>

The formulation of MP2 is intended to mean that, all other things being equal, a particular belief is more valuable if it causes the probability of successful action—that is, action that fulfills the desires of the subject—to increase, rather than if it failed to do so. MP2 does not make any comparative claims about the values of beliefs in unrelated scenarios and presupposes that the subject’s desires are held fixed. MP2 does not intend to state that the caused increase in the probability of successful action, across scenarios, is the sole determinant of a belief’s value. It, however, implies that a particular belief would be strictly less valuable if it did not cause the probability of successful action to increase to the degree it actually does. One might say in accordance with MP2 that, in a given scenario, the value of a belief is a strictly monotonous function of the caused increase in the probability of successful action.<sup>2</sup>

Another clarification addresses a presupposition of MP3 that one might call, in memory of William James’s (1907/1949) pragmatist theory of truth, a weak pragmatist principle. It links the truth of a belief to the probability (P) of successful action:

WPP. *Weak Pragmatist Principle*.<sup>3</sup> Let it be a causal background assumption,  $c$ , that a person’s belief that  $p$ ,  $b[p]$ , and her desire that  $q$ ,  $d[q]$ , causally explain the person’s behavior in a given case. Then the satisfaction (Sat) of the desire and the truth (True) of the belief are probabilistically related in the following way:

---

1. What strikes me as wrong with Kutschera’s argument is that in knowledge, rather than in merely true belief the subjective and the objective components are linked to each other in an appropriate way. As Nozick (1981) and Dretske (1971) have pointed out, in the case of knowledge, the person believes something *because* it is true, where the because-relation is spelled out in terms of counterfactual dependency or in some other way. To the contextualists, one might reply that even in contexts of low epistemic standards the latter can be raised in a reasonable way such that knowledge and true belief fall apart.

2. If we could assume that the value  $v$  of a particular belief, the caused increase  $D$  in the probability of successful action, and all other factors  $x_1, x_2, \dots$  determining the value of the belief are metrizable and  $v = v(D, x_1, x_2, \dots)$  is an in  $D$  partially differentiable function, MP2 would come down to the claim that the partial derivative  $\partial v / \partial D$  is strictly greater than 0 at all points of the domain of  $v$ . This would not be equivalent to the sole-determinant claim that if  $v(D, \dots) = v(D', \dots)$ , then  $D = D'$ .

3. I call this pragmatist principle weak because, unlike principles favored by classical pragmatists who *define* truth in terms of success, our formulation is fully consistent with the principle being a factual rather than an analytical claim.

$$P(\text{Sat}(d[q])/c \wedge \text{True}(b[p])) > P(\text{Sat}(d[q])/c \wedge \neg \text{True}(b[p])).$$

The principle thus states that the probabilities that a person's desires be satisfied are strictly greater given her beliefs are true than given her beliefs are false when in both cases the beliefs and desires explain the person's behavior.

A successful action is one whose consequences satisfy the person's desires or, to put it in terms of decision theory, maximize subjective utility. Now there certainly are particular situations in which acting on the basis of a false belief *de facto* has better consequences than acting on the basis of a true belief would. Driving a car I have the desire to cross an intersection safely. I have the true belief that the traffic light is on red and slam on the brakes. The driver behind me did not see the red light and bumps into my car. Had I falsely believed that the light is on green, I would not have stopped and would have crossed the intersection safely—no other cars were passing by. The probabilistic formulation of the principle, however, abstracts from the circumstances of a specific situation. The probability of successful action in the weak pragmatist principle is conditioned solely on the behavioral relevance of the beliefs and desires and on the truth in contrast to the falsity of a belief with a certain content.

### *Epistemic values*

The introduction of the Weak Pragmatist Principle leads us to the question of whether beliefs, as the maxim of rational belief evaluation (MP2) purports, are to be evaluated principally according to how much they increase the probability of successful action. Many epistemologists would claim that there is a specific epistemic value and what determines the epistemic value of beliefs is the intrinsic goal of truth. This attitude is captured by the principle of veritism as stated by G&O:

VP. *Veritist Principle*. All that matters in inquiry is the acquisition of true belief.

The introduction of truth as an intrinsically valuable epistemic goal does, however, not solve Meno's Problem, but simply transforms it into what is a variant of the so-called Swamping Problem. It has been put forward among other by Jones (1997), Swinburne (1999), and Zagzebski (2003). It

was named so by Kvanvig (2003). The Swamping Problem can be regarded as another inconsistent set:

- SP1. *Extra epistemic value of knowledge*. A person's knowledge is epistemically more valuable than a person's merely true belief with the same content.
- SP2. *Epistemic belief evaluation (derived from veritism)*. The epistemic value of a person's belief is determined by its closeness to the goal of truth.
- SP3. *No difference in closeness to truth*. Knowledge is no closer to the goal of truth than is merely true belief with the same content.

The core of the Swamping Problem is that the property of being knowledge does not add any epistemic value to true belief if epistemic value consists in closeness to the goal of truth. If one accepts the Weak Pragmatist Principle the main difference between the Swamping Problem and Meno's Problem is whether truth is regarded as intrinsically or instrumentally valuable, a difference that is important, but won't be of much concern to us in this paper. It seems that only if one were to reject the probabilistic link between the truth of beliefs and the success of action, the two problems would fall apart substantially. But if this link is assumed to hold the two problems can be dealt with roughly in parallel.

Contrasting epistemic rationality with practical rationality opens up the option to explain the extra value of knowledge by introducing further epistemic values in addition to truth. The teleological goal of maximizing coherence or the deontological compliance with certain epistemic obligations like that of avoiding contradictions might be good candidates here. As Sartwell (1992), however, points out, this strategy leads to a dilemma: Either those epistemic values are instrumentally valuable with regard to truth or they are intrinsically valuable in their own right. In the first case, the goal of maximizing coherence or the obligation to avoid contradictions would be regarded as means to approach the goal of truth. Being guided by those goals always brings you closer to the truth. Here the value of those aims is derived from the value of truth, SP2 stays in place and the Swamping Problem remains unsolved. In the second case, the additional epistemic values would be regarded as valuable whether following them brings you closer to the truth or takes you farther away from it, depending on the particular situation. Since the latter possibility is not excluded, a potential conflict of epistemic values is lurking. As Sartwell puts it, there

would “no longer [be] a coherent concept of knowledge” (180). The dilemma indicates that appealing to further epistemic values and thereby dissolving the Swamping Problem (rejecting SP2) and—via WPP—also Meno’s Problem (rejecting MP2) fails to be a promising option.

### *The conditional probability solution*

Goldman and Olsson propose a reliabilist way of solving the Swamping Problem and thus indirectly also Meno’s Problem. Their proposal at first glance seems so attractive because they apparently refrain from introducing additional epistemic values. A reliable process is one that leads to true belief with some threshold probability (Goldman 1986). The reliabilist analysis that knowledge is reliably produced true belief (plus X) on the one hand implies a difference between knowledge and merely true belief.<sup>4</sup> It, on the other hand, does so apparently without further epistemic values because the reliability of a belief producing process is equivalent to its truth-conduciveness.<sup>5</sup> But where does the extra value come from? Even though I agree with their overall attitude, I believe that the explanation G&O propose for the extra value of knowledge doesn’t go through.

G&O’s proposal, in fact, consists of two solutions of which they claim that they “are independent, but [...] also compatible with one another and perhaps complementary” (11). They call their first proposed solution the conditional probability solution (CP). The idea is the following:

If a true belief is produced by a reliable process, the composite state of affairs has a certain property that would be missing if the same true belief weren’t so produced. Moreover, this property is a valuable one to have—indeed, an epistemically valuable one. Therefore, *ceteris paribus*, knowing that p is more

---

4. The addition “plus X” is intended as a placeholder for some condition apt to counter Gettier-style examples against the classical definition of knowledge as justified true belief (Gettier 1963, Lehrer 1965, Goldman 1976). In this context it is important to notice that Goldman’s (1986) process reliabilism identifies justified belief with reliably produced belief. By G&O, the choice of X is regarded as largely irrelevant for the discussion of the Swamping Problem. For, it is the justification of a belief that is supposed to raise the value of knowledge beyond that of merely true belief.

5. Some authors use the term “truth-conducive” in an absolute sense as “leading to truth”. Accordingly, a belief producing process would be called truth-conducive only if it always produces beliefs that are true. In an infinite domain this is harder to attain than even 100%-reliability. In this paper I opt for a probabilistic or statistical interpretation of “truth-conduciveness”, which levels out the contrast to the notion of reliability.

valuable than truly believing that *p*. What is this extra valuable property that distinguishes knowledge from true belief? It is the property of *making it likely* that one's future beliefs of a similar kind will also be true. More, precisely, under Reliabilism, the probability of having more true belief (of a similar kind) in the future is greater conditional on *S*'s knowing that *p* than conditional on *S*'s merely truly believing that *p*. (12)

To analyze the CP solution, one has to distinguish between two statements:

- a) The probability of *S*'s having more true beliefs of a similar kind in the future is greater conditional on *S*'s having the reliably produced true belief that *p* than conditional on *S*'s merely truthfully believing that *p*.
- b) *S*'s reliably produced true belief that *p* makes it likely that *S*'s future beliefs of a similar kind will be true.

There is a subtle, but important difference between the formulations (a) and (b). Whereas (a) is just a comparative statement about conditional probabilities, (b) in addition to (a) makes a direct causal claim. The phrase "makes it likely" of (b) in its most common interpretation implies that *S*'s reliably produced true belief that *p* causes the probability of *S*'s future beliefs of a similar kind being true to increase. While (b) presupposes a specific direction of the causal arrow, (a) is perfectly consistent with the assumption of a common cause and thus only an indirect causal link.

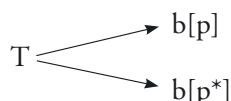
The truth of (a) follows from the definition of a reliable process as one that leads to true beliefs with a probability greater or equal to some threshold probability  $P_r$  ( $P_r > 0.5$ ). The reasoning goes as follows: If *S* has a reliably produced true belief that *p*, *S* has implemented some process type *T* that governs beliefs of kind *K*. The belief that *p* is of kind *K*. Every belief of kind *K* of *S* that is the outcome of a process of type *T* is true with a probability greater or equal  $P_r$ , because processes of type *T* are reliable. If one now makes the slightly oversimplifying assumption that beliefs of the same kind always are produced by belief forming processes of the same type, future beliefs of *S* of kind *K* will be true with a probability greater or equal to  $P_r$ . Since  $P_r$  is strictly greater than 0.5 and the prior probability of a belief being true is at most 0.5, we can conclude:<sup>6</sup> The probability of *S*'s

---

6. The prior probability of a statement depends on the "most natural" partition of the domain. Consider the statement *p* expressed by the sentence "The most expensive evening dress



future beliefs of kind K being true is greater conditional on S's having the reliably produced true belief that p than conditional on S's merely truthfully believing that p. The good thing with (a) is that it is true, but the fly in the ointment is that the scenario is one of common cause:<sup>7</sup>



It is S's access to reliable processes of type T that explains the positive probabilistic correlation between having true beliefs in the future and having a reliably produced true belief now, provided the beliefs are of the same kind. A process of type T is the common cause of the two beliefs and its reliability is the common explanation of the likely truth of the two beliefs. However, it is not the likely truth of the present belief b[p] that explains the likely truth of the future belief b[p\*]. Nor is it the particular event that b[p] was produced by some process of type T that explains the likely truth of the future belief b[p\*]. For, there is no causal dependency of the future belief on the present belief. Nor is there any direct causal dependency between the two particular events of belief production in the present and the future.

This is evident if we consider the following scenario: Assume that on Monday Kim forms the perceptual belief b[p] that her brother wears a red shirt. On Tuesday she produces the perceptual belief b[p\*] that her sister wears a red hat (precisely the same shade of red). We may assume that the belief producing process t on Monday was of the same type T as the belief producing process t\* on Tuesday. Now, it certainly need not be true that if S had not had the belief b[p], she would not have had the belief b[p\*]. It might well have happened that she met her sister on Tuesday, but missed her brother on Monday. Likewise it need not be true that if S's belief b[p] had not been produced by a tokening t of the process type T, S's belief b[p\*] would not have been produced by a tokening t\* of the same process

---

sold in Paris is red". If we chose the partition {p, ¬p}, the prior probability of p would be 0.5. However, if we had chosen the partition {red, green, yellow, blue, white, black}, the prior probability would have been 1/6. Regardless of those problems, the prior probability of positive, non-disjunctive statements with natural predicates is never greater than 0.5.

7. The drawing has only heuristic value. It will in particular be made explicit in the text when we speak of type causation (T causally explains b[p] and b[p\*]) and token causation (t and t\* of type T cause b[p] and, respectively, b[p\*]).

type T. The two beliefs do not causally depend on each other and neither do the two events of production.

The common cause scenario is excluded in formulation (b). The bad thing, however, is that this apparently makes (b) false. S's having the reliably produced belief that p does not cause future beliefs of the same kind to be true in the sense that S's having the reliably produced belief that p makes it likely that future beliefs of the same kind will be true. Unless there is an inferential or some other causal link between the present belief  $b[p]$  and the future belief  $b[p^*]$ , which is normally not the case for two arbitrary beliefs of a kind K,  $b[p^*]$  being true is not causally grounded in  $b[p]$  being a reliably produced true belief:

$$b[p] \not\longrightarrow b[p^*]$$

What does this mean for our evaluative question? If the likely truth of future beliefs of the same kind were indeed causally grounded in the present true belief being reliably produced, we could argue that the present true belief is more valuable than had it not been reliably produced. The rationale would be one of a means-to-end relation. If an end is valuable, whatever helps to bring about the end, i.e., makes the end more likely, is also valuable (provided that other values are not violated). If curing people with antibiotics is valuable, then producing antibiotics is also valuable. If the end of having true beliefs is valuable, then whatever makes it more likely to have true beliefs is also valuable. As we have seen, however, the present true belief that p being reliably produced does not make it likely in this causal sense that future beliefs of the same kind will be true. The extra value of reliably produced true beliefs as opposed to simply true beliefs cannot be accounted for by the means-to-end relation. What is valuable is the reliable process of type T, the common cause. The evaluation of the process itself is not at issue, though.

A defender of the CP solution might eventually want to propose an interpretation of "makes it likely" that (i) avoids the unavailable causal reading and (ii) still explains why the likely truth of the future belief increases the epistemic value of the reliably produced present belief. Olsson (this volume) seems to appeal to an epistemic reading of "makes it likely" as "is indicative of":

- c) S's reliably produced true belief that p is indicative of S's future beliefs being true, provided the beliefs are of a similar kind.

The question now is whether an epistemic, but non-causal relation is apt to transfer value as a proper causal relation in a means-to-end scenario would. My suspicion is that it isn't. To remain in the example given above, an increase of antibiotics production helps to increase antibiotics treatment. An increase in antibiotics production also causes the pollution of drinking water with antibiotics to increase. Due to this common cause scenario, the increase of antibiotics pollution is indicative of an increase of antibiotics treatment. Whereas the increase of the treatment, however, is a good thing, the increase in pollution certainly is not. We usually say that the latter is a negative side effect of the increase in production, which itself is a proper means to increase treatment. The relation "is indicative of" does not *per se* transfer value. In many cases it certainly does, but only in those where it is grounded in a direct causal relation between its relata. It is hence rather questionable whether G&O's CP solution indeed shows how the identification of knowledge with reliably produced true belief (plus X) explains the extra value of knowledge.

#### *Value autonomization*

G&O's proposal to explain the extra value of knowledge includes a theory of value autonomization, which—as they acknowledge—might be regarded as "complementary" (11) to the conditional probability solution. They do so for two reasons: First, whether the truth-conduciveness of a belief producing process can be projected onto future cases depends among others on the "non-uniqueness, cross-temporal accessibility [...] and generality" (G&O, 14) of the process.<sup>8</sup> If a process that once led to a true belief happens to be unique, only accessible at a certain moment in time, or simply too specific to be repeated, its reliability would not imply anything factual about the likely truth of any future belief. Thus, a necessary condition for a reliably produced belief to even be indicative of the likely truth of future beliefs is that the process, which led to the present belief, is non-unique, cross-temporally accessible and sufficiently general. G&O concede that those conditions are not always fulfilled, but only normally.

The second reason is that in a means-to-end scenario, if the end is intrinsically valuable, the means may inherit value from the end, but the

---

8. They also include learnability as a further constraint. There may, however, as well be innate belief forming process with projectible reliability, which are neither learned nor learnable.

means still is not intrinsically, but only instrumentally valuable. If veritism holds, the truth of a future belief is intrinsically value. Now, even if it were the case—which as we saw is questionable—that a person’s reliably produced true belief makes it likely (in this causal sense) that her future beliefs of a similar kind will be true, the present belief would gain additional epistemic value, but only of an instrumental sort.

Value autonomization now is supposed to be a psychological mechanism that bridges the gap between “normally” and “always”, on the one hand, and between “instrumentally valuable” and “intrinsically valuable”, on the other hand. Whereas G&O so far only purport to have shown that reliably produced true belief normally has additional instrumental value with respect to the goal of truth as compared to merely true belief. They concede that there still is an explanatory gap to our practice of value attribution. As far as our practice is concerned, we *always* attribute more epistemic value to knowledge than to merely true belief and we regard knowledge as *intrinsically* more valuable than merely true belief.<sup>9</sup>

G&O hold that there is a mechanism of promotion that starts off with an initial assignment of instrumental value in normal cases and leads to a general assignment of non-instrumental value:

The main possibility we suggest is that a certain type of state that *initially* has merely (type-) instrumental value eventually acquires independent, or autonomous, value status. We call such a process value autonomization. [...] The value autonomization hypothesis allows that some states of affairs that at one time are assigned merely instrumental value are ‘*promoted*’ to the status of independent, or fundamental, value. (17–19, my emphasis)

I have no doubt that mechanisms of value autonomization might indeed exist. It is plausible to assume that what is normally instrumentally valuable often will after some habituation be generally regarded as intrinsically valuable. However, the shortcomings of G&O’s treatment of the Swamping Problem are quite independent of the viability of a theory of value autonomization. The deficits are with the conditional probability solution, on which G&O’s story of value autonomization seems to build. This is because the attribution of non-instrumental general value to knowledge as a result of some mechanism of value autonomization causally presup-

---

9. As we have seen in an earlier section of this paper, there might be epistemic contexts—e.g., a quiz show—where true belief and knowledge fall together. Our statement here might thus need some qualification. The problem of weak epistemic standards is nevertheless orthogonal to the problem of non-projectible processes.

poses that instrumental value is normally attributed to knowledge in an initial phase. The initial attribution of instrumental value in normal cases, however, is sufficiently explained by G&O only if the CP solution suffices to explain the instrumental value of knowledge in normal cases. I have argued that this is not the case because we face a common-cause rather than a means-to-end scenario with regard to future true belief. G&O haven't shown how, in the first place, the property of being reliably produced adds any value to a true belief, even normally when non-uniqueness, cross-temporal accessibility and sufficient generality of the belief forming process are granted. For, a belief being reliably produced does simple not cause the probability of future beliefs being true to increase, even when the reliability of the process is projectible onto future cases. And since there is no direct causal relation, there is also no means-to-end relation between the present belief being reliably produced and future beliefs being true. Consequently, there fails to be a gain even of instrumental value that is grounded in the property of being reliably produced. G&O's value autonomization account of the extra value of knowledge is not a second, independent solution to the Swamping Problem—contrary to what they claim—but stands and falls with their conditional probability solution, a solution whose adequacy we found to be questionable.

### *The evolution of knowers*

To shed some light on Meno's Problem and the related Swamping Problem, I would like to turn to a structurally analogous problem that we face in the evolution of human cognition. Disregarding sceptical doubts for a moment, when we look around us, we find lots of knowers rather than mere truthful believers. Even though some beliefs of ours might be false, much of what we believe is true, and moreover, the things we truthfully believe, most of the time, are also things we know. True beliefs that are based on lucky guesses, reading tea leaves, wishful thinking and the like are rare, at least, if we hold them firmly. We might with some justification say that human brains are knowledge producing machines. If being a knower, however, is a widespread trait among the human species, that trait should have an evolutionary explanation. It is very unlikely that the trait of being a knower is a mere epiphenomenon of evolution as the beauty of butterflies perhaps is. The trait of being a knower must have had some evolutionary advantage for our predecessors.

When we go back in evolution to prehuman species, it is questionable whether the concepts of knowledge, belief and truth still apply. Davidson (1999), e.g., argues that those concepts must not be applied to animate beings without a language capacity. To avoid objections of this kind, I will henceforth capitalize the relevant expressions and talk about KNOWLEDGE, BELIEF and TRUTH thereby referring to mental states and properties of mental states in prehuman species that come closest to knowing, belief and truth in the context of humans. When considering the evolution of human knowers and prehuman KNOWERS we face a problem, the Evolutionary Problem, which is analogous to Meno's Problem. It consists of the following apparently inconsistent set of propositions:

- EP1. The trait of being a KNOWER is evolutionary more successful than the trait of being merely a TRUTHFUL BELIEVER.
- EP2. A trait within a species is evolutionarily the more successful, the more it increases fitness.
- EP3. The trait of being a KNOWER increases fitness only insofar as the trait of being a TRUTHFUL BELIEVER would increase fitness.

EP1 is apparently justified by the fact that being a KNOWER is an evolutionary successful trait that goes beyond that of being a TRUTHFUL BELIEVER. The trait is not epiphenomenal and hence should be evolutionary advantageous. EP2 is a quite general principle of post-Darwinian evolutionary theory where the evolutionary success of a trait might be measured by the frequency with which it occurs among the species members. EP3 finally is a restatement of MP3 where successful action or behavior is biologically interpreted as evolutionary fitness.

The reason for me to propose the Evolutionary Problem is that it has a straightforward solution, which by analogy might be transferred to Meno's Problem and the related Swamping Problem. In a relatively unknown paper, "The Need to Know", Fred Dretske (1989) compares the having of true beliefs to the having of synchronous representations. In order to survive and reproduce, animals—maybe a little more complex than protozoans—need to find food, flee predators, and mate conspecifics. To succeed in doing so, an animal has to coordinate its behavior with its environment. Fleeing a predator means: run *when* a predator approaches. Continuously running, whether or not there is a predator in the environment would be extremely inefficient. It would exhaust the organism and

as likely lead to its death as if it did not run away. Finding food means: go *where* edible things grow and eat *what* is nutritious. Eating everything alike would lead to intoxication, eating nothing to starvation. Mating is good, but not regardless with *whom*. Passing on one's genes will only succeed if the mating partner is fertile, of the opposite sex, and apt in many other respects. To survive, reproduce and finally succeed in evolution, the organism must have the relevant information on when, where, what, and who and this information must result in appropriate on-time behavior. The organism has to achieve a complex task of synchronization. The external target, be it food, a predator, or a mating partner, and the appropriate behavior have to be brought into synchrony. This is typically done by an internal state, for which Millikan (1996) coined the term pushmi-pullyu representation. This trigger-like state has both a descriptive and a directive aspect.<sup>10</sup> It is of great survival, reproductive and evolutionary value for the organism that those pushmi-pullyu representations be synchronous with their respective targets: If a chimp is swinging from liana to liana, his fingers must close at exactly the moment the targeted liana lies in between them. The success of virtually all behavior of non-human animals is dependent on the possession of synchronous pushmi-pullyu representations. It is fair to say that synchronous pushmi-pullyu representations are probably the simplest biologically realistic model of TRUE BELIEFS. The true beliefs of humans might be more complex as far as content and logical structure are concerned, and more decoupled from their targets, but they, very likely, stand in a continuous evolutionary line with synchronous pushmi-pullyu representations.

Now, if synchronous pushmi-pullyu representations are so important for evolutionary success, how are they transmitted? The problem obviously is that synchrony is impossible to transfer from one generation to another. Since the environment is continuously changing, a representation that is synchronous with its target now might be asynchronous with its target in a second. The obvious answer is: what can be transmitted isn't synchrony, but mechanisms of synchronization—not TRUTH, but TRUTH-condu-

---

10. A good example for pushmi-pullyu representations are monkey alarm cries. Those simple, syntactically unstructured, but target-specific cries have a directive component, "run and hide!", and a descriptive component, "a lion is near". The two components aren't separated. Neuroscientific studies indicate that many cortically realized representations of substances like tools and fruits are pushmi-pullyu: both, descriptive features (form, color, etc.) and directive motor affordances (to be peeled in the case of a banana or to be turned in the case of a screwdriver) are part of those representations (Martin et. al. 1996, Werning 2009).

cive processes. The whole purpose of perception is to synchronize certain internal states of the cortex with the corresponding external target objects. The blueprints for the mechanisms of synchronizations, be it the architecture of the optic nerve or the anatomy of the ear, may well be encoded in the genes. Where the mechanisms of synchronization would be too coarse and stiff when encoded in the genes directly, at least, routines to acquire mechanisms of synchronization in development could be inherited.

The solution to the Evolutionary Problem hence is that the trait of being a TRUTHFUL BELIEVER can only be inherited as the trait of being a KNOWER. We here presuppose the reliabilist assumption that KNOWLEDGE is TRUE BELIEF produced by a TRUTH-conducive process. Since synchrony/TRUTH cannot be transmitted from one generation to another, the only way to increase the chance for the next generation to have synchronous/TRUE representations is to transmit synchronizing, i.e., TRUTH-conducive mechanisms. The evolutionary relevant trait of being a TRUTHFUL BELIEVER is to be identified with the trait of being a KNOWER. Thus the inconsistency of EP1 to EP3 is resolved.

#### *Keeping up with truth across time*

Let's return to Meno's Problem and our original question: Why is it rational to value knowledge more than merely true belief? The solution of the Evolutionary Problem offers us two main options to deal with Meno's Problem.

The most radical analogy to draw would be to say that Meno's Problem just is a disguised version of the Evolutionary Problem: success in action is to be interpreted as evolutionary success and to value a person's knowledge is to value her as having the trait of being a knower. Being merely a truthful believer is not an evolutionary relevant trait at all because it cannot be passed on—this holds true at least for all sorts of time-bound beliefs, most importantly perceptual ones. The trait of being a truthful believer—as regards time-bound beliefs—can only be passed on as the trait of being a knower. What counts is the possession and transmission of truth-conducive processes. Each generation has to use their truth-conducive processes anew to build up representations of their ever changing environment. Taking this option, knowledge would be more valuable than merely true belief because it is part of a valuable trait in evolution, whereas merely true belief is not.



What speaks against drawing this radical analogy are a number of important disanalogies between the evolutionary scenario and the rational choice scenario proposed by Plato. First, increasing biological fitness and increasing utility, i.e., the degree to which desires are fulfilled, are quite independent aims. Many of our desires have nothing to do with reproduction or survival, which are the main *definienda* of biological fitness. Sometimes the two aims even are in conflict: In modern societies many people, e.g., have the explicit desire not to reproduce. Second, beliefs in general must not be identified with pushmi-pullyu representations. As Millikan (2004) has worked out in detail, beliefs unlike pushmi-pullyus are typically deprived of directive aspects and often locally or temporally detached from their targets. The behavioral role of beliefs thus goes beyond a mere coordination of on-time behavior with the presence of a certain target. The truth of a belief cannot in general be reduced to synchrony with a target. Third, the solution to the Evolutionary Problem focuses on processes that can be transmitted in a genetic or, at least, mimetic way. Even though a large deal of our knowledge, e.g., perceptual or grammatical knowledge, might indeed be dependent on processes that are passed on in either of the two ways, there are probably many reliable belief-forming processes that are not genetically transmitted. The solution to the Evolutionary Problem thus may be regarded a valuable contribution to answering the question how we have come to be knowers, but it is not a comprehensive explanation why knowledge is more valuable than true belief.

A more indirect lesson to draw from the solution of the Evolutionary Problem is that the extra value of knowledge might have something to do with keeping up with truth across time. In the evolutionary scenario this has an intergenerational interpretation, but there might also be an interpersonal, social understanding. A large part of our knowledge depends on the testimony of others: reports, gossip, narrations, newspaper articles, TV news etc., all taken as expressions of beliefs.<sup>11</sup> But again: why should we value a reported belief that's true, but based on an unreliable source less than a reported belief that's true and based on a reliable source? From the epistemic perspective of truth as well as the pragmatic perspective of rational choice, both reported beliefs, it seems, are on a par. Giving credence to the first brings us as close to the truth as giving credence to the second. However, would valuing a reliably produced reported true belief more than an unreliably produced one really make no difference with

---

11. For an elaboration of the notion of testimony in epistemology see (Coady 1992).

regard to the goal of truth?—Perhaps not in the present, but very likely in the future. By valuing reliably produced beliefs more, we have a chance to manipulate our testimonial environment in a positive way. The underlying assumption is that valued practices are more likely to be repeated in the future than unvalued ones. There is a manifold of mechanisms that seem to support this assumption. They range from psychological enforcement over social sanctions to economic market dynamics. When a child's belief is evaluated and the assignment of value is expressed by praise if the belief is based on evidence rather than hearsay, we enforce certain doxastic dispositions. So future beliefs of the child are more likely to be true and the child reporting its beliefs will bring us closer to the truth ourselves. If we favor uttered reliably formed beliefs of our friends over those that are due to guessing and other unreliable processes, we may, on the long run, attract certain friends to us more than others and thus, by a kind of social selection, make it more likely to arrive at true beliefs in the future when those will be based on our friends' testimony again. Finally, valuing certain beliefs over others may even have economic consequences. If we value a newspaper whose authors most often form their beliefs in reliable ways more than a newspaper whose authors less often do so, we might be ready to pay a higher price (and we should if truth is our primary doxastic goal!). The truth-conducively produced newspaper will more likely flourish and its distribution will spread. Our likelihood to arrive at true beliefs in the future will increase.

It is useful to compare the social scenario to the evolutionary scenario. In the latter the factor responsible for the spread of a trait in a species is an increase in fitness. In the social case, the factor responsible for the spread of a practice in society is an increase in value. As TRUTH/synchrony, being an on-time property of pushmi-pullyu representations, cannot spread within the species directly by a mechanism of inheritance, but only through the inheritance of mechanisms of synchronization, what can be enforced by psychological, social, and economic mechanisms in a society is usually not the truth of present beliefs, but truth-conducive processes that will be in effect in the future.

Since the success of our own truth-seeking activities strongly depends on testimonial beliefs of others being reliably produced, we have developed a culture of positive and negative sanctions regarding the production history of beliefs. The extra value of knowledge is manifest in a practice of providing positive and negative reinforcement which favors reliably produced beliefs over others. Valuing instances of knowledge more than instances

of merely true belief is itself a means to make our own beliefs more likely to be true—in the strong causal sense of increasing their probability of being true. The extra value of knowledge is instrumentally grounded in the ultimate goal of truth, which we want to achieve not only now, but also in the future. The key idea to solve Meno’s Problem and via the bridge of the Weak Pragmatist Principle also the Swamping Problem is to regard value not as a causally inert property of doxastic states, but as a property that in psychological, social, and economic ways has behavioral effects in the choices we make.<sup>12</sup>

## REFERENCES

- Beckermann, Ansgar 1997: “Wissen und wahre Meinung”. In Wolfgang Lenzen (Ed.), *Das weite Spektrum der Analytischen Philosophie: Festschrift für Franz von Kutschera*. Berlin: Walter de Gruyter, 24–43.
- Coady, C. A. J. 1992: *Testimony*. Oxford: Oxford University Press.
- Davidson, Donald 1999: “The Emergence of Thought”. *Erkenntnis* 51(1), 511–21.
- DeRose, Keith 1995: “Solving the Skeptical Problem”. *Philosophical Review* 104, 8–51.
- Dretske, Fred 1971: “Conclusive Reasons”. *Australian Journal of Philosophy* 49, 1–22.
- 1989: “The Need to Know”. In Marjorie Clay & Keith Lehrer (eds.), *Knowledge and Skepticism*. Boulder: Westview Press, 89–100.
- Gettier, Edmund 1963: “Is Justified True Belief Knowledge?”. *Analysis* 23, 121–3.
- Goldman, Alvin I. 1976: “Discrimination and Perceptual Knowledge”. *The Journal of Philosophy* 73, 771–91.
- 1986: *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- 1999: *Knowledge in a Social World*. Oxford: Oxford University Press.

---

12. It should be noted that the position defended here is not one that equates or even equivocates “being valuable” and “being valued”. It is rather a position of epistemic value realism which as is common for realist positions assumes that the epistemic value of doxastic states has causal power with regard to human preferences. A doxastic state’s property of being valuable is the cause of our valuing that state. The property is somehow analogous to a fruit’s property of being sweet. A certain fruit being sweet may justly be said to be the cause of an animal’s preference for this fruit over other fruits that are less sweet. In the case of sweetness, uncovering the underlying causal mechanisms may be a rather ambitious aim that may involve a complex evolutionary story. A similarly complex story remains to be told in the case of epistemic value.

- Goldman, Alvin I. & Erik J. Olsson 2009: "Reliabilism and the Value of Knowledge". In Duncan Pritchard, Alan Millar, & Adrian Haddock (eds.), *Epistemic Value*. Oxford: Oxford University Press, 19–41. Page references are to the manuscript version of this paper.)
- James, William 1949: "Lecture VI: Pragmatism's conception of truth". In William James, *Pragmatism*. New York: Longmans Green & Co. (Year of original: 1907), 197–236.
- Jones, Ward E. 1997: "Why Do We value Knowledge?". *American Philosophical Quarterly* 34, 423–39.
- Koppelberg, Dirk 2005: "Zum Wert des Wissens: 'Das Menon-Problem' ". *Logical Analysis and History of Philosophy* 8, 46–55.
- Kvanvig, Jonathan L. 1998: "Why Should Inquiring Minds Want to Know? Meno Problems and Epistemological Axiology". *The Monist* 81, 426–51.
- 2003: *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
- Lehrer, Keith 1965: "Knowledge, Truth, and Evidence". *Analysis* 25, 168–75.
- Martin, Alex, Cheri L. Wiggs, Leslie G. Ungerleider, & James V. Haxby 1996: "Neural Correlates of Category-Specific Knowledge. *Nature* 379, 649–52.
- Millikan, Ruth 1996: "Pushmi-pullyu Representations". *Philosophical Perspectives* 9, 185–200.
- 2004: *Varieties of Meaning*. Cambridge, MA: MIT Press.
- Nozick, Robert 1981: *Philosophical Explanations*. Oxford: Clarendon Press.
- Platon 1968: *Œuvres complètes, Tome III, 2e Partie: Gorgias, Ménon* (A. Croiset & L. Bodin, eds.). Paris: Les Belles Lettres.
- Sartwell, Crispin 1992: "Why Knowledge Is Merely True Belief". *Journal of Philosophy* 89, 167–80.
- Swinburne, Richard 1999: *Providence and the Problem of Evil*. Oxford: Oxford University Press.
- von Kutschera, Franz 1981: *Grundfragen der Erkenntnistheorie*. Berlin: de Gruyter.
- Werning, Markus 1997: *Erkenntnis und Schlußfolgerung*. Master thesis, Free University of Berlin, Berlin. (<http://tinyurl.com/5mgqnk>)
- 2009: "Complex First? On the Evolutionary and Developmental Priority of Semantically Thick Words". *Philosophy of Science* 76(5). (In press)
- Zagzebski, Linda 2003: "The Search for the Source of Epistemic Good". *Metaphilosophy* 34, 12–28.