

## Can Evolution Explain Rationality?

[... draft... a really rough draft... still a work in progress ... ]

James Sage

### Introduction

In “The Evolution of Rationality” Elliot Sober criticizes Daniel Dennett’s claim that it is a priori true that (human) believers are rational. In order to cause difficulties for Dennett, Sober introduces an analogy between cognition and respiration. I argue, however, that such an analogy can be safely absorbed by Dennett’s position, thus Sober’s criticism against Dennett is without force.

I then criticize Dennett by claiming that such an account has no business appealing to evolution by natural selection to support his views regarding rationality. In short, if it is true a priori that believers are rational, then there is no reason to appeal to evolutionary considerations, which are both empirical and a posteriori.

### Dennett and Cognitive Systems

In the intellectual footsteps of Davidson and Quine, Daniel Dennett has articulated and defended a view known as “the intentional stance” which utilizes the concept of an intentional system and the application of an intentional interpretation. The application of an intentional interpretation provides a basis for the ascription of mental content to others, including beliefs and other propositional attitudes. And this allows us to explain and predict their actions.

According to Dennett, what it is for a system to have beliefs is to be able to apply an interpretation to its behaviors, according to which beliefs, desires, and other examples of intentional content figure into the explanation, and ultimately a prediction, of their behavior. I will refer generally to such organisms, persons or intentional systems as *cognitive organisms*.<sup>1</sup> All cognitive organisms, including human persons, are believers. Clearly, there are some organisms that do not have beliefs. These organisms fail to be *cognitive organisms*; such organisms fail to be cognitive (mushrooms, worms, etc.).<sup>2</sup> I will use “interpretationist” to refer to the kind of account, such as Dennett’s account, which understands cognitive organisms in terms of the ascription of mental content.

---

<sup>1</sup> Of course, Dennett intends his account to apply generally to systems, not just (carbon-based) organisms. I will, however, use “organisms” to refer to systems generally, and “cognitive organisms” to refer to intentional systems. I am interested in that subset of intentional systems that are evolved organisms (as opposed to or robots, which might be intentional systems, though are not organisms).

<sup>2</sup> According to Dennett, such systems must be interpreted from the “physical stance” or the “design stance”.

One of the central tenets of Dennett's interpretationist account is the following Holist insight: in order to apply an intentional interpretation to the behaviors of an organism (an organism whose beliefs we are attempting to decipher), we must be able to make sense of the organism's entire range of potential beliefs. That is, in order to make sense of a someone's beliefs at all (to make sense of someone as a *believer*), we must assume that they have *lots* of beliefs, and that most of their beliefs are *true*. But why think that a believer has numerous, and mostly true, beliefs? Isn't it possible for a believer to have very few, or mostly false, beliefs? According to Dennett, the answer is: No.

Dennett argues that a believer cannot have mostly false beliefs. Consider the following example. Suppose A is assessing the (potential) beliefs of B. In order for A to make sense of any of B's beliefs (that is, in order to make sense of B as an intentional system), A adopts the intentional stance with respect to B. As a result, A cannot interpret B as having mostly false beliefs. Why not?

First, on Dennett's account, B won't have just one belief on a particular topic. In order to have even one belief about something, B must have numerous other beliefs as well. So, in order to make sense of B having any *single* belief, A will need to ascribe to B a whole host of other beliefs (for example, beliefs about particular concepts, various facts, as well as beliefs about the meanings of the terms used, their reference, etc.).

Now, consider the following exchange between A and B:

B: "Ted Williams was the greatest hitter of all time."

A: "So, you don't think Hank Aaron was the greatest hitter of all time?"

B: "Who's Hank Aaron?"

A: "What about Mickey Mantle or Babe Ruth, or Sammy Sosa or Barry Bonds?"

B: "Who are they?"

A: "You've never heard of these famous baseball players?"

B: "What's 'baseball'?"

A: "You mean you are unfamiliar with the sport of baseball?"

B: "Baseball is a sport? Huh. I wouldn't have guessed."

A: "Forget it. You're an idiot."

B: "What's an 'idiot'?"

A: "Ugh."

By this time, A probably starts to suspect that B is simply repeating what he heard someone else say when B utters "Ted Williams was the greatest hitter of all time". Based on this evidence, B has no idea what it *means*, and in any case, it is not clear that B has the corresponding *belief* that *Ted Williams was the greatest hitter of all*

*time*. What undermines A's grounds for attributing this belief to B is that A cannot be sure that B has proper grasp of the concepts required to make sense of B's having such a belief. So, not only does A have reason to doubt whether B understands what it *means* to say that "Ted Williams was the greatest hitter of all time", A also has reason to doubt whether B has any corresponding *belief* at all.

This brings us to a second point, which is also common to both Holist accounts and Dennett's interpretationist account: in order for A to understand what B *means*, A will need to attribute to B lots of other beliefs, including beliefs held by B regarding what his utterance means.<sup>3</sup> In the example above, in order to make sense of B's belief that *Ted Williams was the greatest hitter of all time*, A would need to ascribe to B the belief that *Ted Williams was a baseball player*, that *some baseball players are hitters*, that *Ted Williams was a hitter*, and so on. A would also have to attribute to B various beliefs about the meanings of the terms used in B's utterance: a belief about what "greatest" means, what "hitter" means, and to whom "Ted Williams" refers. This turns out to be lots of true beliefs that A would have to attribute to B.

In order for A to make sense of B having the relevant concepts, A must be able to attribute to B a number of true beliefs about those concepts. Without these other beliefs about concepts and meanings, there is nothing to help A understand what B means when B utters "Ted Williams was the greatest hitter of all time" and so there's little reason for A to attribute *any* corresponding belief to B about Ted Williams.

In order to make sense of Dennett's claim that believers will have mostly true beliefs (i.e., that it is impossible for a believer to have mostly false beliefs and still be a believer), let's consider what occurs when A attempts to ascribe to B lots of *false* beliefs—perhaps A attempts to make sense of B as having *mostly false* beliefs.<sup>4</sup>

For every false belief that A attributes to B, A must be able to make sense of B having the specific concepts which figure into that false belief. Otherwise, it wouldn't be proper to say that B even had a belief—reasonably, B must understand what he himself means in order to have a belief. In order to ascribe a *false* belief to B, A must be able to attribute to B a number of beliefs that A takes to be true.<sup>5</sup> Thus, for every *false* belief A ascribes to B, A must also ascribe to B *numerous true* beliefs. So, in order for A to understand B as a cognitive organism (as a *believer*), B's false beliefs could never outnumber B's true beliefs. So, the idea that B has *mostly false* beliefs only makes sense if A *ignores* all of those true beliefs that would have to be attributed to B just to make

---

<sup>3</sup> This idea, that beliefs fix the meanings of the terms we use is roughly in line with the constellation of views put forward by Davidson and Quine, sometimes known as Meaning Holism or Concept Holism

<sup>4</sup> I've hurried past an issue here: it's not clear that "having mostly true beliefs" and "having mostly false beliefs" exhausts the options. However, I think such a simplification is warranted (leaving room for the unique and unusual case in which an intentional system has exactly 50% true and 50% false beliefs). I think it is proper to say of a flipped coin which has landed heads up 51 times out of 100 that it has landed "mostly heads up"... in the same way, we can treat a person's beliefs, such that mostly true and mostly false essentially exhaust all of the reasonable options.

<sup>5</sup> And I think that A must not only take the beliefs as true (for A), but I think that A must take the beliefs as true for B (i.e., that B reasonably thinks that the belief(s) in question are true).

sense of B's false beliefs. However, it is unacceptable for A to ignore all of these true beliefs. It follows that, in order for A to understand B as a *believer*, B must be interpreted by A as having *mostly true* beliefs, and this precludes A from asserting that B has mostly false beliefs.

Redundant wrap up: Understanding certain concepts, and having true beliefs about those concepts, is a necessary condition for having beliefs in which those concepts figure. Moreover, having numerous (true) beliefs is a necessary condition for having just one false belief. Hence, a believer must have mostly true beliefs.

Dennett goes on to say that if we have an intentional system with mostly true beliefs, then we have a rational system. In other words, we must assume, as an intrinsic part of our interpretation of B, that B is rational. [This seems to be Dennett's only criteria for rationality... but I'm not sure if that's correct... In the *Intentional Stance*, Dennett is clear that he offers no positive account of what, in particular, constitutes rationality for a believer... he claims that rationality is NOT deductive closure, nor is rationality understood in terms of the elimination of all contradictions... rationality, for Dennett, seems to be a feature of the interpretationist framework under which we make sense of the intentional (propositional) content of cognitive organisms/intentional systems... rationality is part of the interpretationist "package" ... where we have a believer we have cognitive organism with mostly true beliefs, and whenever we have a cognitive organism with mostly true beliefs, we have a rational organism... it's just part of our interpretation].

For Dennett, the notion of rationality as an enabling assumption (a background condition, constitutive of the intentional stance (it might also be called the "rational stance"). Thus it is constitutive of all intentional systems. Thus, Dennett's rationality assumption is a pre-condition for interpreting a system as a cognitive system – as a system with a mental life, as a believer.

So, in order to make sense of the claim that B has beliefs at all, A must adopt the intentional stance. In so doing, A must treat B as rational. As such, it must be assumed that most of B's beliefs are true. So, to have beliefs at all, an organism must be rational. To this extent, Dennett offers a position according to which believers are *necessarily* rational – believers cannot fail to be rational unless they also fail to be believers.

Both of the following claims are necessarily true and both are intrinsic features of adopting the intentional stance: (i) believers have mostly true beliefs; (ii) believers are rational. In other words, Dennett claims that it is true a priori that believers are rational. This can also be summarized as follows: if an organism has beliefs at all, then that organism is rational. Perhaps it is best expressed as a biconditional: an organism is a believer *if and*

*only if* that organism is rational. This biconditional is *necessarily* true (given Dennett's interpretationist framework).<sup>6</sup>

If a system fails to be rational, then the system ceases to be a believer, and so it ceases to be an *intentional* system (at this point, in order to make sense of its behavior, we must adopt some other interpretive "stance" such as the "physical stance" or the "design stance"). When we adopt the intentional stance, we inherit a package deal: when the conditions have been met to be a believer, then that believer has mostly true beliefs *and* that believer is rational. According to Dennett, it doesn't even make sense to speak of believers who do not have mostly true beliefs. Similarly, it doesn't even make sense to speak of believers who are not rational. These are point to which I will return.

### Sober's Criticism

As we have just seen, Dennett claims that it is true a priori that cognitive organisms (believers) are rational. In "The Evolution of Rationality", Sober makes three claims, the first two of which are immediately contrary to Dennett's view. A full treatment of the third claim, while in mild agreement with Dennett, would take us too far beyond the scope of this paper.

First, Sober claims that the issue of whether or not a believer is rational should be investigated according to two, independent questions. Contrary to Dennett, Sober claims that the issue of whether or not an organism has beliefs and whether or not that same organism is rational must be understood as two distinct questions. For example:

(Q1) Does organism O have beliefs?

(Q2) Is O rational?

Are believers rational?

Sober objects that Dennett treats the above questions as *one* question, not two. According to Dennett, this single question is answerable a priori (or, at least, it is a priori true that believers are rational). Dennett's question might look something like:

(Q<sub>D</sub>) Is organism O a *cognitive* organism?

---

<sup>6</sup> This appears to be slightly different than saying it's true *a priori*, as Sober seems to stress. (see below)

According to Dennett, if an organism is a cognitive organism, then it is also true that the organism in question is has mostly true beliefs and that the organism in question is rational. Thus, there is no need to ask the further question regarding O's rationality: if O is a cognitive organism, then O is (necessarily) rational.

Of course, Dennett must still take an empirical look at O in order to figure out O's status as a believer (as a cognitive organism). That is, whether or not we can make sense of O as a cognitive organism must be established empirically: whether or not O (in particular) is a *believer* is not known a priori. However, it is known a priori that *if* O is a believer, *then* O is rational and O has mostly true beliefs. According to Dennett, it is a priori true that believers are rational and this is precisely what Sober means to criticize.

Regarding Sober's first claim, questions (Q1) and (Q2) are distinct questions. Minimally, this entails that the truth of (Q1) is logically distinct from the truth of (Q2). On Sober's account, it's possible for O to be a believer and yet *fail* to be rational. That an organism has beliefs at all is independent of whether or not the organism is rational (believers can fail to be rational). This is plainly in disagreement with Dennett's interpretationist account.

Second, Sober claims that if it happens to be *true* that organism O has beliefs *and* that O is rational, then these are true a posteriori (they are true independently, however). Where an organism satisfies the conditions to be a believer, it could also be the case (or fail to be the case) that O satisfies the conditions for being rational. Moreover, these conditions (the belief condition and the rationality condition) are not necessarily identical: on Sober's account, there is nothing about the nature of being a believer that guarantees, *a priori*, that believers are also rational. Again, this is contrary to Dennett's interpretationist account.

Finally, Sober claims that if it is true that O is a cognitive organism and it is true that O is rational, then evolutionary considerations might be able to provide an explanation of how/why these a posteriori truths came about. Here's what Sober has in mind: that being rational (in some "core" sense of rationality) is adaptive (and that organisms with adaptive traits (cognitive or otherwise) tend to reproduce their kind with more frequency).

To the extent Sober thinks that evolutionary considerations play some role in explaining why believers are rational), there is mild agreement with Dennett. This mild agreement occurs because Dennett, too, thinks that evolutionary considerations help to establish that rational cognitive organisms evolved.

Variation among rational organisms required for evolutionary explanation, variation in the trait we're investigating... and just how would such measurement of variation take place? Well, in the case of rationality, Sober suggests, it won't occur a priori.

However, in the final section of the paper, I will claim that Dennett should deny that evolution is *required* to explain the conceptual connection linking believers with rationality. Nor should Dennett say that evolution guarantees such a connection. Evolution guarantees no such thing. If there is any “guarantee”, it comes from the conceptual derivation of the interpretationist account.

[In a separate paper, I treat Sober’s claim that evolutionary considerations make it likely that organisms possess some minimum (or “core”) rationality... the a posteriori claim that believers are (minimally) rational...]

### Sober’s Project

Sober also distinguishes two claims:

Strong claim: that belief implies rationality.

Weak claim: that belief implies some minimum (or “core”) rationality.

And we can imagine the weakest claim:

Weakest claim: that belief is wholly independent of rationality.

According to Sober, neither the Strong claim nor the Weak claim is true a priori (if either is true). Yet, Sober seems to think that natural selection might in fact make the weaker claim true (natural selection might favor organisms who routinely make at least *some* of the inferences from existing beliefs... this is a nice insight into what Sober takes to be “rational” for an organism: from existing beliefs to make some portion of all of the (deductively) licensed inferences. Perfectly rational organisms will make *all* these inferences... perfectly irrational organisms make *none* of these inferences... So, evolution by natural selection might favor organisms who make at least *some* of the inferences (again, this sounds as though “making an inference” is something that occurs, objectively, in someone’s head... like “releasing insulin” or “removing chemical waste from the bloodstream”).

Sober invites us to imagine a continuum of organisms: on the far left are those that infer *none* of the logical consequences of a typical belief. On the far right are those organisms who infer *all* of the logical consequences of a typical belief (see Table 1).

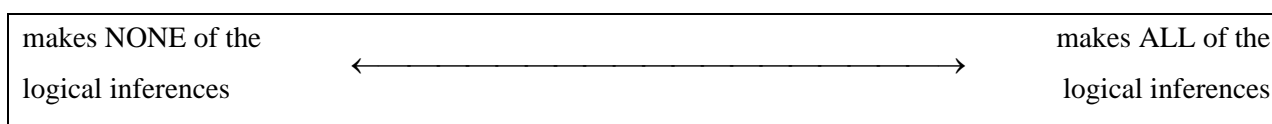


Figure 1

A concrete example might be useful. Consider the following single, typical belief:

(b) Jones owns a Ford and Brown is in Barcelona.

Now, a perfectly rational believer will also infer each of the following beliefs from (b):

(c) Jones owns a Ford.

(d) Brown is in Barcelona.

Call this inference rule “AND-ELIMINATION”. Because AND-ELIMINATION is a deductively licensed inference, perfectly rational agents (located to the far right of the continuum) will always believe each conjunct (c) and (d). AND-ELIMINATION is a licensed inference for all such conjunctive beliefs. However, some cognitive organisms will not make such inferences in all cases. Those organisms at the far left end of the continuum will never make such inferences.

Humans are not at either end, but somewhere in the middle. Sober claims that Dennett denies that it is logically possible for *cognitive* organisms to populate all locations of this continuum: on Dennett’s view, at some (perhaps fuzzy) point approaching the far left end of the continuum, the organism in question ceases to be a *cognitive* organism at all—its behavior (for example, failing to perform AND-ELIMINATION) suggests that it is making very few inferences if any, and so its behavior leaves us questioning whether it has beliefs at all (it leaves us questioning its status as an intentional system, much like we started to doubt whether or not it was reasonable to attribute to B the belief that *Ted Williams was the greatest hitter of all time*). Consider the following exchange I might have with Bubba:

Me: “Hey Bubba. Do you have any beliefs about Linda?”

Bubba: “Yea. I believe that Linda is a bank teller and that Linda is a feminist.”

Me: “So, you believe that Linda is a feminist.”

Bubba: “No. I don’t believe that!”

Me: “But you said that you believe that Linda is a bank teller and that Linda is a feminist.”

Bubba: “Yes, *that is* my belief. But I *don’t* believe that Linda is a feminist.”

Me: “What are you talking about? When you uttered that you believe that Linda is a bank teller and that Linda is a feminist, doesn’t it follow that you also believe that Linda is a feminist? It’s a simple logical inference!”

Bubba: “Well, I don’t know about all that. But I do know that I don’t have the belief that Linda is a feminist. By the way, what’s a ‘logical inference’?”

Me: “Forget it, you’re an idiot.”

Bubba: “What’s an ‘idiot’?”

There may be organisms at the far left end of the continuum, but according to Dennett, they are merely organisms, not *cognitive* organisms (their behavior must be predicted and interpreted based on the design stance or the physical stance). At some point (again, perhaps a fuzzy point, indicated by the “T” in Figure 2) moving toward the left end of the continuum, such that whenever an organism fails to be rational, the organism in question also fails to be cognitive (it fails to be a believer). According to Dennett, belief and rationality go hand-in-hand.

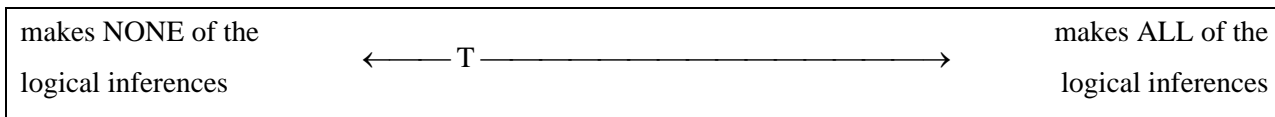


Figure 2

Let the portion of the continuum to the left of the “T” in Figure 2 represent the (possibly fuzzy) point at which organisms cease to be rational. On Dennett’s account, organisms occupying the continuum to the left of the “T” in will not be rational, nor will such organisms be believers (where rationality is abandoned, so is belief).

And because a system’s behaviors are all that we can appeal to in order to establish whether or not the organism even has beliefs, there is no other “litmus test” for an organism’s beliefs—there’s no way to make sense of mental content other than the behaviors exhibited by a system. On Dennett’s view, you cannot stick a probe into an organism’s skull in order to discover its beliefs.

Contra Dennett, Sober claims that all points on the continuum may be occupied by cognitive organisms (believers). On Sober’s account, cognitive organisms can (conceivably) occupy the portion of the continuum left of the “T”. Such organisms are believers, but they fail to be rational. Sober might even agree that there is some (perhaps fuzzy) point, such as “T” which indicates the cut-off for rationality. What Sober would deny, however, is that organisms to the left of the “T” cease to be cognitive organisms. If they DO cease to be cognitive organisms at the very same place as they cease to be rational, such a coincidence would be an empirical, a posteriori happenstance.

No matter how “irrational” we take an organism to be, failing to perform even the easiest inferences (say, AND-ELIMINATION) will never suffice, a priori, to conclude that the organism is not a believer (at least, we cannot conclude solely on the basis of an organism’s rationality whether or not it is a believer). Having a belief, on Sober’s representational account, does not amount to an observer’s application of an interpretation along with the background assumption of rationality. Sober seems to want to say that having beliefs is independent of the ascription of beliefs and the assumption of rationality. Dennett claims that this is the only way to understand intentional systems.

Regarding Sober’s “weak” and “strong” claims, it might be true that natural selection favors those organisms at some minimum level of inference-making (Sober’s rough measure of rationality, where the minimum level can be represented by “T”). And so, as a matter of empirical happenstance, we don’t find cognitive organisms occupying the far left end of the continuum, to the left of the “T” (we can *imagine* such organisms (there’s nothing inconceivable about them), but they just aren’t actually around because they were penalized by the impersonal hand of natural selection). Again, the point might be fuzzy, but Sober seems to think that the far left will *in fact* be unoccupied, and this is because such organisms will be (on the average) less fit than organism’s occupying other regions of the continuum. (Organism’s occupying the far right end of the continuum might be quite rare as well: the biological cost associated with computing ALL of the logical consequences of an organism’s beliefs could render such organisms as biologically unfit... due to the lack of informational economy (efficiency) that such a design would yield... and so if such creatures ever DID appear on the face of the earth (inferential idiots or inferential geniuses), they may have been eliminated soon after, without reproducing their kind...)

Clearly, Dennett and Sober disagree about how to understand the relationship of believers and those who are rational (and their status as rational beings). Part of this is because Sober is assuming that what it is to have a belief is not to be characterized as Dennett and the interpretationists have argued. For Sober, to have a belief has everything to do with what’s going on inside an organism’s head (i.e., with the “wiring in the head”, or what might be understood as mental representations that manifested by this wiring). The extent to which an organism makes various inferences... where “making an inference” is, again, in some way an objective manipulation or transformation of that wiring in the head and the corresponding representations... and not an interpretation of the organism’s (linguistic) behavior.

In other words, Sober is approaching this issue with what I call a “representationist” view of belief. Representationist views do not *rely* on the adoption of the intentional stance or an application of an

interpretation—beliefs are a kind of real thing (observer independent; not the product of an attribution; but *really* there<sup>7</sup>) or mental representation.

Sober tries to show that Dennett’s a priori claim about the relation between believers and rationality is inadequate—to show that it is incorrect to conclude that organisms who fail to make the easiest inferences (those organisms located to the left of the “T”) are no longer believers.<sup>8</sup> Sober introduces an analogy with a non-cognitive trait (choose any other functionally complex, non-cognitive, evolved trait). Sober chooses to compare cognition (or cogitation) with respiration. Consider Figure 3:

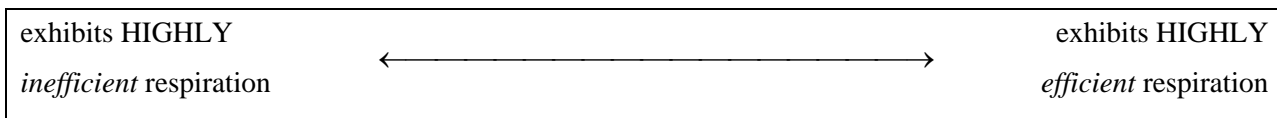


Figure 3

Here, the intuition is that respiration is a kind of “real” event (observer-independent), for example, an objectively measurable conversion of ambient oxygen to physiologically useful oxygenated blood. Some organisms are able to achieve highly *efficient* respiration (extracting lots of oxygen from the air and producing highly oxygenated blood, all with very low physiological cost to the organism). Suppose that these organisms, such as elite marathon runners or triathletes, occupy the far right end of the continuum in Figure 3. Organisms with highly *inefficient* forms of respiration (organisms with lung disease, emphysema, etc.), are located at the far left end of the continuum. Such organisms are utterly incapable of transforming ambient oxygen into useful oxygenated blood. Sober wants to say that they are respiring organisms, they just aren’t respiring very well.

Presumably, Sober chose this example precisely because it is meant to suggest that cognition (beliefs) is just this kind of natural, biological, measurable trait, such as respiration.

Sober concludes that we shouldn’t (wouldn’t) think that it is a priori true that if an organism engages in respiration at all, that it must engage in respiration of at least some basic (if not quite elevated) level of efficiency. We shouldn’t demand that in order to qualify as a respiring organism that organisms must achieve a certain level of respiration (say, the level that normal, healthy human adults typically achieve—which would be someplace in the middle). To place a cut off on this continuum based on what level of respiration typical, healthy *humans* happen to be capable of, and to say that below such a cut off organisms fail to be respiring, is to impose our own (familiar, yet arbitrary) standards of respiration. Just because WE happen to engage in respiration at fairly high levels of efficiency, it does not follow that this level thereby sets the level at which

---

<sup>7</sup> Dennett, of course, would insist that the interpretation allows us to say that mental content is REALLY there too.

<sup>8</sup> Question: how does one show that a claim is not true a priori? You could show that the claim is false, or you could show that it is true a posteriori. How else?

organisms are respirers. (and below which others fail to be respirers... or are they just really inefficient? At some point, really inefficient respirers simply cease to be respirers at all...)

According to Sober, to be a respiring organism, it is not required to engage in some basic (or even highly elevated) level of efficient respiration (where efficiency might be measured in terms of converting ambient oxygen to physiologically useful oxygenated blood, etc.). And so it is with cognition as well. Applying this reasoning to cognition: to be a cognitive organism, it is not required to engage in some basic (or even highly elevated) level of efficient cognition (where efficiency might be measured as Sober depicts rationality: in terms of making logically-licensed inferences, or removing contradictions in your belief-set, etc.)

This analogy is supposed to challenge Dennett's claim that organisms fail to be believers when they drop below a certain (minimal) level of rationality. There are at least two ways to respond to such an analogy.

First, Dennett might reply that cognition is NOT the same as respiration in at least one respect: we do not measure inference-making, belief-formation, and derivation of logical consequence in the same way that we measure conversion of ambient oxygen to oxygenated blood. In order to identify cognition, we must adopt the intentional stance (and there is no other litmus test for cognition, no other scientific measurement of beliefs). By adopting the intentional stance we are engaging in a kind of practice in order to make sense of behaviorally complex systems. This is called the rationality game (we do not adopt this practice when investigating respiration). In other words, it's not clear that the analogy is a good analogy... hence, it is not a good criticism... Hence, the comparison with respiration that Sober has introduced is *disanalogous*.

A second way Dennett might reply is by claiming that the analogy is perfectly appropriate: just as we identify beliefs and rationality, we identify and measure the occurrence of respiration in terms of complex input-output functions. Appropriate inputs for respiration include oxygenated air along with a system of organs and tissues. These organs and tissues transform the air into relevant output: air that has less oxygen and more carbon dioxide, and blood with oxygen-saturated hemoglobin molecules. *It is on the basis of these input-output pairs that we measure the occurrence of respiration.* If these input-output pairs did not occur, then we would have little reason to suppose respiration was taking place. (we would not just simply say that respiration was occurring at some very low level)

If the inputs were sodium chloride and carbon monoxide, and the outputs were calcium bicarbonate or nitrite sulfide, such a system would not be a *respiring* system. It's not just that the system is engaging in really

inefficient respiration, it's not a respiring system at all (perhaps it's a system engaged in calcification or nitrification, etc.).<sup>9</sup>

Or consider the following: if the system took in oxygenated air, but the output was (nearly) equally oxygenated air (little, if any, change in the air), we would begin to question whether the system was respiring at all. (we have little evidence / no reason to think that the system is respiring...) (Imagine inflating and deflating the lungs of a human corpse ... this activity doesn't qualify as respiration because the relevant changes in physiology and exhaled air (the outputs) are not occurring...)

So, Dennett would be free to claim that the functional transformation of inputs into appropriate outputs seems to be the way we identify the occurrence of respiration, and that this functional approach is no different, in principle, than what we do when we identify the occurrence of cognition when adopting the intentional stance (we're making our assessment of the system's beliefs on the basis of the various input-output pairs... in this case, they are behavioral and linguistic events). In order to say of a system that it has beliefs or that it is rational, a certain range of input-output pairs must obtain. The failure to produce the correct input-output pairs (e.g., behaviors, including linguistic behaviors) leaves us wondering whether the transformation in question is occurring (whether or not this particular *transformation* has to do with respiration or cognition). So, ... the respiration analogy can be absorbed by Dennett's interpretationist account.

Thus, Sober's criticism fails to carry any weight: either the analogy fails (say, because the two phenomenon are not similar enough, thus such an appeal carries no weight). Or else, the analogy can be absorbed by Dennett's interpretationist account (interpreting respiration as a complex input-output function, thus rendering the analogy without force).

[I guess the other possibility is that the analogy is accurate in terms of respiration and cognition are objectively measurable phenomenon... but this would be an example of Sober simply registering his disagreement with Dennett... it's not an argument against Dennett's position... it's just claiming Dennett is wrong, supposing the interpretationist account has got it wrong...]

So, Dennett can escape Sober's criticism... But why is Dennett appealing to evolution?

[final section suggesting that Dennett has no business appealing to evolution to support the claim that believers are rational]

---

<sup>9</sup> I might as well say that I am an astronaut, I just lack a many of the relevant behavioral input-output pairs that make be a really good astronaut... I think it is reasonable to say that I lack so many of the input-output pairs that I'm not just a poor astronaut, I'm not even an astronaut.

Let's look at Dennett's claim that believers are rational is true a priori.

To the best of my knowledge, the truth of evolution does not help to support this claim. According to the interpretationist account offered by Dennett, the possibility of a believer not being rational or not having mostly true beliefs is ruled out conceptually. It is ruled out a priori. The fact (or truth) of evolution—if it is a fact (if it is a truth) – is a posteriori and lends no support for the claim that believers are rational. How COULD it lend any support? What could it contribute? Presumably, the fact (or truth) of evolution is empirical, contingent. Do empirical truths lend rational, epistemic support to a priori truths? At most, empirical truths aid us in discovering various truths, some of which might be true a priori. Empirical truths, however, are not a priori truth-makers (they don't function as truth-makers).

Consider the example: All bachelors are unmarried. We could employ empirical means to discover that this statement is true (we could interview all bachelors and find out that all of them are unmarried). However, such empirical discoveries could only lead us to believe the truth of the statement in question (the empirical results do not make it true that all bachelors are unmarried). Also, in this way, empirical results (truths) do not provide evidential support for a priori truths (our empirical inquiry might have been inconclusive... but this would not change the status of the belief, as a belief true a priori). Empirical truths are, at best, guides for belief formation – they might play a role in the realm of discovery. Empirical truths are not the makers of a priori truths, and thus they cannot lend evidential support.

In the same way, the fact (or truth) of evolution does not lend evidential support to the claim that believers are rational. Evolution might serve as a guide for our affirmation of the statement (though in this particular case, I doubt it), but the empirical fact of evolution does not make it true that believers are rational.

All that we can say is that believers are rational and that believers have mostly true beliefs. So, I'm willing to accept that much. However, even if we accept that believers have mostly true beliefs and that believers are rational, this position is neutral with respect to whether any of those beliefs are justified; whether the believer in question is a knower; and whether such knowledge is very interesting or robust.

This, of course, leaves open the possibility that believers fail to be knowers in a variety of doxastic realms (unless all that is required to be a knower is that one is rational). While being rational (in Dennett's sense), believers could still fail to have interesting true beliefs, especially if our interest is in scientific knowledge of the external world. In other words, the sub-class of a rational believer's beliefs that are true may or may not intersect with the subclass of a believer's beliefs that are justified (hence, there's no reason to think that rational believers are (necessarily, a priori) knowers. So far, nothing has been said that suggests that believers must be

knowers (that this is guaranteed by the interpretationist framework). And certainly, an appeal to the truth (or fact) of evolution is utterly irrelevant in these matters.

It is important to note that, for Dennett, “believers are rational” is true a priori. And if this is so, then there’s no need to appeal to the fact/truth of evolution and natural selection (for the same reasons as above). What light, after all, could evolution shine on this matter? Could evolution falsify the statement that “believers have mostly true beliefs”? Could an empirical survey of bachelors EVER falsify the statement that “all bachelors are unmarried”? Surely, we’d be more likely to think that our empirical data were incorrect, and this is because the nature of “bachelor” is not settled empirically. Similarly, what investigation into the evolutionary origins of cognitive organisms could EVER falsify the statement that “believers have mostly true beliefs”? So, it’s not at all clear how evolution could ever come to bear on the conceptual issue in question and the a priori truth that believers have mostly true beliefs (are rational).

[essentially, the end of the paper and it’s way too long for a conference paper... but it gets the ideas out there... what follows is speculation leading to further papers...]

For Dennett, the notion of rationality is an “enabler”, it is a background assumption that enables us to make sense of a system as a cognitive system. Rationality, for Dennett, is *constitutive* of cognition. Dennett would say that it only makes sense within the context of presuming rationality to speak of a system as cognitive.

For Sober, rationality is an *evaluative measure* of a cognitive system (specifically, of the inferences made by a cognitive system). Sober would say that rationality can apply (in degrees) or fail to apply (in degrees) to cognitive systems.

These differences begin to shed light on the fundamental difference between Dennett and Sober with respect to rationality.

Second, we need to ask: in virtue of WHAT are believers rational? (Cf.: In virtue of what are bachelors unmarried?). A number of proposals.

1. Perhaps any one with a least one belief is rational? That doesn’t seem right (and it’s far from interesting). It’s not even clear that believers could ever have just one belief. There seems to be some minimal, critical mass of beliefs.
2. Perhaps any one with mostly true beliefs (that is, on Dennett’s account, any believer) is rational. Would it be proper to say that a believer is rational in virtue of having mostly true beliefs? If this is so, then

someone with 51% true beliefs is a rational agent, regardless of which inferences, if any, they make; regardless of how many contradictory beliefs they hold; regardless of how poor they are at choosing effective (practical) means to reach their ends; regardless of whether or not their sheer stupidity gets them killed; etc. (See Cherniak's *Minimal Rationality*). All of these possibilities are compatible with Dennett's proposal. There's something deeply unsatisfying about this simple answer. It also seems plainly wrong. It would force us to accept that all believers are, necessarily, rational. But it seems that the way we use the words "rational" and "irrational" does not follow this. If Dennett is correct, then our use of "rational" and "irrational" should parallel our use of "believer" and "non-believer", respectively. I don't think our current use reflects such a parallel application of terms. For instance, Dennett's view does not allow for a particular group of agents, namely, *irrational believers*. This, on Dennett's view, is an oxymoron. We might as well say of a figure that it is a round square, or of a man that he is a married bachelor, or some such absurdity. Unlike round squares and married bachelors, however, the idea of an irrational believer doesn't conjure up such absurdity. And because of this, there's something unsettling about Dennett's proposal.

3. Rationality might be a label used to identify people (believers) who are approved sources of information. (In addition to being believers, the label 'rational' might pick out belief-forming processes, inferences, beliefs, etc.) Rationality, then, becomes a social practice—a game—of sorts. We, collectively, set the rules for the application of the label 'rational', 'irrational', and so on. This "rationality game" is worthwhile only if the rules for the application of such terms are rules that we can satisfy. (for example, the rule that "rational" applies only to believers with no contradictions in their belief-set would not be a useful rule for the rationality game—it would result in identifying no believers, except, perhaps God, as a rational agent who is an approved source of information. We must be capable of playing the Rationality Game, and so the rules that we (collectively) establish will be rules that we can satisfy. (if we are not able to satisfy the rules, the game loses its value, and goes out of existence... we can be pretty sure that the rules of valuable games will be of a certain sort).

A variation of proposal 3 is to ask: How "low" can you go? That is, part of filling out the proposal in 3 is to identify the lower threshold for rationality—for example, which inferences are minimally required (again, see Cherniak's *Minimal Rationality*).

But the overall approach of 3 addresses the question: What does the practice of identifying instances of rationality (whether believers or processes) *do* for us? What is it SUPPOSED to do for us?

And finally, I want to ask: What role, if any, does an appeal to evolution play in the Rationality Game? I suspect the answer to this is: not much, if any.

Some Thoughts about the “Rationality Game” (prelude to next chapter/paper where I offer my own positive account of rationality, ala Edward Craig (new paradigm stuff) and Ram Neta’s talk on “what do we want from a theory/concept of perception”)

So, the analogy to introduce, then, is that of a game. Take baseball for example. Baseball is a game in which agreed-upon rules and procedures are followed. There is a threshold according to which some activities or behaviors count as playing baseball (this threshold may be fuzzy and may change over time), other do not count as playing baseball.

But it seems correct to say that for the game of baseball, there is some threshold below which the activity taking place ceases to be baseball. Say, if the activity includes 16 players on each team, or if five bases and two home plates are used, or if there are four outs per inning and a 180 degree playing field, etc., then the activity is simply not an activity we call “playing baseball”.

Watching a tee-ball game brings this to light even more. A common alteration nowadays is for innings to be played such that all nine players get a turn at bat (in this case, hitting the ball from the tee). Put-outs do not matter, except as a way to prevent the other team from scoring runs. The team with the most runs at the end of five innings is the winner.

In this sense, the activities that these children engage in are not baseball—it is another game (“tee-ball” perhaps, but it is not *baseball* proper). It is not baseball because there is sufficient alteration of the rules and procedures (and even, perhaps, of the aim of the game itself). In other words, the behavioral input-output pairs are sufficiently different than those associated with the game of baseball. Of course, there’s a sense in which tee-ball helps youngsters to develop the skills required to play baseball someday, as well as instilling the virtues of sportsmanship and competition... but this is to get ahead of ourselves and begin to offer an answer to the question: Why do we have our children engage in the behaviors known as “tee-ball”?

So, it seems reasonable to maintain that there is some minimum (though perhaps vague) threshold below which some set of activities fails to count as *playing baseball*, and this is because WE are responsible for coming up with this designation of “baseball” (as well as its rules). The game of baseball is OUR game, OUR invention—it has been refined slowly and socially. The rules that constitute *playing baseball* is a social construction, and THAT is why we can (collectively) establish some threshold or other for what counts as playing baseball. We may also conclude, a priori, that activity falling outside or below some standard is simply not baseball—I needn’t observe the activities common to a hockey game to know that such activities are not instances of playing baseball. Understanding the “grammar” of baseball (and the “grammar” of hockey) allow me to conclude that activities taking place during a hockey game are not instances of playing baseball.

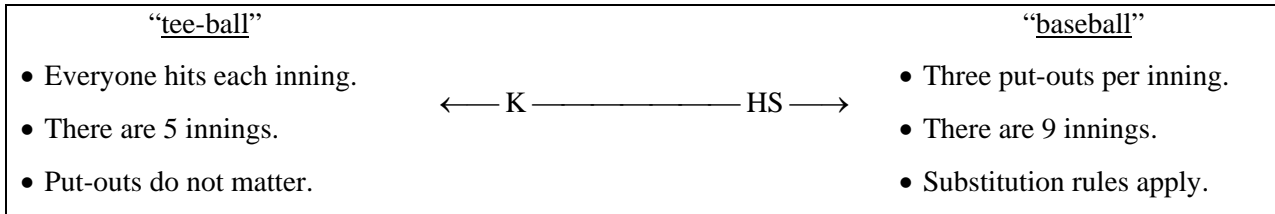


Figure 4.

Figure 4 represents the continuum between “tee-ball” on the far left and Major League “baseball” on the far right. Arguably, a set of activities falling to the left of the “K” not only is a case of playing baseball poorly, it also fails to be a case of playing baseball at all. “K” represents a threshold, to the left of which we no longer say the activities are instances of playing baseball.

“HS” might represent the place on the continuum where we place the activities engaged in by high school athletes. In high school baseball, there are only seven innings, but put-outs matter, and some substitution rules apply. Other rules that apply include regulations as to the total number of innings a high school athlete can pitch each week, etc., and these are rules not found in Major League baseball. So, we might be more likely to accept activities falling at the HS threshold to be “baseball” (because it bears a close enough family resemblance to Major League baseball), and those activities falling to the left of K not to be regarded as “baseball” at all (because such activities are so remotely unfamiliar to baseball, it is another game entirely, called “tee-ball”).

None of this addresses the question: why engage in these activities at all? Why engage in identifying these activities? These concerns will also play a big part in how we interpret various behaviors and activities.

This, I take it, is Dennett’s view with respect to cognition and the ascription of mental content. Belief and rationality are human (social) inventions. What it is to believe and what it is to be rational is for one’s behaviors to be interpretable in a certain way. To ascribe mental content to another organism is to play the “rationality game”. And like baseball, there are certain rules to the rationality game, there is a certain “grammar” to what it takes to be a believer. What it takes, according to Dennett, are the appropriate behavioral input-output pairs.

So, if the set of activities (behaviors) an organism displays fails to comply with the rules and procedures of belief and rationality, or if the activities (behaviors) fall below some minimal (though vague) threshold, then we may conclude that the organism in question is no longer playing the rationality game (and as a consequence, they are no longer playing the belief game either). It’s not just that such an organism (who falls below some threshold) is playing the rationality game poorly. That organism isn’t even playing the rationality game. According to Dennett, we play the rationality game by adopting the intentional stance.

Further, we are justified in rejecting some organism's set of activities as qualifying that organism as a *cognitive* organism (as a believer), and that is because the interpretive framework (the details of it) are OUR creation.

So, we can say, just as we did with baseball, that for some set of observed behaviors or activities generated by an organism, if the behaviors/activities fall below some minimum standard of what it counts to be rational, then that organism ceases to be rational and so it ceases to be a believer... it ceases to be a cognitive organism.

(Just as a child's activity might fail to meet some minimum standard of what it takes to be playing baseball, such that below that standard the child is not merely playing baseball poorly, the child is not even playing baseball in any form.)

(Just as an organism's activity might fail to meet some minimum standard of what it takes to be a rational believer, such that below that standard the organism is not merely failing to make certain inferences (or, inferring poorly, or engaging in irrational behavior), the organism is not even identifiable as having beliefs (and as a system without beliefs, it's not proper to say that it is irrational. There is a threshold in the rationality game below which an organism is no longer reasoning poorly... the organism is not even reasoning anymore.)

There is no further fact of the matter, discoverable by objective measurement, that will overturn this—it is a consequence of accepting that the rules governing the ascription of cognitive content (beliefs) and the determination of rationality are socially constructed. It's OUR game; we developed (and inherited) the rules. The conceptual constraints of such an interpretive framework make it true, a priori, that believers will be rational. If one drops below a certain measure of rationality, then one ceases to be a believer. If a child drops below a certain measure of executing the activities we call playing baseball, then it ceases to be playing baseball—it isn't just playing baseball poorly, it isn't even playing baseball. Similarly, at some level, an organism is not just reasoning poorly, the organism ceases to be reasoning at all (and, on Dennett's view, the organism thereby ceases to be a cognitive organism).

it is this connection that Dennett claims obtains via a priori argument (no amount of evolutionary savvy will establish this *conceptual* connection... what evolution provides is a behaviorally complex organism and the basic ability to apply an interpretation—ironically, the “ability to apply an interpretation” is something that we must ascribe to another organism... it's not something established outside of the rationality game itself). Given this, it is a mystery why Dennett asserts that “Natural selection guarantees that *most* of an organism's beliefs will be true, *most* of its strategies rational” (1987, p. 75).

So, according to the rationality game, it is a priori true that believers will be rational... and that is because it is a consequence of applying an interpretation (a consequence of adopting the intentional stance). What is the minimal point at which an organism ceases to be a believer? Well, that can be debated... it's okay for it to be fuzzy... and neither I nor Dennett needs to answer that question here.

What this shows, however, is that the game of rationality is socially constructed... and there is no deeper truth, beyond the scope of the game and the rules of interpretation, according to which organisms can be said to be believers or said to be rational. "Believer" and "rational" are labels that make sense only within the interpretive framework (i.e., the intentional stance). These labels, according to Dennett, don't belong to the biological sciences in the way that Sober assumes.

So, for mental traits such as cognition and mental representation and belief and rationality, there is no fact of the matter that transcends the interpretive framework being applied to the organism in question. Sober, however, tries to smuggle cognition into the biological realm; he tries to treat cognition on a par with respiration. Sober doesn't so much as refute Dennett as he changes the subject. So, Sober has not provided us with any reason to reject the interpretationist framework endorsed by Dennett and Davidson—Sober has simply registered his disagreement.

So, what is this game we call rationality? And how is it related to belief, justification, knowledge, etc.?

Why engage in this game at all? What does it allow us to do? How might engaging in this game increase our inclusive fitness? If engaging in such a game DOES increase fitness, what role does evolution play in such a proposal?