A Defence of AI-Functionalism Against Brandom's Arguments from Holism and the Frame Problem

REINER SCHAEFER *University of Guelph*

ABSTRACT: Brandom argues that functionalism must ultimately fail because it will not be able to explain how we can holistically update our beliefs solely in terms of abilities possessed by non-linguistic things. In this paper I respond to this argument by arguing that non-linguistic animals encounter and overcome an analogous sort of holistic updating problem. I will also try to demystify holism and de-intellectualize language use/reasoning.

RÉSUMÉ: Brandom soutient que le fonctionnalisme doit ultimement échouer parce qu'il ne saurait expliquer comment nous pouvons actualiser nos croyances de façon holistique uniquement en termes de capacités possédées par des objets non linguistiques. Dans cet article, je réponds à cet argument en soutenant que les animaux non-linguistiques rencontrent et surmontent un problème semblable d'actualisation holistique. Je vais aussi tenter de démystifier le holisme et de désintellectualiser l'utilisation du langage et le raisonnement.

Brandom argues that functionalism must fail ultimately because it will not be able to explain how we holistically update our beliefs solely in terms of abilities possessed by non-linguistic things. The sort of holism that language users encounter is supposedly unique to language users, and therefore, only language users will have the abilities needed to overcome the problems (such as the frame problem) that arise from this holism. In this paper, I will argue that non-linguistic things do in fact engage in a sort of holistic updating that is closely analogous

***Dialogue* 50 (2011), 741–750.**

doi:10.1017/S0012217312000030

to the holism involved in belief updating. Therefore, any difficulties relating to holism that must be overcome by linguistic things must also be overcome by non-linguistic things, and that means that functionalists should be able to appeal to this non-linguistic holistic updating when explaining reasoning and language use. Along the way, I also hope to demystify holism and de-intellectualize reasoning and language use.[1]

AI Functionalism and Necessary Abilities for Language Use

Functionalism (or specifically what Brandom calls "pragmatic AI functionalism") consists of the claim that we can analyze the ability to use language by decomposing it into an arrangement of primitive non-linguistic abilities. Brandom calls this sort of decomposition, *algorithmic decomposition* (or inversely algorithmic *elaboration*). The primitive non-linguistic abilities being elaborated should not be thought of as unanalyzable; they are merely primitive in relation to the complex language-using ability. It is necessary for AI functionalism that the primitive abilities it appeals to be non-linguistic in the sense that non-linguistic things could possess them. If language use is decomposed into primitive abilities that are themselves linguistic (abilities that only language users have), then the resulting decomposition cannot be considered *substantive*—they would simply amount to explaining language use in terms of language using abilities and not be philosophically interesting on their own.[2]

Brandom argues against AI functionalism in two steps. First, he argues that there is a certain ability that must be a member of any set of primitive abilities that can be algorithmically elaborated into the complex language using ability. Second, he argues that only linguistic things have that ability. If Brandom is correct on both these points, then he will have shown that there is no set of non-linguistic abilities that can be algorithmically elaborated into the ability to use language, thereby falsifying AI-functionalism.[3]

As we delve into the first stage of Brandom's argument, we find that there are at least three abilities that he takes to be necessary for language use. First, if someone is able to use language, then she must be able to distinguish some performances as having the significance of assertions. But it is not possible for someone to recognize a performance as an assertion without also treating it as having an inferential significance. What it means to treat a performance as having inferential significance, then, is treating it as justifying certain other assertional commitments and as sometimes requiring justification. For example, a vocalization of "The cat is on the mat" is treated as an assertion by taking it to justify assertions like "There is a mammal on the mat" or "There is probably fur on the mat," and by treating it as sometimes justified if it has also been claimed that "I hear meowing at the door and there is a mat at the door." Language users can of course disagree about which inferences are correct and which are not, but for our purposes (and Brandom's) it does not matter whether the judgments made about inferences are correct by some higher standard. What matters is that something can only be a language user if it treats some inferences as correct and others as incorrect through the way it demands, gives, and accepts justification for assertions. We see then for Brandom that asserting and inferring come as a package.[4] They amount to the first two of the three abilities necessary for language use.

But while Brandom makes heavy use of what he calls 'material inference' in most of his philosophical accounts, his argument against AI-functionalism does not depend upon it or upon his inferentialist semantics. For our purposes, we can understand inferential relations simply as relations of evidential support between assertions (or between beliefs). The argument here only requires that inference and assertion-making abilities be necessary for language use. They need not be sufficient. But what is additionally required for Brandom's argument is that inferential relations and beliefs be holistically interrelated, because, as we shall see, the third of the abilities necessary for language use involves overcoming problems that arise from holism. And again, while holism is a consequence of Brandom's inferentialism, one can accept the relevant sort of holism without being an inferentialist (Quine is a good example).

Holism and the Frame Problem

What makes inferential relations holistically interrelated is the nonmonotonicity (or defeasibility) of much of our reasoning. For example, whether it is correct to make the inference from "The dry, well-made match was struck" to "The match ignited" depends upon what else is the case. If the match was in a strong electromagnetic field then the inference would not be good, unless it was also in a Faraday cage, but the Faraday cage won't help if there is no oxygen in the room. Most of the inferences we accept have defeasibility conditions, and these also have defeasibility conditions and so on, until every one of our beliefs could make a difference to the appropriateness of every inference we endorse or could endorse. Therefore, if someone is to treat certain performances as having inferential significance, she must make some distinction between correct and incorrect inferences relative to her set of background commitments (beliefs), which act as collateral premises. With one set of collateral commitments, she would treat the inference from *p* to *q* as good, but there are other sets of collateral commitments in the context of which she would not treat the inference as good.

This sort of holism has, of course, been a worry for many in the AI field, because it suggests (to some people at least) that any time we learn something new, we would have to evaluate the impact that the new belief has upon every other actual or potential belief, and upon the inferential relations between these beliefs. It may turn out that the new belief is a defeasor of an inference that was previously taken to be good and which justified our holding one or more of our other beliefs. Here is where the frame problem begins to rear its head.

Holism on its own might not be too much of a problem if the algorithmic model we were constructing had a very limited range of possible beliefs (or a limited ontology), and had preset limitations and heuristics for what could be inferentially relevant to what else. This is frequently what is done in AI, which is why the frame problem is more of a philosophical problem than a technical problem today.[5] But Brandom, following Fodor somewhat, further characterizes the ability to use language in a way that rules out being able to impose such convenient limits. (BSD 81-82) Part of being able to use language is, in principle, being able to generate infinitely many new predicates which could in turn allow us to relate anything to anything else. Fodor, for example, writes,

> Consider a certain relational property that physical particles have from time to time: the property of BEING A FRIDGEON. I define 'x is a fridgeon at t' as follows: *x is a fridgeon at t iff x is a particle at t and my fridge is on at t*. It is of course a consequence of this definition that, when I turn my fridge on, I CHANGE THE STATE OF EVERY PHYSICAL OBJECT IN THE UNIVERSE; namely, every physical particle becomes a fridgeon. (Fodor, 144)

The concern is that the same combinatorial productive resources that allow us to generate useful concepts like 'match' and 'electromagnetic field' also allow us to generate useless concepts like 'fridgeon.' Given that it is a necessary condition for something to have these combinatorial productive resources if it is to be a language user, it follows that we cannot in advance limit the range of possible beliefs and the inferential relations between them. Every change in belief about something can have an impact, in principle, on every other belief we have about anything else and the inferential relations between them. Because of inferential holism, it follows that the inferential relations that our concepts stand in include those relating to some (if not all) of the infinitely many complex relational properties we could generate.

We can call this sort of holism 'unbounded holism,' because there can be no principled limits on what could be inferentially relevant to what else, and because there is no specifiable limit on the set of considerations that one could attend to in principle. This is largely the result of there not being any limits on what collateral premises could be in play. Unbounded holism is contrasted with the less threatening 'bounded holism,' which specifically limits the range of considerations that can be inferentially related to others, or limits the number of considerations that could in principle be attended to.[6]

It is of course practically impossible for language users to explicitly reassess every belief and inference every time any change in belief occurs. No doubt many of the complex relational properties that a language user can generate will not be relevant to the goodness of many of the inferential relations that actually matter to the agent. This, argues Brandom, suggests that language users must have an additional ability such that for many of the inferences that they could attend to, they can distinguish in practice which beliefs are relevant to its goodness and which are not, thereby allowing language users to ignore much of what they could in principle attend to when updating beliefs. If I cannot ignore most of the complex relational properties that I could attend to, then I . . .

> . . . am accordingly obliged to check every one of my beliefs and the inferences that support them to see whether they are infirmed by those facts–to be sure that my conclusion that the solid floor will bear my weight is not affected by its suddenly consisting of fridgeons…For any complex relational property such as being a fridgeon or having old-Provo-colored eyes, we can describe *some* inferential circumstances (however outré) in which the credentials of some significant claim would turn precisely on the presence or absence of that property. (BSD 82)

This introduces the third ability that Brandom takes to be necessary for using a language. Very generally, the relevant ability is overcoming unbounded holism. More specifically in the case of the frame-problem, the ability is ignoring many (probably most) of the considerations that one could attend to in principle. But Brandom's argument against AI functionalism is not that language use cannot be algorithmically decomposed because of the frame problem (or more generally, unbounded holism). In fact, Brandom claims that such an algorithmic elaboration can be given, but only if one of the primitive abilities involved in the elaboration is the ability to ignore many of the complex relational properties that one could attend to. Rather, Brandom's argument is that that ability is only possessed by linguistic creatures, and therefore, the AI functionalist cannot give a *substantive* algorithmic decomposition of language use.

But this raises two questions: why must ignoring be a *primitive* ability? and why is ignoring a *linguistic* ability? I do not think Brandom gives a very clear answer to the first question. The relevant sort of ignoring involved in language use, according to Brandom, amounts to being able to distinguish in practice which changes in beliefs would affect the correctness of an inferential relation and which would not affect it. In a footnote, he suggests that we have no idea how to algorithmically decompose this ability into simpler non-linguistic ones. (BSD 83) But beyond this, he says little else. I suspect he says what he says because he does not think that we can algorithmically decompose the abilities that generate and manage unbounded holism into non-holistic abilities. But this is not something I will directly address in this paper.

Brandom gives a clearer answer to why the relevant sort of ignoring is specifically a linguistic ability. He argues that

> Only something that can *talk* can [ignore a vast variety of considerations one is capable of attending to], since one cannot *ignore* what one cannot *attend* to (a PP-necessity claim), and for many complex relational properties, only those with access to the combinatorial productive resources of a *language* can pick them out and respond differentially to them. (BSD 82)

Roughly, the idea is that only things that can generate complex relational predicates can generate and attend to an indefinitely large class of considerations that are holistically interrelated. Because only linguistic things can generate complex relational predicates, only linguistic things can encounter the unbounded holism that gives rise to challenges such as the frame problem. As a result, only linguistic things will have the abilities (such as ignoring) that are necessary for overcoming these challenges.

Updating our Beliefs about Animals' Updating Practices

I will now rebut Brandom's argument by demonstrating that non-linguistic things can plausibly be thought to generate and attend to an indefinitely large class of considerations that are holistically interrelated. It would follow that whatever difficulties (such as the frame problem) are encountered and overcome by linguistic things because of unbounded holism should *analogously* be encountered and overcome by intelligent non-linguistic things. Therefore, whatever abilities non-linguistic creatures have that allow them to overcome or avoid the problems of unbounded holism can be used as primitive abilities in a substantive algorithmic decomposition of language use. If I am correct, then it makes no difference whether non-linguistic things are able to generate and attend to complex relational predicates.

In presenting a case for my claim that intelligent but non-linguistic animals can generate and attend to an indefinitely large class of considerations that are holistically related to each other, I will make use of some of the under-appreciated ideas developed in Mark Bickhard's and Loren Terveen's interactivist approach to cognitive science.[7] Presumably intelligent animals will interact with their complex environments in real time and must "choose" their interactions well if they are to survive and reproduce. For our purposes, we can say of what is normally called "action" and "perception" that they are both organism interactions (hereafter, O-interactions). For example, a lioness may move her head and focus her eyes or she may move her paw against the ground, but in either situation, she is O-interacting with her environment. O-Interactions typically give feedback that allow the creature to differentiate its environments, and intelligent creatures can learn which sorts of feedback are reliable indicators of whether it is appropriate to initiate some other O-interaction in order to achieve some goal.

The relationship between the feedback of some O-interaction and the appropriateness of initiating some other O-interaction (a relationship that can be called an O-interaction strategy) can be seen as closely analogous to an inference. Just as many inferences are defeasible, so too are O-interaction strategies. For the lioness it is often a good strategy to pounce on a young mammal nearby in order to get food, but if the lioness differentiates her environment as one that also includes the large protective parents (such as adult elephants) then the strategy will be defeated (no longer considered appropriate). In principle, any feedback from any O-interaction may be relevant to the appropriateness of any given O-interactive strategy. The result is a sort of holism that is analogous to that of inferential relations.[8]

Now the crucial question is whether non-linguistic creatures can generate an indefinitely large class of considerations to which they are able to attend (or differentiate). If they cannot, then the sort of holism they encounter might be of the bounded sort, and therefore, might not raise the sort of difficulties we are concerned with here, such as the frame-problem. But there seems to be good reason to think that intelligent animals like lions or dogs can generate an indefinitely large class of considerations, and therefore, must deal with an unbounded holism. The range of O-interactions that a lioness can initiate is very large, and she can combine various O-interactions to form more complex ones whose feedback can be related to the initiating of any other complex of O-interactions. A pet dog can learn that it is not allowed to be on the bed and that it will not be punished if its master is not home, and that the slamming of a certain kind of car door is an indicator that the master has returned home. It seems that the dog must have learned this largely by combining and relating various O-interactions—that is by generating complex relational O-interactions.

If all this is correct, then non-linguistic things must generate and manage a sort of unbounded holism that is closely analogous to the sort involved in language use. We also know that non-linguistic animals are not incapacitated by their dealings with unbounded holism and therefore, if this holism does raise any serious difficulties, then non-linguistic things have the abilities necessary for overcoming them. This can be shown rather clearly in the case of the frame problem, where a plausible story can be told such that intelligent but non-linguistic creatures can be said to overcome the frame problem by being able to appropriately ignore many of the considerations to which they could in principle attend.

For example, while stalking a young zebra, the lioness does not consider whether she should spin around in a circle three times or whether it is relevant that the pebble next to her left paw is smaller than the zebra. The lioness *could* attend to such considerations in the sense that she *would* attend to them if she were unfortunately captured and put in a circus, where she had to spin in circles in order to get food. This ability to ignore is not relevantly different from the layperson's ability to not consider fridgeons unless they are lured into philosophy classes where they must learn to discuss them if they are to pass their philosophy exam, or participate in some philosophical discussion about the frame problem. The lioness does and should ignore the possibility of spinning around in circles in most circumstances, just as language users in most circumstances do and should ignore the possibility that something is composed of fridgeons.

De-Intellectualizing Language Use

We should not be surprised at the conclusion that non-linguistic animals do not attend to everything that they could in principle attend to. Did we ever think that non-linguistic things were even slightly inclined to do otherwise—even if they couldn't "ignore" many of the possible considerations that they could attend to? Probably not. Then why should we think rational language users would be so inclined? I suspect it is because we philosophers often want to understand us humans as exhibiting some sort of ideal rationality, but do not expect the same from non-linguistic (and hence nonrational) animals. We are okay with the idea that intelligent, non-linguistic, nonrational creatures attend to the considerations that they do as a result of biological processes and conditioning which cannot necessarily be justified by some higher standard of reasoning. But we loath to think that we humans' attending to some considerations rather than others could be unprincipled in just the same way. By and large I think this is largely because, as Brandom points out, a significant part of our being language-users consists of our taking one another to have obligations to justify our beliefs and actions when appropriately challenged to do so.[9]

Problems like the frame problem start to arise when we not only want to be able to provide reasons to justify our beliefs, but want to be able to justify the operations of the various primitive abilities that give rise to the more complex ability to use language and update our beliefs. From this perspective, it is not enough that we are able to update our beliefs; we must be able to update them in an epistemically principled way. Holism, as the view that every consideration is potentially relevant to every other consideration, seems threatening precisely because it suggests that we have no principled reason to attend to some considerations rather than others when we update our beliefs. Because we have no principled reasons for attending to some considerations rather than others, we are rationally obliged to attend to all of them—to make sure that we don't miss anything.

But AI functionalism does not require that language use be algorithmically decomposable into primitive abilities that act according to rationally justifiable principles. It does not even require that the primitive abilities look anything like abilities that we could intentionally and reflectively carry out. Furthermore, it runs against the grain of much of Brandom's work to assume otherwise. For instance, on Brandom's account, the primary way that language users endorse an inference is simply by being disposed to update beliefs in particular ways.[10] Whether or not these sorts of dispositions are the result of processes or abilities operating according to rationally justifiable principles is irrelevant.

We should therefore set aside the more epistemological concerns about whether or not we can rationally justify the operations of the primitive abilities involved in our attending to the considerations that we do when we update our beliefs. A more promising approach consists of treating the language users' ability to update beliefs as being a sophisticated extension of the non-linguistic ability to organize interactions with an environment to satisfy various interests (which for language users will be largely social). Once we adopt this approach then holism ceases to be a problem for AI functionalism, and there is no longer a need for an ability to *ignore* that goes beyond merely *not attending*. This is because there is no longer a presumption that language users have to have principled reasons for not *actually* attending to any of the considerations that they *could* have attended to. While language users *could* attend to any consideration when updating their beliefs, what considerations they will *actually* attend to (and thereby take to be relevant) will likely be determined by their environment, their conditioning, and their biological (including social) interests—not principles of good reasoning that they might espouse to justify themselves in conversations or reflections about what should be attended to.

In a nutshell, if we want to try to algorithmically decompose the ability to use language into non-linguistic abilities, then we should think of language users as being intelligent animals responding to their largely social environment, not epistemologists responding to an imaginary sceptic who claims that we cannot ultimately justify our attending to some considerations rather than others when we update our beliefs. I am not claiming that de-intellectualizing the primitive abilities comprising language use (including belief updating) will *solve* the very difficult question of why language users attend to the considerations that they do when updating their beliefs. But I think it does put AI functionalism on a much more promising path.[11]

Notes

[1] The argument being addressed is found in Brandom's *Between Saying & Doing: Towards an Analytic Pragmatism*. For brevity, I will later refer to this text as *BSD*. On pages 234-235, Brandom states that BSD represents a distinct project from the one he undertakes in *Making It Explicit*, but he also says it overlaps in many ways. The argument I present here is directed solely at what is said in BSD (particularly chapter three) but is influenced by my readings of *Making It Explicit*.

[2] I qualify that they are not interesting *on their own* because Brandom does think we can give a useful analysis of language use by characterizing our practices that deploy objective modal vocabulary in terms of subjective normative vocabulary. Such an analysis is enlightening, claims Brandom, because normative vocabulary is sufficiently different from modal vocabulary (even if it is not strictly expressively weaker). See BSD Chapter Six.

[3] To be fair, Brandom actually admits that he cannot give a knock-down argument against AI-functionalism. (BSD, 79) Rather, he is attempting to show why we should be pessimistic about its success.

[4] See BSD Chapter Two Section 3, especially page 42.

[5] See Shanahan, 2009.

[6] The distinction between unbounded and bounded holism is not intended to correspond with the distinction between bounded rationality and unbounded rationality. The latter distinction is primarily a matter of reasoning according to ideal standards of rationality or merely reasoning the best one can with the limited computational resources available.

[7] I could have given a very similar presentation using Brandom's own Test-Operate-Test-Exit cycle presented in BSD chapter six (and in the appendix to chapter two), but the TOTE account involves some problematic commitments and requires more explanation than is presently worth giving.

[8] It is only for ease of discussion that I say that the lioness differentiates her environment *as having this or that feature*. I deny that non-linguistic things have our concepts like "parent," "large," "young," or "stock market crash" in terms of which to categorize

their environments. Animals (or anything for that matter) can reliably react to stuff without having our concept of the thing to which they are reacting. O-interactions can be characterized purely in terms of their function within the organism. Something need not know why or how they obtain the feedback they do. All that matters is that the feedback of an O-interaction is related in certain ways with other O-interactions (and their feedback). This differs from the externalism of Brandom's TOTE cycles, which would require that interactions be specified in terms of the objects of the interaction.

[9] Brandom does not emphasize this as much in BSD, but it is central to his more thorough account of language in *Making It Explicit* (Chapter 3 in particular).

[10] This is crucial to Brandom's more thorough account of language in *Making It Explicit*, discussed in great detail in the first three chapters. It is present, though perhaps less central, in BSD (for instance see pp. 44-45 and pp. 119-120).

[11] Much gratitude goes to Mark McCullagh for various helpful comments on an earlier draft of this document.

References

Bickhard, Mark, and Terveen, Loren

1995 *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. Elsevier Scientific.

Brandom, Robert

1994 *Making It Explicit: Reasoning, Representing, & Discursive Commitment*. Cambridge: Harvard University Press.

2008 *Between Saying & Doing: Towards an Analytic Pragmatism*. New York: Oxford University Press.

Fodor, Jerry

1987 "Modules, Frames, Fridgeons, Sleeping Dogs, & the Music of Spheres." In *The Robot's Dilemma*. Ed. Zenon Pylyshyn. Norwood: Ablex. pp.139–149.

Shanahan, Murry

2011 "The Frame Problem." In *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta. Winter 2009 Edition. Web.