Mind Games

A Reply to Daniela Hill

Chris Eliasmith

In her discussion of my original article, Hill reconstructs an argument I may have been making, argues that the distinction between natural and artificial minds is not exclusive, and suggests that my reliance on behaviour as a determiner of "mindedness" is a dangerous slip back to philosophical behaviourism. In reply, I note that the logical fallacy of which I'm accused (circular reasoning) is not the one present in the reconstruction of my argument (besides the point), and offer a non-fallacious reconstruction. More importantly, I note that logical analysis does not seem appropriate for the discussion in the target article. I then agree that natural and artificial minds do not make up two discrete categories for mindedness. Finally, I note that my research program belies any behaviourist motivations, and reiterate that even though behaviour is typically important for identifying minds, I do not suggest that it is a substitute for theory. However, the target article is not about such theory, but about the near-term likelihood of sophisticated artificial minds

Keywords
Artificial minds | Behaviourism | Logical analysis | Minds

Author

Chris Eliasmith
celiasmith@uwaterloo.ca
University of Waterloo
Waterloo, ON, Canada

Commentator

Daniela Hill
dhill@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

I think Hill is right to wonder aloud about my methodology in the target article. After all, I just ignored the hard philosophical issue of saying what minds really are. I pretended (somewhat self-consciously) that we all know what minds are, and so that we will simply be able to tell when someone has created one, if they ever do. But, I did that for a reason. The reason was this: I did not want to get lost in the minutiae of metaphysics when my focus was on a technological revolution—one with significant philosophical consequences (which is also not to say I don't like such minutiae in their proper time and place).

2 A failure of logic

However, Hill was also not especially taken by the reasons I provided for expecting such developments either. Hill's suggestion is that the best reasonable argument you could construct from my original considerations was fallacious. Though, Hill is quick to point out that I didn't take myself to be constructing an argument: "... not to claim that Eliasmith really argues for the emergence of artificial minds" (this collection, p. 4).

Nevertheless, her analysis is that what I have provided is best understood as a *petitio principii* (aka circular argument): "this means that the conclusion drawn at the end of the ar-

Eliasmith, C. (2015). Mind Games - A Reply to Daniela Hill.
In T. Metzinger & J. M. Windt (Eds). *Open MIND:* 12(R). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570788

gumentative line is identical with at least one of the implicit premises" (Hill this collection, p. 4). Unfortunately, the technical analysis offered (p. 5) is a *non sequitur* (i.e., there is no logical connection between the premises and conclusion). Regardless, one fallacy is as embarassing as the other.

However, I'd like to suggest that if we wanted to recast the original paper as a logical argument, then a simple *modus ponens* will do: if we have a good theory and the technological innovations necessary to implement the theory, then we can build a minded agent. We have good (and improving) theory and will have the proper technological innovations (in the next 50 years), therefore we will be able to build a minded agent (in the next 50 years). Indeed, most of the paper is arguing for the plausibility of these premises.

More to the point, however, I think that we can take this as an object lesson for when logical inference is really just the wrong kind of analysis of a paper. Instead of trying to provide a logical argument from which the conclusion necessarily follows from the premises, I am providing series of considerations that I believe make the conclusion likely given both the current state of affairs, and expected changes. In short, I think of the original paper as providing something more like a series of inferences to the best explanation: all of which are, technically, fallacious; and all of which are directed at establishing premises.

3 Back to minds

Despite disagreeing with the analysis of the logical structure of the paper, I do appreciate the emphasis that Hill has placed on philosophical and ethical aspects of our attempts to construct minds. In the original article, I only very briefly touch on those issues. However, I would be quick to point out that I do not think, and never intended to suggest, that the distinction between "natural" and "artificial minds" was an absolute, "exhaustive," or "exclusive" one (Hill this collection, p. 3). Like most interesting and complex features, possession of 'mindedness' no doubt comes in degrees. In fact, I think that our attempts to construct artificial minds will provide a much better sense of the dimensions along which such a continuum is best defined.

Finally, I must admit that I find it somewhat alarming that I'm being characterized as a behaviourist in Hill's article—*that* has definitely never happened before: "Let us see how Eliasmith characterizes artificial minds. One can see this as a judgment based on the similarity of behaviour originating from two types of agents: humans and artificial" (this collection, p. 5). Hence, I was espousing "analytical behaviorism… a failed philosophical research program" (Hill this collection, p. 6). Indeed, I, like all behavioural scientists, believe that behaviour is one important metric for characterizing the systems of interest. However, the reason I focus on internal mechanisms in my own research – all the way down to the neural – is that I believe those mechanisms give us critical additional constraints for identifying the right class of algorithms that give rise to behaviours. Consequently, I wholeheartedly agree with Hill that "There could be much more to mindedness than behaviour" (this collection, p. 6). So, for the record, I believe that our best theories for how to build minds are going to be highly informed by low-level mechanisms.

That being said, I also think that most people's judgments of whether or not something counts as being minded is going to come down largely to their being convinced of the naturalness, or "cognitiveness" of the behaviour that is exhibited by agents we construct. Notice that there is a difference between a claim of how people will judge mindedness, and a claim about theories of mindedness or how we ought to best achieve that judgment. Turing was, after all, onto *something* with his test.

4 Conclusion

I noted in the original article (Eliasmith this collection) that I was attempting to avoid becoming mired in tangential debates regarding what it is to have a mind by simplifying the criteria for mindedness (for the purposes of that article). Exactly the kinds of debates I was attempting to avoid are raised in Hill's comment-

ary. For example, I don't think we know if there is a clean contrast between a "fully minded agent" and a "merely cognitive agent" (Hill this collection, p. 6). Perhaps there is, and perhaps it is that a fully minded agent can "experience herself as a cognitive agent", (Hill this collection, p. 6) but perhaps not. This does not strike me as a decidable question at present.

So, perhaps my unwillingness to venture into the murky waters of necessary and sufficient conditions for having a mind came off as making me look like a behaviourist. But in truth, my purpose was rather to focus on providing a variety of evidence that I think suggests that artificial minds are not as far away as some have assumed. There is, I believe, a historically unique confluence of theory, technology, and capital happening as we speak.

References

Eliasmith, C. (2015). On the eve of artificial minds. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Hill, D. (2015). Future games - A commentary on Chris Eliasmith. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.