Argument against Strong AI formalised

Samuel Tscharner

University of Basel, Switzerland

22. September 2018

The following proof is an attempt to formalise an argument from Peter Godfrey-Smith, which he brings forward in his article "Mind, Matter, and Metabolism" [1]. To put it in a nutshell, he argues that *there is no AI with human-like subjective experience* because the AI-systems usually envisaged in functionalistic thought experiments concerning consciousness (resp. subjective experience) have obviously different fine-grained functional properties than humans have.

It is to mention that the proof does not coincide accurately with the shape of Godfrey-Smith's argument, but is more a logical model of it. Nevertheless I tried to keep it close to the original text while only using the most important coherences. Therefore, the two premises are not to find exactly like this in the text, but are an essence from what is written in it.
The assumption of the first subproof is the definition of strong AI in the relevant sense for the argument. On one hand it seems to be a specific criterion put forward by Godfrey-Smith that strong AI should have human-like subjective experience. On the other hand it seems that Godfrey-Smith only read the criterion out of the functionalistic thought experiments. So it might be a premise or an assumption. If we regard the AI-definition as a premise the conclusion would be that *there is no strong AI-system with human-like subjective experiences*. Now, with taking the definition as an assumption the conclusion comes out as a conditional that *if the definition holds, then there is no strong AI-system with human-like subjective experiences*. Both ways make the point, the latter is just more humble. The second assumption is, according to Godfrey-Smith's article, a premise about strong AI which many functionalists hold in their thought experiments. All in all, the premises (step 1 and 2) should be understood as propositions which are hold to be facts within Godfrey-Smith's theory. The definition of AI, like it is placed now as an assumption (step 3), can be understood as consensual definition, which probably still can be changed without a big impact on the whole theory. So, there can still be something which is called strong AI but is defined differently. The second assumption (step 4) then has to be interpreted as a thesis of a competing theory which is shown to be incompatible with Godfrey-Smith's theory.
The rest is simple calculation. I applied mostly the rules of fitch-style calculus of natural deduction. However, to abbreviate the proof I used some other deduction rules which are well-known to be valid in classical first order predicate logic. It is surely possible to develop a calculus-style that makes the different roles of the set and assumed propositions better recognisable but that is a problem for another paper.

The spirit of this undertaking is to find ultimately clear rules for the formalisation of natural language arguments. The attempt at hand might be a interesting example for further inquiry. Many question can be raised on its basis.
The idea is that clear rules for the formalisation of natural language arguments might facilitate the whole procedure of science by providing a much better overview of debates. Not to mention that a lot of argument or even theory examination could eventually be done by computer programs. However, for this it has to be clear what are premises and what are assumptions of an argument or a theory.

[1] Godfrey-Smith, Peter, 2016, Mind, Matter, and Metabolism, in *The Journal of Philosophy* 113, 481-506.

Definition of used Predicates

Predicate (x)	Meaning
HSE(x)	x has human-like subjective experience
HFG(x)	x has human fine-grained functional properties
HCG(x)	x has human coarse-grained functional properties
AI(x)	x is a system with strong AI
Met(x)	x has a metabolism

Proof:

1. $\forall x(\mathsf{HSE}(x) \rightarrow (\mathsf{HFG}(x) \wedge \mathsf{HCG}(x)))$	
2. $\forall x(\mathsf{HFG}(x) \rightarrow \mathsf{Met}(x)$	
3. $\forall x(\mathsf{AI}(x) \leftrightarrow (\mathsf{HSE}(x) \wedge \neg\mathsf{Met}(x)))$	
4. $\exists x(\mathsf{AI}(x) \wedge \mathsf{HCG}(x) \wedge \neg\mathsf{HFG}(x))$	
5. $\boxed{a}$ $\mathsf{AI}(a) \wedge \mathsf{HCG}(a) \wedge \neg\mathsf{HFG}(a)$	
6. $\mathsf{AI}(a)$	$\wedge$ **Elim:** 5
7. $\mathsf{AI}(a) \leftrightarrow (\mathsf{HSE}(a) \wedge \neg\mathsf{Met}(a))$	$\forall$ **Elim:** 3
8. $\mathsf{HSE}(a) \wedge \neg\mathsf{Met}(a)$	$\leftrightarrow$ **Elim:** 7, 6
9. $\mathsf{HSE}(a)$	$\wedge$ **Elim:** 8
10. $\mathsf{HSE}(a) \rightarrow (\mathsf{HFG}(a) \wedge \mathsf{HCG}(a))$	$\forall$ **Elim:** 1
11. $\mathsf{HFG}(a) \wedge \mathsf{HCG}(a)$	$\rightarrow$ **Elim:** 10, 9
12. $\mathsf{HFG}(a)$	$\wedge$ **Elim:** 11
13. $\neg\mathsf{HFG}(a)$	$\wedge$ **Elim:** 5
14. $\perp$	$\perp$ **Intro:** 13, 12
15. $\perp$	$\exists$ **Elim:** 4, 5–14
16. $\neg\exists x(\mathsf{AI}(x) \wedge \mathsf{HCG}(x) \wedge \neg\mathsf{HFG}(x))$	$\neg$ **Intro:** 4–15
17. $\forall x\neg(\mathsf{AI}(x) \wedge \mathsf{HCG}(x) \wedge \neg\mathsf{HFG}(x))$	**DeMorgan1:** 16
18. $\forall x(\neg\mathsf{AI}(x) \vee \neg\mathsf{HCG}(x) \vee \mathsf{HFG}(x))$	**DeMorgan2:** 17
19. $\forall x(\mathsf{AI}(x) \rightarrow (\mathsf{HCG}(x) \rightarrow \mathsf{HFG}(x)))$	2x$\rightarrow$ **Def:**18
20. $\boxed{a}$	
21. $\mathsf{AI}(a)$	
22. $\mathsf{AI}(a) \rightarrow (\mathsf{HCG}(a) \rightarrow \mathsf{HFG}(a)$	$\forall$ **Elim:** 19
23. $\mathsf{HCG}(a) \rightarrow \mathsf{HFG}(a)$	$\rightarrow$ **Elim:** 22, 21
24. $\mathsf{AI}(a) \leftrightarrow (\mathsf{HSE}(a) \wedge \neg\mathsf{Met}(a))$	$\forall$ **Elim:** 3
25. $\mathsf{HSE}(a) \wedge \neg\mathsf{Met}(a)$	$\leftrightarrow$ **Elim:** 24, 21
26. $\neg\mathsf{Met}(a)$	$\wedge$ **Elim:** 25
27. $\mathsf{HFG}(a) \rightarrow \mathsf{Met}(a)$	$\forall$ **Elim:** 2
28. $\neg\mathsf{HFG}(a)$	**modus tolles:** 27,26
29. $\neg\mathsf{HFG}(a) \vee \neg\mathsf{HCG}(a)$	$\vee$ **Intro:** 28
30. $\neg(\mathsf{HFG}(a) \wedge \mathsf{HCG}(a))$	**DeMorgan3:** 29
31. $\mathsf{HSE}(a) \rightarrow (\mathsf{HFG}(a) \wedge \mathsf{HCG}(a))$	$\forall$ **Elim:** 1
32. $\neg\mathsf{HSE}(a)$	**modus tolles:** 31,30
33. $\mathsf{AI}(a) \rightarrow \neg\mathsf{HSE}(a)$	$\rightarrow$ **Intro:** 21–32
34. $\forall x(\mathsf{AI}(x) \rightarrow \neg\mathsf{HSE}(a))$	$\forall$ **Intro:** 20–33
35. $\forall x(\neg\mathsf{AI}(x) \vee \neg\mathsf{HSE}(x))$	$\rightarrow$ **Def:**34
36. $\forall\neg(\mathsf{AI}(x) \wedge \mathsf{HSE}(x))$	**DeMorgan3:** 35
37. $\neg\exists x(\mathsf{AI}(x) \wedge \mathsf{HSE}(x))$	**DeMorgan4:** 36
38. $\forall x(\mathsf{AI}(x) \leftrightarrow (\mathsf{HSE}(x)\neg\mathsf{Met}(x)) \rightarrow \neg\exists x(\mathsf{AI}(x) \wedge \mathsf{HSE}(x))$	$\rightarrow$ **Intro:** 3–37

Q.E.D.