Why the embryo rescue case is a bad argument against embryonic personhood

1. Introduction

The “Embryo Rescue Case” (ERC) refers to a thought experiment that is used to argue against the view that embryos have a right to life (i.e. are persons). I will argue that cognitive science undermines the intuition elicited by the ERC; I will show that whether or not embryos have a right to life, our mental tools will make it very difficult to believe that embryos have said right. This suggests that the intuition elicited by the ERC is not truth indicative. The upshot of this is that we have an undercutting defeater for our intuition that embryos do not have a right to life. In what follows, I will first explain the ERC and how it is used to argue that embryos lack a right to life. After this, I will briefly explain the relevant cognitive science, and argue this undermines the intuition produced by the ERC.

2. The Embryo Rescue Case

The Embryo Rescue Case (ERC) is an thought experiment meant to show that embryos are not persons; it is meant to show that embryos do not have a right to life.1 It goes something like this:

You are in a hospital. The building is on fire. On one end of the building, there is one 5 year old child, on the other end, there is a petri dish holding 100 embryos. You only have enough time to save one of them. What should you do?2

The answer that is (alleged to be) obvious is that we should save the child. Therefore, we ought to conclude that a child is far more valuable than an embryo. But, if the embryo had a right to life, then surely we should do the opposite: if given the choice between saving 100 persons and saving 1 person, we should save 100.

A skeptic might respond to the ERC by granting that we should save the child, but explaining it in a way that is consistent with embryos being persons. Christopher Kaczor has

---

1 For the purposes of this article, I will take an organism to be a person if and only if it has a right to life.
done precisely this. He argues that contextual factors can make it such that we should save one human over another, but that this does not show that the human who we should not save does not have a right to life. For example, if given the choice between saving the president or saving a civilian (or a group of civilians), we should save the president. But this does not entail that the civilian is not a person; rather, it merely shows that there can be reasons to save one person over another. Applied to the ERC, this means that showing that we should save the child over the embryos does not entail that embryos do not have a right to life, for there may be reasons to do so (e.g. the child can feel pain whereas the embryos cannot). And hence the ERC does not tell us anything interesting about the moral status of embryos. However, Kate Greasley has argued that this reply misses the point. She claims that

given the strength and prevalence of the judgment that you must rescue the [child], the favored response by the “personhood-from-conception” theory is to attempt to explain the ERC intuition in ways that are consistent with the moral equivalence of embryos and human babies...[but this response fails] to dislodge the very compelling notion that the reason for our common conclusion that one ought to save the [child]...rather than the embryos is our belief that the moral status of the [child] vastly outweighs that of any embryo...And if the ERC intuition is far more plausibly explained by our conviction that embryos lack equal moral standing with born humans, then, insofar as we find that conviction irresistible, we will have grounds to disbelieve the “personhood-from-conception” thesis.5

So, Greasley thinks that rendering the idea that we should save the child over the embryos consistent with the embryos being persons is irrelevant. Rather, what we must do is explain why we have the intuition or conviction6 that there is a vast difference between the moral status of born humans and embryos; this strong conviction, she thinks, gives us reason to doubt that embryos are persons. Of course, when Greasley talks about ‘our conviction’ that there is not an equal moral standing between embryos and born humans, she assumes that this conviction, or intuition, is held by nearly all persons. How she knows what convictions everyone has, we are not told (has she conducted a study to find this result, or is it just convenient for her to assume?).

---

4 In this case, the reasons might include that killing the president would destabilize the country whereas killing the civilian (or group of civilians) would not.
6 I will use these terms interchangeably.
Putting that issue aside, I will grant, for the sake of argument, that everyone has an intuition about there not being an equal moral standing between embryos and born humans. However, I will argue below that this intuition is undercut by facts about the cognitive makeup of human beings, and hence this gives us no reason to think that embryos are not persons.

3. Undermining the ERC Intuition

For the sake of argument, I have granted that there is, indeed, a widely held intuition (produced by the ERC) that there is a vast difference in the moral status between born humans and embryos. In what follows, I will briefly describe the nature of undercutting defeaters. After this, I will briefly explain some contemporary cognitive science, and argue that it provides an undercutting defeater for the intuition produced by the ERC.

3.1 Undercutting Defeaters

Suppose that a person, call her Sally, walks into a widget factory. She walks over to the assembly line to admire the widgets being produced. The widgets appear to her to be red, and she forms the belief “these widgets are red” are the basis of this appearance. After Sally forms this belief, the owner of the factory, Sarah, walks over to Sally and tells her about the new red lights that she had installed over the assembly line. Sally notices the lights, and notices that they are shining directly on the widgets. Once Sally realizes this, she acquires a defeater for her belief that the widgets are red, for she has reason to think that the basis of her belief (i.e. the widget appearing red) is not indicative of its truth (i.e. the widget actually being red). Let us call this (i.e. having reason to think that the basis of one’s belief is not truth-indicative) an undercutting defeater. An undercutting defeater does not give one reason to think her belief is false; rather, it means that one should withhold belief. More generally, if S believes B on account of it appearing that B, and S later comes to hold another belief Q that entails that it appearing that B to S is not indicative of B’s truth, then S has an undercutting defeater for B, and she should give up (but not reject) B.

3.2 Cognitive Science, Defeaters, and the ERC

Persons - beings with a right to life - are at least agents. That is, persons are things that do not merely operate mechanistically, but are capable of self-propulsion for the sake of their own goals. Or, more modestly, the persons that human beings are acquainted with (through sense perception) are agents; we do not know of any non-agent persons. So, if humans have a hard

---

7 Here, I make use of Michael Bergmann’s concept of an undercutting defeater, see Bergmann, Michael. (2005). "Defeaters and Higher-Level Requirements" The Philosophical Quarterly (55): 419-436. The argument I offer below is compatible with other models of undercutting defeaters, such as John Pollock’s. See Pollock, John. (1986). Contemporary Theories of Knowledge. Rowman and Littlefield.
time attributing *agency* to a thing T, they will inevitably have a hard time attributing *personhood* to T.\(^8\)

### 3.2.1 Cognitive Science

In cognitive science, there are two different types of beliefs distinguished: reflective and non-reflective.\(^9\) The former are held explicitly and are what we most often refer to as ‘beliefs’. For example, my belief that Canada borders America is a reflective belief; or, my belief that the sum of two even numbers will always be an even number is a reflective belief. Non-reflective beliefs, conversely, are those beliefs that are formed automatically, without conscious effort. For example, my belief that I am not able to walk through walls is non-reflective; or, my belief that things that move in a goal-oriented manner are agents is non-reflective. The relationship between reflective and non-reflective beliefs may seem distant, but this is not so. Our non-reflective beliefs will often be a *default* for our reflective beliefs. For example, if we were to present a new object to a group of persons and were to ask them if it would fall down when released, this group would form the reflective belief (and would answer) that it would fall down. However, their reflective belief would simply be the result of them explicitly stating their non-reflective belief, which is that objects fall down if they are not being held up.\(^10\)

Another mental tool of ours, theory of mind, ascribes mental states to objects that are deemed by our hypersensitive agency detection device (explained below) to be agents. So, when we see a person leave a room and come back with an object, we form the non-reflective belief that they left the room to retrieve that object. For example, if Sally saw Sarah leave the room and return with a drink she would, if asked why Sarah left the room, reply by reading off her non-reflective belief: she would say that Sarah left the room to get a drink; her theory of mind attributed mental states to Sarah to account for her leaving the room and returning with a drink. In this case, Sally will have read off her non-reflective beliefs that were produced by her theory of mind.\(^11\) So, reflective beliefs often turn out to be based on non-reflective beliefs. However,

---

\(^8\) This is a testable hypothesis. I claim it on its *prima facie* plausibility, but it is certainly possible that it be falsified. However, given the mental tools that human beings possess (explained below), it seems very unlikely that this is incorrect.

\(^9\) Sometimes, these are called Type 1 and Type 2 beliefs, and sometimes non-reflective beliefs are simply referred to as ‘intuitions’.


non-reflective beliefs also play a further role: they make reflective beliefs seem more plausible. That is, if a reflective belief matches well with many of our non-reflective beliefs, it will strike us as being plausible; the more non-reflective beliefs it matches, the more plausible it will seem. So, why do we hold certain reflective beliefs? According to Barrett, when “no obvious reason presents itself to discard the nonreflective belief, we accept it as our reflective belief.”

Now, given our mental tools, human beings have an easy time believing that there are agents; indeed, it is so easy to believe that there are agents that Barrett calls this mental tool the hypersensitive agency detection device, or HADD. So, HADD produces beliefs in other agents extremely easily (illustrated below). But when does it produce belief in other agents? To detect invisible agency, all that is needed is for an event to occur that the subject cannot easily and readily explain by appealing to non-agental factors. For example, if a person hears a bush rustle at night, HADD will quickly produce the belief that it was caused by an agent, making it likely that the subject will flee the scene. But being produced by HADD does not entail that the belief will be sustained, for if HADD produces the belief that an agent was the cause of x, but none (or few) of our other mental tools generate non-reflective beliefs about this agent, then it will be easily discarded. For example, if I only hear the bush rustle once and notice no other signs of agency, I will have little need to attribute agency to the noise; my theory of mind will not produce non-reflective beliefs about the purported agent responsible, and so it will have little support. So, if I notice that the wind is blowing, I will easily discard my belief that an agent caused the rustling and attribute the noise to the wind (a non-agent). But of more relevance to our topic is agent attribution to visible objects. What really turns the gears of HADD is (apparent) self-propelled movement. That is, if an object appears to move itself in a purposeful manner, HADD will kick in and produce the belief that the object is an agent, and then theory of mind will attribute various mental states to the object to account for its movement. Waytz et al, instead of speaking of self-propelled movement, speak of causal uncertainty. They say “[c]ausal uncertainty is one of the basic triggers of the goal for prediction and control. When a car behaves in a perfectly predictable way in response to your actions, it seems mindless; but when it starts lurching forward while braking, or stalling while starting, then your car might seem to have a


mind of its own.” So, apparent self-propulsion or causal uncertainty will cause HADD to kick in. Indeed, apparent self-propulsion is so effective at activating HADD that humans will attribute agency to *shapes* that move in such a manner. As Barrett succinctly puts it, “when HADD perceives an object violating the intuitive assumptions for the movement of ordinary physical objects (such as moving on non-inertial paths, changing direction inexplicably, or launching itself from a standstill) and the object seems to be moving in a goal-directed manner, HADD detects agency.” Another factor that contributes to attributing agency is *fear*. This is because, Epley et al argue, fear makes us more attentive to our surroundings, and more willing to attribute agency to things (or to the causes of events). So, in summary, HADD makes it extremely easy to attribute agency to things that move (or make noise) in a seemingly self-propelled, goal-oriented manner, and it is even easier if we are in a fearful state.

As mentioned in section 3.2, persons are (at least) agents - or, at least, the persons that we know of are agents. So, if an organism is a person, then it is an agent. And, on the other hand, if an organism is not an agent, then it is not a person - or, at least, it will be hard to view it as a person. This is where HADD becomes crucial: it is very difficult for humans to attribute agency to *embryos*. This is for (at least) two reasons: (a) embryos do not appear (to human eyes) to be self-propelled or goal-oriented and (b) humans do not need to appeal to the agency of embryos to explain phenomena. In light of (a) and (b), it is to be expected that HADD will not attribute agency to embryos; while HADD is indeed *hypersensitive*, it still requires a trigger to activate it, and embryos do not act as a trigger, even in thought experiments like the ERC. Indeed, given (a) it is highly counterintuitive to attribute agency to embryos; it violates our non-reflective belief that agents are self-propelled and goal-oriented (since an embryo does not appear to be so), and it finds no support in our other non-reflective beliefs (our theory of mind will not attribute mental states to embryos since they are not non-reflectively thought to be agents). Thus, the thesis that the embryo is a person will inevitably strike us as implausible and hence it will strike us as intuitive that the embryo is not a person. So, given our mental tools and the nature of embryos, it is to be expected that it will be intuitive that embryos do not have a right to life. And this result will transfer directly over to the ERC: when imagining the scenario, it will seem highly implausible, given our mental tools, that the child and the embryos are of equal value. Thus,

---

19 Perhaps a developed fetus will be able to activate HADD through screening technology (or through imagination in thought experiments). My point here strictly concerns embryos (and perhaps fetuses very early on).
crucially, even if embryos have a right to life, we will find this thesis implausible; whether or not embryos are of equal worth to children, the ERC will elicit the intuition that there is a vast difference in value between the two. So, granting that Greasley’s claim is right, it is not at all surprising to find that people have a ‘strong intuition’ that there is a vast difference in the moral status of embryos and of born-humans: given our mental tools, it is to be expected.

3.2.2 An Undercutting Defeater for the ERC

We saw above that, given our cognitive tools, it will be highly intuitive to think that embryos are not persons even if they actually are persons; that is, even if the embryo is a person, our cognitive tools will make it seem intuitive that they are not. Moreover, there is no good reason to think that the only beings that have a right to life (i.e. persons) that exist are those that our cognitive tools easily recognize.20 This, however, gives us an undercutting defeater for the intuition produced by the ERC: it gives us reason to doubt that the intuition produced by the ERC is truth-indicative. Thus, cognitive science undercuts the evidential force of the intuition elicited by the ERC. So, in the same way that Sally should not believe the widgets are red based on their appearing red to her, we should not believe that embryos are not persons based on the intuition produced by the ERC: in both cases, the basis of belief is dubious. Put in terms of evidence, let “E” refer to the intuition (produced by the ERC) that there is a vast moral difference between a born human and an embryo, and let “H” refer to the hypothesis that the embryo has a right to life. On the relevance theory of evidence, a proposition p is said to be evidence for another proposition q if and only if the probability of p given q is greater than the probability p given the denial of q.21 However, given our mental tools, the probability of E given H is not greater than the probability of E given the denial of H;22 rather, they are equal. This is because, given our cognitive makeup, E will occur on both H and the denial of H. And hence, in terms of the relevance theory of evidence, the ERC does not provide evidence that an embryo is not a person; the ERC offers no evidence against the thesis that embryos are persons.

Abstracting away from cognitive science, the general point is this: given that embryos do not appear to us to move in a self-propelled way (or give us any other signs of agency) and do not resemble born humans,23 it is to be expected that it will be difficult to attribute equal moral

\footnote{20 I.e. there is no reason to think that all beings with a right to life that exist will trigger HADD.}
\footnote{21 It can be formally stated as: Pr(p/q) > Pr(p/~q).}
\footnote{22 I.e Pr(E/H) is not greater than Pr(E/~H).}
\footnote{23 Francis Beckwith (Beckwith, Francis J. (2007) Defending Life: A moral and Legal Case Against Abortion Choice. Cambridge University Press) argues that the reason people find it intuitive that we should not save the embryos is that they do not look like us. Dean Stretton (Stretton, D. (2008). "Critical Notice-Defending Life: A moral and legal case against abortion choice by Francis J Beckwith" Journal of Medical Ethics (34): 793-797) replies to this by making the ERC even more bizarre: it is now between adult aliens (that are cognitively like adult humans) that look like turnips. He claims that since we would still choose to save an alien over a petri dish of embryos, it cannot be that not looking like a born human is what elicits the intuition produced by the ERC. It is not clear how Stretton
worth to them. But this does not tell us anything interesting about their actual moral worth since we have no reason to think that the only things that have moral worth are those things that resemble humans or (detectably) move in a self-propelled way; rather, it just tells us about our own cognitive tools. Of course, none of this shows that an embryo does, in fact, have a right to life. Rather, it is just to say that the ERC is a bad argument for the non-personhood of embryos. Opponents of embryonic personhood need to look elsewhere for support for their position.\textsuperscript{24}

4. Conclusion

In this article, I have argued that the contemporary cognitive science suggests that whether or not the embryo is a person it will appear that it is not. This is because if an object is not an agent, then it is not a person. And, furthermore, it is very difficult to attribute agency to an embryo, so it will be very difficult to attribute personhood to it. And hence that we find it difficult to believe that embryos are persons is no surprise; indeed, it is expected. I argued that this provides us with an undercutting defeater for the intuition elicited by the ERC: we have reason to doubt that that intuition is truth-indicative. Therefore, I conclude, the embryo rescue case is a bad argument against embryonic personhood; it gives us no reason to think that embryos are not persons. While it very well may be that embryos are not persons, the ERC gives us no reason to endorse this view.\textsuperscript{25}

5. References\textsuperscript{26}

knows how most people would react to his turnip example (has he done a study to find out, or is it just convenient for him to assume?), but I digress. What is crucial is that my argument against the ERC differs from Beckwith’s, since I employ cognitive science, and in particular HADD, to undercut the intuition elicited from the ERC. Stretton’s reply to Beckwith, even if successful, does not touch my argument: whereas Stretton’s critique is concerned with perception, my argument is concerned with other cognitive tools. More crucially, his alien/turnip thought experiment does not conflict with any claims that I have made in this article: his aliens would still activate HADD while embryos do not, and hence his thought experiment will face the same problem as the original ERC.\textsuperscript{24} This conclusion is important since personhood is often the central issue of abortion debates: most philosophers seem to think that the morality of abortion is contingent on whether the fetus is a person. Of course, some deny this, e.g. Jarvis-Thomson, Judith. (1971). “A defense of abortion”, \textit{Philosophy and Public Affairs}, 1: 47–66, and Hendricks, Perry. (2018). “Even if the fetus is not a person, abortion is immoral; or, the impairment argument” \textit{Bioethics}: 1-9.

\textsuperscript{25} A referee notes that there is some similarity in my argument to evolutionary debunking arguments about morality (e.g. Street, Sharon. (2006). ‘A Darwinian Dilemma for Realist Theories of Value,’ \textit{Philosophical Studies} (127): 109–166) and other evolutionary debunking arguments, such as Alvin Plantinga’s evolutionary argument against naturalism (e.g. Plantinga, Alvin. (2011) \textit{Where the Conflict Really Lies: Science, Religion, and Naturalism}. Oxford University Press). The former threatens our moral knowledge since, it is argued, evolution via natural selection was not concerned with producing true moral beliefs, only beliefs that improve fitness. The latter threatens naturalists (those who affirm that there is no God or anything at all like him) that accept evolution with self-defeat, since evolution, it is argued, does not care about belief content (and hence truth), but only survival. The difference between my own argument and these evolutionary debunking ones is that I have not claimed that all moral intuitions are undercut, nor is evolution is doing any work in my argument. Rather, I have argued only that the intuition elicited by the ERC is not good evidence for the non-personhood of embryos.

\textsuperscript{26} Thanks to G.L.G. - Colin Patrick Mitchell - for particularly valuable comments on this article.


De Cruz, Helen, and J. De Smedt. (2015). A Natural History of Natural Theology. MIT.


