AI & Soc (2012) 27:431–436
DOI 10.1007/s00146-011-0333-7

ORIGINAL ARTICLE

On the irrationality of mind-uploading: a rely to Neil Levy

Nicholas Agar

Received: 12 March 2011 / Accepted: 20 July 2011 / Published online: 20 August 2011

Abstract In a paper in this journal, Neil Levy challenges Nicholas Agar's argument for the irrationality of mind-uploading. Mind-uploading is a futuristic process that involves scanning brains and recording relevant information which is then transferred into a computer. Its advocates suppose that mind-uploading transfers both human minds and identities from biological brains into computers. According to Agar's original argument, mind-uploading is prudentially irrational. Success relies on the soundness of the program of Strong AI—the view that it may someday be possible to build a computer that is capable of thought. Strong AI may in fact be false, an eventuality with dire consequences for mind-uploading. Levy argues that Agar's argument relies on mistakes about the probability of failed mind-uploading and underestimates what is to be gained from successfully mind-uploading. This paper clarifies Agar's original claims about the likelihood of mind-uploading failure and offers further defense of a pessimistic evaluation of success.

Keywords Mind-uploading · Strong AI · Pascal's Wager

1 On the irrationality of mind-uploading: a rely to Neil Levy

In a paper in this journal, Neil Levy (Levy 2011) challenges my argument (Agar 2010: chapter 4) for the prudential irrationality of mind-uploading. Mind-uploading is a futuristic process that involves scanning brains and recording relevant information which is then transferred into a computer.[1] According to its advocates, uploading transfers both human minds and identities from biological brains into computers. Uploaded humans will enjoy benefits of enhanced cognition unavailable to those who retain their biological brains. While mind-uploading is not currently possible, advances in computer technology could make it available soon.

I call my argument against the rationality of mind-uploading, Searle's Wager. This name acknowledges two philosophical precedents. First, it recognizes John Searle (Searle 1980), the best-known critic of the program of Strong AI—the view that it may someday be possible to build a computer that is capable of thought. It also acknowledges Blaise Pascal (Pascal 1995) who, lacking proof of God's existence, presented his (in)famous Wager Argument for the prudential rationality of belief even when in doubt about God's ontological status.

Searle's Wager imagines candidates for mind-uploading being asked to place a bet. The success of mind-uploading is contingent on the truth of Strong AI. If Strong AI is a correct view then the procedure may work. Uploaded humans can enjoy a variety of enhancements denied to biological humans. Conversely, if Strong AI is a false view, then no computer could ever serve as a receptacle for a human mind. Mind-uploading inevitably fails. I argue that even those convinced by the philosophical arguments for Strong AI and therefore of the possibility of mind-uploading should allow that there is a non-negligible chance that they are, in fact, mistaken. I combine the claim that there is a significant chance that mind-uploading will

This paper is a response to Neil Levy's paper: "Searle's wager" (AIS Vol. 26.4).

N. Agar (✉)
Philosophy Programme, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
e-mail: nicholas.agar@vuw.ac.nz

[1] For recent advocacy of mind-uploading, see Kurzweil (2005) and Sandberg and Bostrom (2008).