AI & Soc (2016) 31:171–190
DOI 10.1007/s00146-015-0620-9

ORIGINAL ARTICLE

Simulation, self-extinction, and philosophy in the service of human civilization

Jeffrey White[1]

Received: 16 February 2015 / Accepted: 25 September 2015 / Published online: 30 October 2015

Abstract Nick Bostrom's recently patched "simulation argument" (Bostrom in Philos Q 53:243–255, 2003; Bostrom and Kulczycki in Analysis 71:54–61, 2011) purports to demonstrate the probability that we "live" now in an "ancestor simulation"—that is as a simulation of a period prior to that in which a civilization more advanced than our own—"post-human"—becomes able to simulate such a state of affairs as ours. As such simulations under consideration resemble "brains in vats" (BIVs) and may appear open to similar objections, the paper begins by reviewing objections to BIV-type proposals, specifically those due a presumed mad envatter. In counter example, we explore the motivating rationale behind current work in the development of psychologically realistic social simulations. Further concerns about rendering human cognition in a computational medium are confronted through review of current dynamic systems models of cognitive agency. In these models, aspects of the human condition are reproduced that may in other forms be considered incomputable, i.e., political voice, predictive planning, and consciousness. The paper then argues that simulations afford a unique potential to secure a post-human future, and may be necessary for a pre-post-human civilization like our own to achieve and to maintain a post-human situation. Long-standing philosophical interest in tools of this nature for Aristotle's "statesman" and more recently for E.O. Wilson in the 1990s is observed. Self-extinction-level threats from State and individual levels of organization are compared, and a likely dependence on large-scale psychologically realistic simulations to get past self-extinction-level threats is projected. In the end, Bostrom's basic argument for the conviction that we exist now in a simulation is reaffirmed.

Keywords Simulation · Model · Cognitive social science · Democracy · Brain in a vat · Skepticism · Global coordination problem

1 Introduction

> We must declare war on war, so the outcome will be peace upon peace.
> -Obama (2014)

> Fathers, provoke not your children to anger.
> -Fetyukovich[1]

Current events have my head spinning. I wake daily to read on the affairs of the world, only to wonder if it is all a dream. But for the suffering, from wedding parties bombed to civil activists singled out for government assassination, evidence invites speculation that our globe might well be part of a particularly violent and extraordinarily realistic video game, the product of a team of evil demons for their exclusive enjoyment at the expense of anyone with compassion and an eye to a flourishing future in which such existential threats are finally resolved. At least, in imagining it so, perceived reality is easier to accept.

As fanciful as this may seem, Nick Bostrom's recently patched "simulation argument" (Bostrom 2003; Bostrom and Kulczycki 2011) purports to demonstrate the probability that we exist now in a simulation of similar

✉ Jeffrey White
drwhite@kaist.ac.kr

[1] Humanities and Social Sciences, Korean Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

[1] The straight-talking counselor from The Brothers Karamazov as translated in Dostoyevsky and McDuff (2003, pp. 949–950).

complexity, an "ancestor simulation" of a period prior to that in which a civilization more advanced than our own—"post-human"—becomes able to simulate such a state of affairs as ours—"human." By this argument, one (or more) of the following propositions is (are) true:

> (1) The human species is very likely to go extinct before reaching a post-human stage.
> (2) The fraction of post-human civilizations that are interested in running a significant number of ancestor simulations is extremely small.
> (3) We are almost certainly living in a computer simulation (Bostrom and Kulczycki 2011, p. 54).

Proposition 1 expresses confidence in humanity's potential to destroy itself, as "we are likely to go extinct as a result of the development of some powerful but dangerous technology" (Bostrom 2003, pp. 251–252). Accepting 1, however, also implies that some human civilizations may not go extinct or extinguish themselves. They become "post-human," representing a situation in which "humankind has acquired most of the technological capabilities that one can currently show to be consistent with physical laws and with material and energy constraints" (Bostrom 2003, p. 245). From within such a situation, it follows easily that realistic simulations of "ancestor" low-tech level 1 worlds such as our own are a matter of course. Post-human civilizations interested will create limitless simulations, and as their simulations might also create simulations, the ratio of simulated to natural worlds spun out accordingly confirms near certainty in the third proposition, that this world, ours, is simulated.

A "patch" offered in 2011 (Bostrom and Kulczycki 2011) aims to correct for concerns that too few simulating populations survive proposition 1, and become 2, relative those surviving without such simulations. The concession that the patch offers is simply to accept the original conclusion so long as simulating post-human worlds are not unduly outnumbered by non-simulating worlds, in which case the simple probability delivers either ambiguous or opposite results. And with this, the reasoning driving Bostrom's simulation argument is quite clear—though there may be few civilizations which achieve a capacity to produce simulations of the complexity in terms of which we find ourselves today, some which do achieve that capacity will exercise it, and in exercising it will create many more simulated ancestor worlds like our own than could ever have existed otherwise. So, given the probability that simulated ancestor worlds far exceed natural alternatives, we must accept proposition number 3.

This is a contentious result for the skepticism that it seems to represent, skepticism that may be analyzed into two types, epistemic and axiological (Cogburn and Silcox 2014). Epistemic skepticism "is the traditional sort of worry about our knowledge of the external world" and axiological skepticism "is the concern that no genuine value could attach" to a simulated existence (Cogburn and Silcox 2014, p. 563). Together, these forms of skepticism seem to encourage an attitude that simulated life is—on the probability conferred by Bostrom's simulation argument—not real, and this may encourage the sentiment that suffering is—so long as it is not my own—inconsequential. Indeed, the very word "simulation" carries the sense that everything exists only as an "if" and is accompanied by the feeling "as-if" life mattered, "as-if" justice mattered, and so on and nothing more. We are "as-if" transported into the *Brothers Karamazov*, confronted by a maddening—and murderous—moral nihilism. Bostrom himself understands this implication, but reasons that "properly understood … the truth of (3) should have no tendency to make us 'go crazy'" (Bostrom 2003, p. 255). What we have taken for good should remain taken for good, and life should continue as it has, simulated or not.

My suspicions are that Bostrom does not take our simulated condition seriously enough. The following paper argues that the Bostrom-scale simulations under consideration may prove a necessity rather than a luxury for a civilization with post-human aspirations. The next step in the fugue confronts objections deriving from similarities between simulations and brains in vats (BIVs). The third and fourth steps build the complimentary case for psychologically realistic simulations in the near term, while the fifth confronts a different form of skepticism with a review of promising work in neurorobotics. Finally, the second half of the paper argues that the second of Bostrom's three propositions should be modified on grounds that pre-post-human civilizations like our own, "human," probably *require* the ongoing development of large-scale psychologically realistic social simulations in order to achieve and then to maintain a post-human status, thereby affirming his conclusion.

2 Skepticism and reality

> …the scientific spirit does not require us to blind ourselves to the practical consequences which hang upon the solution to not a few scientific problems.
> -Rashdall (1914, p. 199)

> How does it happen that a whole generation of scientific experts is blind to obvious facts?
> -Dyson (2015)

Confronted with the proposition that we exist as simulations, one may respond in a number of ways. One may

object to Bostrom's characterization of the nature of the simulated condition, for example. He writes that "it would suffice for the generation of subjective experiences that the computational processes of a human brain are structurally replicated in suitably fine-grained detail, such as on the level of individual synapses" (Bostrom 2003, p. 244).[2] This characterization is close enough to envatted brain archetypes to suffer similar faults, perhaps ending any further discussion about any simulated condition before it gets started.

To this end, Putnam (1982) famously argued for the necessary falsity of the proposition that we are brains in vats (BIVs). The crux is that a brain in a vat must not be able to think itself a brain in a vat, as to do so would be to deny the purpose behind putting brains in vats in the first place, this being to fool them into thinking themselves not envatted. So understood, to be able to resolve one's self as an envatted brain becomes confirmation of its impossibility, instead inviting the diagnosis of a different kind of delusion. In a similar spirit, Birch (2013) has attacked Bostrom for confidently projecting from an understanding of computation to the probability of a simulation. He charges Bostrom with exercising a "selective skepticism" toward more directly available information about the physical world that rather disconfirms its *ir*reality. Birch puts his criticism this way:

> In the early stages of the argument, Bostrom draws on empirical evidence to defend speculative claims about the potential power of post-human computing. In the latter stages, he assumes that my evidential situation with respect to the physical reality of my own hands is no better than my evidential situation with respect to the hypothesis that my cells contain some random stretch of junk DNA. To save his argument, Bostrom needs to explain how this remarkable conjunction of scientific realism and limb skepticism can be sustained (Birch, p. 101).

Expecting that such an explanation in not forthcoming, in other words, any belief in a simulated condition is delusion.

This sentiment is confirmed on the analysis of Huemer (2000). Huemer reminds us that any "claim of superiority" afforded a hypothesis is always relative alternative hypotheses (Huemer 2000, p. 409). Huemer's strategy is to demonstrate that BIV skepticism presumes an indirect rather than a direct realism in order to maintain the delusion that a BIV hypothesis is superior. He argues from direct realism against BIV skepticism, that we have "no grounds for suspecting that I'm a brain in a vat" as would be available given a common perceptual hallucination, and thus that "the presumption in favor of my perceptual belief that I have two hands" for example "remains undefeated" (Huemer 2000, p. 411). In the end, it is up to the skeptic to address this challenge from direct realism. My two hands and what I make with them are much clearer and more distinct than is any simulation; thus, the simulationist is either deluded, or wrong.

But, what if the simulationist weakens the argument, in some way? Anticipating such a move, Davies (1997) has recalled Graeme Forbes' (1985) contention that "the skeptic can evade Putnam's argument and achieve all she has ever wanted by switching to the hypothesis that I am 'relevantly like a brain in a vat'" (as quoted in Davies 1997, p. 51). This "revised skeptical hypothesis" reduces the relationship between our own condition and that of a BIV to one of intelligible analogy whereupon "it is enough that we can point to an 'intelligible instance' of the contrast by describing the situation of a BIV" (Davies, p. 52).

Similar to Huemer's approach, Forbes works from a comparison between two "epistemic positions" from which truths are differently evident, one for us and one for a BIV, with the view from one affording a grasp of certain truths which, due the nature of the other, are from it unassailable (cf. Davies, p. 58). This "privileged" position proffers Huemer's "claim of superiority" over less privileged hypotheses concerning an envatted condition. And on this model, again not unlike Huemers, Davies argues that a weakened strategy still fails because it fails to provide direct evidence for any relevant analog of an envatted condition.

For us to grant that our condition is relevantly like a BIV's is to accept that we embody the same essentially deluded condition of a BIV, that we stand in a deluded relation with the world, and thus that we are unable to provide reliable evidence for either a simulated or an unsimulated condition either way (cf. Davies pages 52 on the essential delusion of a BIV and also 54-6, objection 3). Thus, Davies concludes that the revised skeptical hypothesis:

> … is as devastating to our epistemic confidence as the original sceptical hypothesis, for, if we were in a situation relevantly similar to a BIV, then, even if we could know that we were not BIVs, we would in fact have 'no real understanding of the universe' in which we exist (p. 52).

In the end, any relationship between these two epistemic positions, ours and a BIV's, is predicated on the condition of a BIV, that any correspondence between a BIV's mental states—i.e., all of them—and the world as it really is—i.e., as we find it—is deluded, if not false then unjustified. Set up in this way, even a weakened BIV proposal makes knowledge impossible. And, as there appears to be no direct evidence that it is not we who inhabit Forbes' "privileged" epistemic position by default, the only

[2] Deepest gratitude to Peter Broedner for pointing to this passage.

alternative is to reject the claim that we are relevantly like brains in vats.

Birch takes a similar tact against Bostrom's simulation argument. Birch notes that "there is no reason to suppose that post-human civilizations would not radically mislead their simulated creations with regard to the true laws of physics, and the true properties of material substrates" with the result being a necessarily religious—and no longer selective—skepticism, "a pervasive *de dicto* scepticism about all aspects of physical reality, including those aspects epistemically relevant to the limits of post-human computation" (Birch 2013, p. 106). Read strongly, thus, Bostrom not only suffers a Cartesian deceiver, but seems to rest his argument in a Cartesian circularity. There is no way to limit his selective skepticism without presuming more than argument allows, the truth of one notion that shines most distinctly, that the simulated realizes the simulation as simulation, and from this recovers the "reality" of the simulation, itself. Bereft of any epistemic upshot, thus, the appeal of the skeptical strategy in any form appears lost.

However, there is a misestimation in Birch's reasoning. Birch asserts that "there is no reason to suppose" that simulators would not purposefully mislead the simulated, consistent with Putnam who frames the envatted brain within the intentions of the evil genius who put it there, and Davies who presumes that this relationship is the most "relevant," binding as it does the two epistemic perspectives constituting the revised skeptical hypothesis as set out by Forbes and as reflected more recently in Huemers. It is from the form of this relationship that Putnam is originally able to construe logically necessary grounds for the rejection of the BIV hypothesis. As brains in vats, we are *supposed* to feel as-if on two young legs in a field of flowers and green grass, not the way that we really are in fact, tethered to a keyboard most of the time, because it is of the purpose of a BIV to be deluded. The trouble with following this line of reasoning against simulations of the sort proposed by Bostrom is that we have at hand less evidence for evil geniuses out to delude us with simulations than we do that we might live in a simulation created by a different sort of scientist for a very different purpose, altogether.

Take for instance "virtual reality" (VR). Cogburn and Silcox (2014) argue against "brain-in-a-vatism"—the tendency to make a BIV out of what is not—by underscoring the value in exactly the kind of voluntary immersion in a simulated reality afforded by VR yet forbidden on the presumption of deceitful simulators. On their analysis, VRs afforded by digital computers have been more successful in "producing knowledge-how" than more passive literary and film vehicles. In part, know-how is afforded by the fact that VR permits the exploration of multiple paths of action—"a good video game allows the player to test the plausibility of a huge number of possible evolutions from a single given setup" (Cogburn and Silcox, p. 577)—while a traditional novel, for example, offers only one. Also on their analysis, this improved know-how not only leads to improved knowledge-that, but *grounds* it. "A good flight simulator teaches one how to fly" and also "gives one more true propositional beliefs" as "a side effect of how one's knowing-how constrains one's knowing-that" (Cogburn and Silcox, p. 577). This is to say that in entraining to an openly *simulated* reality, VR affords the learning agent increased opportunities to form and to test theoretical knowledge as well as practical acumen through proactive participation in an interactive environment, with the result being more robust and better acting agents in the real world. In no way is the common object of the author or inhabitant of these simulations served by deceit in the process.

Simulations teach us how to do things. Real things. Books do too but with books there is an extra step, the translation from symbolic expression to action routine, and here also there remains the trouble with enactive refinement and correction of errors. Books do not think on their own, fix grammar mistakes, rearrange paragraphs to create more coherent compositions. What about virtual cognition? What about simulating something like us? Can VR teach us how to think and act, or how to write better, in ways that older information technologies cannot? Beginning from the understanding that "Modern computers … seem to foreshadow future technologies that will eventually outpace the human sensorium itself" (p. 562), Cogburn and Silcox argue for the possibility of a robust virtual life, projecting "virtual humans" of a complexity similar to actual humans in relevant ways alongside other "intelligent creatures" (p. 571)—"genuinely thinking beings" (p. 572)—potentially imbued with moral standing (cf. p. 577). In such an environment, even real human beings may forge a life worth living, expressing virtues such as bravery and honesty "just as real" as those expressed by real human beings in the "real world" (p. 572). As for the value of the lives of "virtual humans" sharing in real human on-life stories, the authors afford no more speculation. But we may infer that testing for truth is exactly the sort of industry that an intelligent agent pursues, artificial or otherwise. And moreover we may infer that systems optimized for this task—i.e., problem solving—should be best suited to host such an agency. Excellence is habit. Simulation simply affords an arena in terms of which error can be minimized and training perfected whether trained agency is artificial or otherwise.

The next sections examine the limits on and the motivations behind current work that may seed the creation of Bostrom-scale artificially intelligent simulations in the future, solving problems and understanding what we mean when we say "genuinely thinking being." These

motivations are not deceptive, in counterexample to the presumed mad envatter characteristic of BIVism. After that, the fifth section meets objections that some feel cannot be overcome, that something essential to human cognition may be impossible to capture in a simulating medium like that of a digital computer.

3 Virtual realism

> Abbreviation is a necessary evil and the abbreviator's business is to make the best of a job which, though intrinsically bad, is still better than nothing. He must learn to simplify, but not to the point of falsification. He must learn to concentrate upon the essentials of a situation, but without ignoring too many of reality's qualifying side issues. In this way he may be able to tell, not indeed the whole truth (for the whole truth about almost any important subject is incompatible with brevity), but considerably more than the dangerous quarter-truths and half-truths which have always been the current coin of thought.
> -Huxley (1958)[3]

Quickly consider the computational demands of a "realistic" simulation of the sort considered in the last section, a virtual reality potentially outstripping the "human sensorium." Let us initially define a "realistic simulation" as any model operating on the most refined scientific understanding of whatever is under consideration. "Scientific" is doing a lot of work here, but we may allow this simply to mean that it satisfies the most sophisticated available tests for completeness and accuracy of account under practical constraints. Given that realistic simulations so defined are the products of scientists, open to scientific testing, to exist in a simulation is not identical with an existence outside of a simulation. Rather, it is something derivative of the aims of the scientist, and thus it is in terms of achieving this purpose that any "realistic" simulation must be evaluated.

To a degree, the more complex the simulation, the more realistic, and the more realistic the more useful. But in each instance, the judge is also the subject experiencing the simulation, and this judgment can change after entrainment to a more realistic environment representing as it does a movement from a less privileged to a more privileged epistemic position. For illustration, recall the last time that you saw an old sci-fi movie playing on television, and the sense that it conveyed with antiquated special effects. Compared to movies these days, those old effects are not "realistic" at all, though they sure seemed so the first time that we saw them. The same goes for the gamer who moves from chess or go to finely crafted miniatures wargames with various factions and strategic objectives; chess seems rather simple in comparison. In each case, it is in the step back from complexity and the cutting-edge realism that complexity affords that earlier efforts are ever qualified as "not-realistic" in some way. And conversely, it is in the step forward to greater realism that the complexity of the world is revealed.

"Realistic" is not the same as "real"—not for the simulator—it is something less that allows that simulator to do more. Take video games, for example. These are a surprisingly "green" technology allowing us to do more, personally, with less immediate impact on the energetic landscape in terms of which we are commonly immersed. In optimizing for some aspects of human experience while neglecting others, simulations allow us to pilot starships in our living rooms at considerably less cost than would be incurred by daily commutes across the Milky Way for mere amusement. Of course, this "more" that simulations deliver comes at the expense of complexity, with this complexity ideally inessential to the purposes for which the simulation is designed. So, simulations are "less" but they are less for a reason. Simple models representing only essential aspects of target systems—"less"—are more efficient than those representing dimensions inessential to the processes under view—"more."

Insofar as the simulation in question is a simulation of and for something like us, this complexity seems unnecessary during early stages. Simpler environments suffice. As the scale and rate of agent effected changes increase, complexity increases and realism becomes more demanding. Consider a man's skill in identifying the truth about something, before and after training. For a trained lapidary, a synthetic diamond is more or less an obvious fake, but for a naïve groom it may be worth a great deal more. "Not-realistic" is not the object of a "realistic" simulation any more than mistaking fakes is the object of the gemologist, or producing something essentially different from a diamond in the relevant ways is the object of either diamond synthesis or vow induction. "Realistic" is the object, coming at the cost of command over complexity amounting to realism in the relevant ways. Thus, there is a practical limit to realism, i.e., when fake diamonds cost more than real diamonds to produce. And, there is an epistemic limit, when the subject can no longer tell the difference. Within these limits, no matter how advanced a society, as long as this society is interested in what works and in what is true, we may understand "realistic" simulations as those which are close to the best that people can create with the technology available, close to the best that science can facilitate, with the "best" in every case being an ideally efficient and effective (e.g., time and energy saving), complexity-

[3] From the foreword.

reduced certain means to best ends in sight, i.e., delivering the most for the least.

"Close to the best" recognizes that limits emerge at the front of any press for realism in simulations of any sort. For Bostrom, this limit is practical, representing the point at which a simulation may prove "prohibitively expensive" as a post-human status is achieved, so that "we should expect our simulation to be terminated when we are about to become post-human" (Bostrom 2003, p. 253). And, this assessment is troubling. Recall that Bostrom's argument situates contemporary human civilization on the cusp of crisis in the first step, at the precipice between level 1 pre-post-human and level 2 post-human status, facing the likelihood of self-extinction through the mishandling of dangerous technologies. If Bostrom's projected limit is accurate, then there is no way forward. Either we stay as is, a result most worrisome given current global tensions, or we try to become post-human and have the plug pulled by our mad envatter. Which of these simulations are we in? Is there a third option? And, how can we tell the difference, any way?

Understanding that simulations are designed to fulfill a purpose, we may be able to identify which sort of simulation we may inhabit by first divining the likely purpose of our simulation. And to this end we may ask, in terms of our own situation, what are the likely needs that may motivate the development of simulations such as those employed for amusement by level 2 civilizations on Bostrom's account?

John Kultgen, writing for Concerned Philosophers for Peace in 2006, characterizes our "world" as one "in which injustice seems the norm in both international and intra-national affairs" and in which "the absence of armed conflict is only armistice, not genuine peace." In our world, Kultgen stresses that "the need to lay foundations for a stable and permanent peace is urgent." His suggestion? "We must use every acceptable means at our disposal to do so" (p. xvii). "Acceptable" here is doing a lot of work, but we may infer that ideally, "acceptable means" are non-violent, non-coercive, peaceful, and cooperative, in fact quite the opposite of recreational war for the selfish amusement of mad envatters. And, simulations seem to qualify as acceptable means, even simulated nuclear war. Actual war implies self-extinction, global annihilation, or worse, generations suffering by our own hands. Not yet self-extinguished, imagine instead that we develop realistic simulations from atomic to ecological systems, from vegetable to animal, social-political to metaphysical. Marrying these together, imagine that we develop simulations of the scale and resolution of Bostrom's, and we bend them to forecasting possible futures and to holding current situations against ideals. Now, we have a choice of ends. We can see what we are choosing from and what it takes from each of us to get us there, openly, cooperatively, freely. These are acceptable means to ideal ends potentially afforded by simulations, delivering the most for much less. But these are not ancestor simulations, and this is an important point.

At this stage of development, where emerging technologies are directed means the difference between surviving critical periods and graduating to proposition 2 of the simulation argument, or not and dying off.[4] In this situation, we cannot know what it is like to be post-human, but we can see that simulation technologies of the sort that Bostrom conjectures may arise from level 1 people like us working to mount the challenges inherent in being situated at level 1. They are made for a purpose, and this purpose is decidedly forward-looking. Recalling Forbes' privileged epistemic position, this purpose is for level 1 people to approach a level 2 perspective, to get as close as possible in order to best inform themselves exactly how to overcome self-extinction-level threats to existence, ultimately affording a post-human condition.

If we accept Bostrom's limit, then our own post-human ascension is impossible because our advance naturally converges on a single point (self-annihilation either directly or indirectly), or a simple cycle (war, peace, war, etc.). Moreover, if we allow that we exist as a simulation so limited, then we may as well stop trying to achieve a privileged epistemic position in order to sort out current events and plot courses to a post-human condition. Indeed, such a choice may be considered rational contra Bostrom's advice that we keep calm and carry on, regardless. This is a very different future, when one cares not for the way the world turns out. But there is another option.

Neither of the ends afforded by Bostrom's projected limit to post-human simulations is optimal. In light of their suboptimality, the know-how that brings them about and the know-that which follows cannot be counted knowledge, at least not useful knowledge insofar as "genuine peace" is the object. Should Kultgen's call to action be answered through simulations, they will arise in the search for a third way forward. As we shall find in the next section, current work on psychologically realistic simulations indicates that Bostrom-scale simulations most likely originate from a pre-post-human civilization like ours aiming for a better world as we do.

[4] Unable to visit the cores beneath the Fukushima complex, for example, simulations of the immediate environment may provide grounds for hypothesizing technological solutions, and larger simulations may be required in order to facilitate the necessary social coordination and infrastructural as well as scientific development required—perhaps over the span of many generations—to effect any possible solution to the problem of persistent leaching of plutonium into the Pacific ocean.

4 Realistic simulations

> He who thus considers things in their first growth and origin, whether a state or anything else, will obtain the clearest view of them.
> -Aristotle[5]

Should Kultgen's call to action be answered with simulations, this is a massive undertaking, requiring its own call to revolutionary means. Fortunately, this revolution is well under way. Five years prior to Kultgen's call for "any acceptable means" to "genuine peace" and from the same university, Ron Sun crafted a similar call for means in the cognitive sciences to understand, and to computationally model, the essentially social nature of human cognition:

> Cognitive science is in need of new theoretical frameworks and new technical tools, especially for analyzing socio-cultural aspects of cognition and cognitive processes involved in multi-agent interaction (Sun 2001a, p. 6).

And, directed to the development of such frameworks and tools, the "cognitive social sciences" ultimately aim for the eventual construction of "psychologically realistic" models of equally realistic social-political systems (Sun 2006, 2012), i.e., simulations that may be of use in overcoming problems central to Kultgen's assessment. This is not a simple task, bringing with it obstacles not essential to the construction of adequate simulations, themselves, requiring as it does the integration of:

> …at the highest level, sociological/anthropological models of collective human behavior; behavioral models of individual performance; cognitive models involving detailed (hypothesized) mechanisms, representations, and processes; as well as biological/physiological models of neural circuits, brain regions, and other detailed biological processes (Sun et al. 2005, p. 614).

The constructive integration that the cognitive social sciences represent is an obvious stepping-stone to Bostrom-scale simulations. It is an especially broad inquiry recalling E.O. Wilson's seminal call for the *Consilience* of the sciences in the solution of pressing problems facing humanity as a whole (Wilson 1998), the sorts of problems troubling Kultgen, as well. Wilson patiently established that we should aim for a "unity of knowledge" in response to increasingly complex problems, problems arriving on all fronts at once, indissoluble if approached in a non-coordinated, compartmentalized way. For example, ecological problems are simultaneously biological, economic, social, cultural, political, moral, and any lasting solution requires that these horizons be met at once and in mutually coherent terms. There is no sense in trying to solve an ecological problem through solely economic means, for example, by pulling a financial lever only to create friction in other spheres, setting up problems that other scientists will attempt to solve by applying pressures specific to their influence and so the tinkering continues until the meaning is lost. This is a cascade of error, a runaway conflagration, and perhaps closer to contemporary affairs than many would be inclined to confess. New methods are necessary. Acceptable methods.

One way forward is through psychologically realistic simulations. Optimally, simulations need not capture all of the complexity of the simulated, only those dimensions necessary to the inquiry at hand. And this efficiency helps to penetrate long-standing philosophical disputes. For example, Sun (2001b) considers competing interpretations of Adam Smith's "invisible hand" through the use of his CLARION model of cognition. Are these masses spontaneously ordering themselves in great organs of production, in the generation of wealth, through solely self-seeking social agency, or are there other forces at work? Sun addresses this issue by simulating how apparently selfish habits result in equally apparent prosocial consequences, effectively confirming the common appropriation of Smith's calculus in a psychologically plausible computational medium.

It is easy to see how increasingly complex simulations of this sort may help to balance often competing interests in complex problems such as those worrying Kultgen and Wilson. Moreover, such simulations may shed light on just what any individual constituent should do in order to bring ideal ends about. Sun's CLARION model in particular reveals the potential contribution of unique individual agency in the solution to group-level problems. Being a bottom-up hybrid model of individual agent cognition, with implicit and explicit (symbolic) modes of computation corresponding to bottom and top levels, respectively, unique agent positions contribute equally unique symbolic representations of experience. As an individual generates "a particular set of concepts" to account for its interactions, it "puts its own stamp on things and develops its own idiosyncrasy in its dealing with the world" proving that "there are many alternative ways of describing the external world" (Sun 2001a, p. 208). When this unique experience is symbolized, then the individual can inform other agents of the results of its operations from its own unique perspective, as well as be informed by others' unique experiences in the same way. In CLARION, "concepts (as represented in the top level) can be acquired from external

[5] Aristotle and Jowett's 2000 translation from *Politics*, book 1, section 2.

sources" as well as "internally through extraction from the bottom level (explication of implicit knowledge)" demonstrating "a self-generated component in cognition" directly contributing "to the formation and continuous revision of a rich, diverse, and useful set of concepts and beliefs that are shared by a society" (Sun 2001a, p. 18). Thus, on Sun's model, individual idiosyncrasies enrich the conceptual resources available to other agents sharing the space of action, with these resources then useful in the solution of common problems.

Unique agent experience may inform other agent action in different ways. Sun's CLARION model deals with different types of information to reflect this fact.[6] These capacities are expressed through specific sets of subprocesses within the total cognitive architecture. The sort of information that is represented in the current paper, for example, is the result of philosophical reflection. This sort of information processing exists in CLARION as a non-action-centered subroutine, representing "existentially and ecologically significant aspects of the world … that have significant bearings on an agent in its interaction with the world and ultimately in its survival" and which are "not necessarily 'objective' classifications of the world, but the result of the interaction of an agent with its world and the agent's project" (Sun 2013, p. 903). Moral/ethical issues result in a great deal of non-action-centered discourse, for example. There are facts to consider when finding something right or wrong, worth doing or otherwise. These facts may not be common to everyone's experience, but once externalized may influence others' actions.

Consider in this light the issue of the Second Amendment right to bear arms in the USA, an issue with which we shall deal more directly in the seventh section. It has been argued that people no longer require firearms, or weapons of any other sort, because the threats of nature are largely abolished. There are no more, or at least very few, bears and wolves and tigers in Chicago, so people require as few weapons as a result. Besides that, there is the Chicago police department, and it will certainly help to subdue any threat as soon as officers are made aware of the threat, and are free to arrive. Moreover, when the police do arrive, if they see you with a gun then you are more likely to be shot. So, this old habit of associating weapon proficiency and possession with safety must stop, and guns should be forcibly forbidden for the protection of the disarmed public from the armed police at the very least. This is a poignant illustration of non-action-centered information externalized and potentially informing behavior. As such, this sort of information is clearly moral/ethical, and this is to raise two issues, one being the political nature of externalized information—i.e., political voice—and the other being the issue of human motivation, particularly moral motivation. Why are people willing to freely die and kill for rules, laws, principles that they take to be right and wrong? What constitutes a State worth fighting for, and who is entitled to set up such a thing in the first place? More specifically, how are we supposed to computationally model it?

Sun has taken on the issue of motivation, asking why agents do what they do when they do it. His CLARION model represents a two-level theory of motivation, implicit and explicit according to its bottom-up hybrid nature. His account proceeds along the distinction of constitutionally original drives and explicitly refined goals. A "drive" is defined as "the desire to act in accordance with some perceived deficits or needs, which may or may not be physiological," and his model includes eleven distinctly social "high-level primary drives" with complimentary goals formulated according to drives from within and in terms of specific situations (Sun 2009, with discussion of these goals appearing on p. 95). These goals may be shared among agents sharing said situations, may be coordinated for or against, and in this simple exercise multi-agent coalitions may be formed and maintained. In this way, in pointing to goals and informing others of how to get there and why, unique agent-level experience can be seen to affect broader social orders if not found them outright, e.g., Thomas Jefferson and the *Declaration of Independence*.

The difficulty of questions like these about political motivation reminds us of how much work must be done to make Bostrom-scale simulations a reality. As important as it is, resolving this issue of individual creative contribution to multi-agent coalitions in solution to common problems still leaves us a long way from realistic simulations of the sort required in Bostrom's argument. There are many hurdles to overcome, and many of these are due not to technical issues associated with the construction of simulations directly, but rather are due to the context in terms of which this work is carried out.

For example, social systems have their own sciences, their own special languages, with oceans of literature and flowing threads of active research discrete from those of the cognitive sciences. Tying all of these threads together is, again, a massive endeavor, and it had been Sun's hope since at least the turn of the century that social scientists had been equally busy working from their side of the conceptual fence forward. However, since an initial survey in 2001, progress has been slow. The problem appears to be that social scientists have been in the habit of trading explanations of social phenomena framed in distinctly non-cognitive terms, thereby denying any easy translation from one set of models—the social—to the other—the

[6] Information that directly informs action, and that which does not. Consider in this context the traditional analytic distinction between prescriptive and moral reasoning, for example as entertained in Allen (1982).

cognitive—without losing touch with plausible agent-level psychology (cf. Sun's introduction to Sun 2012). In this vein, Don Ross has expressed disappointment in social scientists—especially economists—for having not adopted "the program urged by Sun (2006) for combining cognitive with social modeling" (Sun 2012, p. 297) while at once developing less psychologically realistic models (with possible exceptions for instance in Richetin et al. 2010). And in searching for tools up to the task within the social sciences as a whole, Paul Thagard delivers the following assay:

> … much work in current social science is dominated by two inadequate methodological approaches: the methodological individualism … of rational choice theory; and the postmodernism … in the form of vague discussions of discourse and power relations (Sun 2012, p. 56).

Summarily, integration of efforts toward the solution of global problems through the medium of psychologically realistic social simulations is hindered by conventions specific to often academically insular scholars. And perhaps this is to be expected. After all, in science, it is typical that efforts articulate mechanisms local to areas of study, e.g., vision and pattern recognition, planning and agency. When it comes to the "best" simulations that science can produce, however, we must consider an ideal integration of currently disparate fields. And, while some obstacles are especially stubborn, it is readily apparent that, even in order to realize a rough first draft of such simulations as those proposed by Bostrom, more proactive integration of existing disciplines is required.

5 Virtual cognition

> There are reasons to believe that the goal of understanding the human mind/brain strictly from observations of human behavior is ultimately untenable, except for small and limited task domains.
> -Sun et al. (2005, p. 614)

One of the upshots of Sun's approach is that it lends itself to the simulation of essentially social agents informing one another through a symbolic medium rather than through physical force and violence. In this way, it demonstrates different modes of information processing involved in cognition and action at both individual and social levels of organization, representing a source of privileged insight into the nature of the human condition unavailable to other methods of inquiry. Though it is true that contemporary imaging technologies such as fMRI aid in providing direct correlations between self-reports, behavior, and glial cell metabolism, for example, these remain limited in resolution, in timescale, and are confined to laboratory conditions. Computational models can be used to test hypotheses about cognition and behavior in contexts and resolutions that otherwise resist direct demonstration and without associated risks.

Accordingly, one explicit goal of simulated cognition is the representation of those modes of information processing characteristic of different aspects of the human condition (cf. Gok and Sayan 2012, for a philosophical assessment specifically of Sun's model in this way) including the nature of consciousness and moral sentiment (cf. White 2014). And this is one upshot of model-based reasoning in the main, that it serves the relative evaluation of hypotheses without having to risk the real deal. In this way, computational models of cognition facilitate an especially fine-grained medium for "manipulative abduction" (cf. Magnani 2009) in the effort to articulate target processes. Once the models are set out and refined against related research, they may be tested more directly with more traditional imaging studies for example.

Here, we may respond to skepticism of any project intent on simulating human cognition due to the computationalism apparently inherent in the effort. For example, one may object that simulated intelligence will fail to capture the character of our own experience because much of this character is not computable (in a digital computer). How can unitary propositions represent the fluid nature of consciousness as it is experienced? How can logical expressions be a source of value, or ground anything like a feeling of what it is like to think a logical expression? Is it accurate to consider these to be psychologically realistic simulations, when the ostensible mechanics of computation so obviously differ from the mechanisms of mind as humanly embodied? Finally, what of consciousness, self-awareness, and the unique "mineness" that characterize the human condition?

It is important to emphasize the difference between the thesis that the brain is a digital computer, or can be adequately represented as a digital computer, and broader projects employing computational resources to model processes essential to human cognition, including symbolic expression and even consciousness. Sun's model employs symbols for a number of reasons, one being his focus on individual agent contributions to social systems through this medium. His model is thus open to easy objections of the sort listed above; however, even here they are misplaced. Sure, Sun's model has not distilled all dimensions essential to cognition, only some and perhaps not in the most realistic of ways. I wouldn't want to "be" a CLARION model, frankly, but Sun's is not the only model of cognition under development. It is not supposed to do everything, by itself, right now.

Add to Sun's efforts different approaches to simulated cognition. For example, Murata et al. (2014) demonstrate a dynamic neurorobotics system that is able to spontaneously shift between proactive and reactive stances in minimization of error, with error understood as the less than optimum fit between planned action and perceived outcome. Sun's models also aim to minimize error, adjusting a bottom-level connectionist network in optimizing fit with environment, with an added capacity to extract and to refine symbolic expressions which then serve as guides for self and similar agents. Murata et al. (2014) employs a different architecture, one without a symbolic level, in order to investigate the dynamics involved in transitions from established action routines to actions undertaken in reaction to unpredicted changes in the situation. In the case of this experiment, a robot is wholly dedicated to tracking a moving ball. Its sole purpose is to switch from performance-optimized prediction to reactive tracking of the ball when the ball begins to move unpredictably. Proactively, the agent determines top-down its intended global system state, i.e., best fit with predicted ball location, and achieves optimal results as confirmed by perceived reality, i.e., achieves a best fit as predicted and locates the ball. In the reactive state, unable to predict the location of the ball, the robot responds not as quickly as when patterns are fully predicted, unable as it is to fall into an established attractor and the pattern of activity that this represents. Once such a pattern is established; however, the proactive optimal is again pursued. As simple as this may sound, it demonstrates what only appears incomputable otherwise. After all, routines established and enacted in response to an unpredictably moving ball are not programmable, not directly. Otherwise, they would be predictable. Instead, these dynamics must emerge from the proper functioning of the system as set out. This is again of the upshot of manipulative abduction in the medium of computational models in the main.

We may gain insight into the mechanisms of our own human minds through manipulation of the models that we make of them. Jun Tani has taken this approach in order to afford insight into the dynamics underlying consciousness. A contributor to the preceding research with Murata and others, Tani's multiple-timescales recurrent neural network (MTRNN) model employs time-constants which regulate the lengths of time over which the "neurons" within each subnetwork—high, intermediate, or low—integrate information (see Alnajjar et al. 2013, pages 3 and 4 for a clear summary of the method in the context of cognitive branching and switching). Tani speaks about his robots acting "unconsciously" when the immediate flow of action is uninterrupted and continuous, mediated by established low- and intermediate-level routines without change required in the top level, long time-constant system. Unconscious, the agent inhabits an environment in the strong sense of "inhabit," having embodied the habits suiting its stable situation. Similarly, we seem to sleepwalk through much of life, e.g., in making a cup of coffee (Tani's favorite example), or in locking the door without remembering.

Again, these models aim to minimize error. In minimizing error, they effectively aim for routine, "unconsciousness." And, it is in the failure of this error minimization that Tani has isolated what he takes to be the dynamics of consciousness. It is when the agent finds itself in error, i.e., *not* situated as predicted, that the critical dynamics underlying consciousness and self-awareness arise from the reconciliation of conflicting vectors representing the robotic agent's long-term aims, corresponding action routines, and its immediately perceived reality (see also Tani et al. 2004, on the social dimensions of this process via the mirror system). As in our discussion on Murata, the robot's top-level organization changes, thereby informing future predictions, and the system corrects for current operations by adjustments in top-down intended ends and the patterns of action that secure them. On Tani's understanding, consciousness arises in the correction, in that period characterized by "criticality," "when the intention in the higher order cognitive brain is effortfully modified through the process of searching for the error minimum" until a new habit is formed (Tani 2014, p. 603, see also Tani 1998).[7]

Abnormal consciousness is also accessible on this model. As in human subjects, the intentions of model agents can be confused. In another study using a similar model, Yamashita and Tani (2012) added noise to the information flow bridging the higher (intentional/goal state representing) and lower (responsive to the immediate environment) levels of the network, with results ranging from "spontaneous intermittent increases of prediction error" (bad predictions) and "irregular switching of the intention state of the network" (unpredictable behaviors) in cases of minor interference to "disorganized" behavior that "no longer followed logical rules … characteristic of more severe cases of schizophrenia, such as cataleptic (stopping or freezing in one posture) and stereotypic (repeating the same action many times) behavior" in cases of greater interference.

These experiments demonstrate abnormal conditions emerging as the dynamics of an abnormal network converge on a stable pattern of routine action in the normal way, in the attempted minimization of error. As it is in both human and model being, transitions from one

[7] Consciousness on this model is to adapt a projected future to the pressing reality, realizing how one has gone wrong while establishing how things might be right, again (cf. White 2006, 2010, 2014).

stable condition, one situation to the next are effected continuously with small adjustments over time as the motor system must transition fluidly from one joint angle to another until a new optimal is established, while target states exist as distinct attractors for the system as a whole. These are goal situations representing an optimal fit of a trained agent with its (anticipated) environment. They represent the purpose of that system, the object of its intentions, and consciousness, normal as well as abnormal, arises in the gap between the intended and the currently inhabited situations as they contraindicate one another. This is to say that consciousness is essentially what it feels like to be in error, wrong, somehow ill fit with the environment, in a normal way or otherwise being another question.

As a result, a platform for the testing of hypotheses about the nature of human cognition is emerging, with consciousness naming that "dense interaction" between bottom-up responsive and top-down proactive pathways within an embodied agent for which its own current and future situations are always the matter at hand. What is left in refining our understanding of consciousness is to model increasingly realistic systems and to match these against human self-reports. Insofar as we embody a similar logic, we can make sense of such systems, and this again is part of the appeal of contemporary neurorobotic research. Tani describes his robots as having intended to roll a ball, or pick up the ball. These are models of aspects of our own embodied condition. These are not mere representations of this condition, but they *are* this condition only in less complex ways. They are *virtually* real. Thusly, computational models of cognition, including neurorobotic implementations, afford a privileged perspective on the truth of what is otherwise unobservable, inexpressible, and even apparently incomputable about the human condition at least because they allow us to isolate certain relationships without distracting complexities peripheral to the inquiry at hand.

Regardless, one may object that Tani's, Yamashita's, Anajjar's, and Murata's models lack the distribution of cognition through a symbolic medium, being thus in some ways less realistic than Sun's. Again, however, such an objection neglects the fact that these are not the only models under development. They aren't supposed to do everything by themselves, right now. As different approaches merge and mutually inform one another, future research into the bottom-up basis for symbolic expression (literally, a pushing out of something internal) and natural language may inform the realistic simulation of enacted speech, in response to unpredictable (social) forces, and toward certain (common) ends, i.e., political speech. When married to the cognitive social sciences and stitched into social and political simulations, the results should be large-scale simulations populated with agents surprised in the face of a more or less uncertain future and managing this uncertainty, collectively, coordinating action and establishing standards of action using symbols to identify stable constructs and their relationships with one another in the forms of laws, psychological schemata and systems of ethics. Greater psychological realism, that is the result, and perhaps a simulated civilization bent on becoming and remaining "post-human" at least in part through the use of realistic simulations.

Of course, one may maintain the objection that all such models remain fragmentary, capturing only some aspects perhaps in principle essential to cognition yet being so limited as to remain open at least to objections of irrealism. Yet, this is only to object to the nature of model-based research in general, that it develops against a limit of realism. One may further point out that the materials that encode for information, even as homologous dynamic systems, do not yet change as does the substrate of humanly embodied cognition. As advanced as contemporary work may seem, it remains software models (however unique) running most often on generic hardware via not very flexible "firmware" (controllers of controllers and so on) all together interacting with simple, most often static environments. Even the most advanced robot neurology is not integrated from the molecular level upward through stages of self-transformative interaction with a sometimes chaotic and even malignant natural material environment. This is all true, of course, yet these are also problems for materials science and information technology, and alone do not stand against the potential for a realistic simulation of the human condition with these processes taken into account and articulated in a different medium in the meantime.

One may also object that a "dense interaction" is one thing, but something like phenomenal consciousness, or moral sense, these are different things essential to the human condition and, regardless of apparent evidence to the contrary, beyond computational representation. But, again, such an objection may rest on a misestimation. In his 25 years modeling cognition, Tani has adopted the position that the "hard problem" of consciousness is not really so "hard" (Tani 2015). And, I agree with him [cf. White 2006, 2010, 2012, 2013, which argue in different contexts that consciousness is the felt difference between embodied situations, and White (2014), in which this dynamic is described as the inchworm model (IM) of cognition as an extension of the essential structure of agency]. As it is now, none of our models immediately tell us everything about consciousness all by themselves, but they *do* do a good job of dissolving old problems while opening the way to better ones. And this is the point—the evidence suggests that, regardless of argument to the contrary, the promise of

multiple approaches integrated in articulating cognition in a computational medium is being realized.[8]

The practical potential for the integration of diverse lines of research into singular models of those aspects of the human condition required to render the large-scale psychologically realistic social-political models that may serve in guiding a pre-post-human civilization such as ours past self-extinction-level threats to a more secure, post-human condition seems clear. With further advances integrated around such a purpose, the potential for simulations perhaps indistinguishable from the human condition at least in some focal contexts remains open and, moreover, if a post-human condition may only be achieved through the use of psychologically realistic simulations by pre-post-humans—by people like us—then simulations of the sort that may include something like us may indeed be inevitable. They may *need* to happen, or a civilization goes extinct. So given, then, we must ask about our own epistemic position. What are the relative likelihoods that we are an unsimulated pre-post-human civilization, or that we are a simulated pre-post-human civilization? Considering the limitless potential to simulate civilizations such as ours afforded by projected post-human technology, if we are unsimulated then we may infer that post-human civilizations likely do not exist. If we allow for the possibility of a post-human civilization, including ours, then we most likely exist in a simulation. The trouble, however, is that we have no evidence for a post-human civilization beyond the hypothesis that we currently exist in a simulation created by one. So, the next three sections of this paper pursue more direct evidence to the conclusion that we most likely exist in a simulation.

6 Simulations and political economy

> Why do powerful people not want peace? Because they live from war, from the arms industry. … An old priest I had known years ago said, 'the devil enters through the wallet'. For greed. This is why people do not want peace!
> -Pope Francis with children May 11, 2015[9]

> Come, shave your heads and I will give each of you a red hat and plenty of vodka.
> -The Devil[10]

Problems require solutions adequate to the complexity and urgency of the problem at hand. If we take Bostrom's projections of self-extinction seriously, then the most complex and urgent problems facing us today are self-extinction-level problems dealing with the mismanagement of resources and deadly technologies. With self-extinction, all is lost. In averting self-extinction, all is potentially gained. As solutions to problems of this scale generate the greatest possible good, then these are the problems most worth solving. It is no coincidence, then, that problems of this scale are also the focus of Aristotle's *Politics* (cf. Aristotle and Jowett 2000).

Great problems require tools up to the task. For Aristotle, ensuring the flourishing State is a higher art than that for the flourishing individual, and the science that is bent to the task is that of the statesman, economics. Economics is the practice of "household management" and "thrift" popularly coming to signify the "judicious use of resources"—especially public resources through mechanisms of State—in the seventeenth century as the science of household and State management became an issue for a burgeoning, literate, politically activating middle class.

The *Politics* explores economics in the organization of State and household, with the best measure of leadership being the provision of opportunities to flourish. So, on Aristotle's analysis, the first currency of a sound economy is food, and the ultimate measure of wealth is not in dollars or even in apples but in opportunities. Statesmanship properly understood is opportunity space optimization, and economics is the study of the manipulation of this space, i.e., the ability of the statesman and householder to secure the space of opportunity in terms of which all constituents, self included, succeed or perish.

Aristotle tells us that the best State constitution establishes an economy that balances the character of the constituency with the natural and political environment in terms of which these members must flourish (cf. book 4). He emphasizes the health of the State or household as a plurality—a polity—rather than the wealth for the few at the expense of others. He emphasizes the need for a robust middle class within which all strive to private excellence in common, and so do not contrive to take from one another or through usury to parasite on the productive membership. This portrait is radically different from the financial economy facilitating current global commerce, that type of economy farthest from nature on Aristotle's assay, and led by the worst kinds of people—petty manipulators operating some artificial monopoly at the expense of others (cf. book 1, section 9).[11]

[8] The trick is to not look at your PC and think that you see a brain.

[9] Personal translation.

[10] As translated in Tolstoy (1891, pp. 63–64).

[11] Consider the parallels with Keynes: "Capitalism is the astounding belief that the most wickedest of men will do the most wickedest of things for the greatest good of everyone." Though difficult to pin down, this quote sticks to Keynes because it expresses a sentiment which increased as he saw terms purposefully established in the deal done at the end of WWI directly—and predictably—creating conditions inviting WWII (cf. Backhouse and Bateman 2009).

Aristotle's ideal is not a financial economy such as that in terms of which we exist today. Ideally, a State operates as an economy of virtue in which people rise to recognition through demonstrated practical wisdom, in the creation of opportunities for the good life in common, with burdens of leadership for the administration of what are essentially public resources arising as others willfully follow the more virtuous in this industry (a vision shared by Adam Smith, cf. Thomas Wells 2013). So optimized, in such a system each constituent may most directly contribute to the solution of the most urgent problems facing the society by discovering or creating opportunities to get past them. The problem for the statesman then becomes merely how to enable the voluntary coordination of all of this human potential, ensuring that such opportunities can be secured.

In an ideal economic environment, the role of the statesman is to order the State in such a way as to optimize the constituents' opportunities to create wealth in the form of further opportunities to flourish for the State and all who compose her, including the statesman responsible for setting her up. How far are we from this potential, today? In indirect answer to this question, we need merely evaluate contemporary leadership in light of the following caveat, that it is impossible to imagine a State well-ordered and just by way of the worst leadership, and ill-ordered and unjust by the best (cf. book 4). We might then hope that a civilization may survive self-extinction by chance alone rather than by way of good leadership and enlightened social engineering, that we may somehow "get lucky" as if the suicidal man with an incapacity to form short-term memories everyday slips on the same wet spot on the balcony before acting on his long-term passion to jump, ending up on the sofa rather than the sidewalk from now until the end of time. Of course, this is ridiculous. Logically possible, but following our Aristotelian principles of good governance, this is a false hope and safely ignored.

In the same way, we may ignore politicians borrowing from the future to finance mutually assured destruction before that future ever comes about. But, we cannot safely ignore them. In every case, the role of the statesman is to order the State—economically—in such a way as to facilitate the constituents' creation of wealth. For whom, and at what costs, these are additional questions. Aristotle's statesman, intent on a healthy economy, is concerned with the delivery of the highest goods relative resources necessarily expended by affected parties, him/herself included.[12] It is not the goal of the statesman, insofar as this goal is a healthy State or household, to make the rich richer and the poor poorer. It is rather the goal that all become of their own powers better so that the State as a whole flourishes best.[13]

This seems simple enough, but at the same time current global problems—perhaps due to their top-down nature—remain insoluble. Indeed, without divine intervention, it is difficult to imagine 9 billion inhabitants voluntarily coordinating social transformations necessary to ensure a post-humanity. One way to openly try, however, is bottom-up. Simulate them first. Illustrate how each constituent may benefit from and contribute to possible answers. Guided by open, interactive simulations through which collective futures are first projected and pathways agreed upon, social contracts may be proactively pursued and a lasting peace achieved through acceptable means all by way of enlightened statesmanship, instead.

The statesman who sets out the richest space of opportunity for constituents so that they can in turn create further opportunities for the good life going forward is the best kind of leader. Large-scale psychologically realistic simulations afford a unique medium for the demonstration of this capacity. Serving as an ideally efficient and effective (e.g., time and energy saving), complexity-reduced certain means to set out, test, and secure this landscape of opportunities, simulations provide a pre-post-human people a guiding portrait of political economies friendly to its own post-human potential, with constituents individually and uniquely informing their own collective evolution, proactively overcoming self-extinction-level threats ultimately due to naturally evolved physiologically grounded behavioral routines, habits, e.g., racism, speciesism, tendencies to impatience and violence. Thus, for a level 1 civilization like ours, Bostrom-scale simulations likely begin as tools of statesmanship toward a more or less ideal economy to aid in the transition from unsustainable to sustainable, from unnatural to natural, from IOU to opportunity, from M. A. D. to peaceful, from pre-post-human to post-human.

Following Aristotle, we may constrain our attentions to those political economies that are optimized to maximal self-determination independent of material contingencies, becoming so by way of the best leadership and the efficient management of resources, human resources most of all. These are potentially post-human political economies. There is no living on borrowed time to enrich bankers through debt under an unending threat of nuclear annihilation on this formula. There is only the crushing reality, that these biggest problems must be solved, and only in

[12] From such a common term, different systems of government might be relatively assessed, and this is exactly what Aristotle does in the *Politics*. Aristotle was after all a social scientist, and enjoyed an extensive network of researchers, receiving reports from more than one hundred and fifty independent States.

[13] And insofar as one is intent on anything else, he or she is not a statesman but something else and likely a deceiver.

their solution—not their exacerbation—can economy of any lasting sort remain sound regardless.[14]

Finally, in no way is this objective served by hiding this purpose from the simulated entities within such a construct. Once again, deceivers from tyrant to petty manipulator are the worst of people on Aristotle's assay, and following his principles of good governance we may safely discount any possibility of deceiving simulators achieving a post-human future. Deception represents misspent resources, bad *oikonimia*, and so bad statesmanship, making survival to post-human status all the more unlikely. Moreover, we have direct evidence against a deceiving simulator in current efforts at developing psychologically realistic simulations. Deception is not the motivation for people laying the groundwork for sophisticated future simulations, for the simple reason that in being false they are not realistic, and in their unrealism they are not useful.

Wise men do more important things than replay fictions. They farm possibilities. Thus, if we do exist in a simulation, any worries over "a pervasive *de dicto* scepticism about all aspects of physical reality, including those aspects epistemically relevant to the limits of post-human computation" appear ill-founded. If we exist in a simulation, we do not exist in an ancestor simulation but rather in a possible future. We do not live a lie. We represent hope. So, rejoining the optimistic trajectory afforded by current neurorobotics research, we may benefit by explicitly committing to a course radically different from Birch's BIVist response to Bostrom's simulation argument. Rather than attacking Bostrom's argument, let's allow for it from the beginning and embrace it without reservation aside from its retrospective attitude. Allow that we are in a simulation, a psychologically realistic computational model of social-political systems constructed by a sufficiently advanced post-human civilization to solve its most complex and urgent problems most efficiently. This is a simulated future representing an opportunity to flourish, and so invested we need neither fall to moral nihilism nor remain unaffected and continue to plan as normal per Bostrom's advice to continue "going about our business and making plans and predictions for tomorrow" (Bostrom 2003, p. 255). In fact, we should discover redoubled commitment to a certain scientific way of life, instead, and this is the point of the next section.

7 The virtual future

> I think that until major institutions of society are under the popular control of participants and communities, it's pointless to talk about democracy. … Moreover, I think that's entirely realistic.
> -Chomsky (1973)

> We may still have time to reconsider our courses and to see the world with new eyes.
> -Keynes (1920, p. 296)

We have been reasoning from the premise that the capacity to lead at any scale derives in part from a predictive capacity, an understanding of what is coming and what is necessary, constraining reason to realizable goals in order to ensure the continued flourishing of all concerned. For Aristotle, this understanding is represented through the distinctly human power of political voice. The administration of justice is the natural purpose of this voice, exercised in rational discourse. Everything else is unnatural or at best a diversion and misuse of that faculty essential to the political animal. Ideally thus—when the best of leaders leads the best of these political animals—the State is led solely through the rational discourse of its constituents, informed by the practical wisdom of leaders as we/they lead.[15] And as this discourse has grown so large and complex, in communicating, comparing, and projecting political ends, simulations stand to be uniquely effective in moving this discourse forward.

Let's look once more at the two possible motivations behind our simulated condition. The first motivation is deceptive or worse spurious, an "ancestor simulation." We are a simulation built for amusement, easily let go to laziness like a forgotten, polluted aquarium. If this motivation is most likely, then the end of that simulation is arbitrary. Such a situation indicates that our simulation is the product of a *post*-post-human phase, in which the purpose behind the use and development of simulations has been forgotten, and the society is left without a self-protective appreciation for the powers of prediction that such simulations provide. In that case, easily disconnected along with the heat, turned off like an absent-minded hobbyist on his way to his mother's for a long holiday, life is as senseless as the simulation that supports it. And given this

[14] As opposed to the litany of licentiousness distracting from the good according to the designs of Aristotle's demagogues, bear in mind.

[15] This is to point to a feature central in Plato's dialogues, the relationship between wisdom and rhetoric. "Practical wisdom" is the virtue proper to leadership, required by statesman, king, horse trainer, and scientist alike. All lead by maximizing constituent capacities to self-leadership, i.e., practical wisdom which by the current example includes the capacity to follow in those ways that ensure the flourishing community. In simplest terms, the ideal leader must optimize opportunities for all involved to lead according to expertise, and this means empowering subjects to create and to pursue their own opportunities. This means, effectively, that the leader must follow. The trainer must be informed by the horse, deliver to and anticipate its needs, and vice versa. The statesman must be informed by those with whom he shares residence and vice versa, and this is also ultimately why Aristotle holds friendship to be the bedrock of a healthy State.

information—contrary to Bostrom's advice—one may be inclined to discount the future more dramatically in considerations of lifestyle, if not purposefully emulate the laissez-faire entitlement characteristic of our negligent designer in (perhaps a vengeful) moral nihilism.

The second possible motivation points to our being in a simulation built for a purpose, to solve a problem, a situation with interesting implications. Ideally, a simulation optimized for the solution of a problem requires only that complexity necessary to solve that problem, nothing more and nothing less. That we—as simulations—feel interested in understanding the global purpose of things accords with the view that our simulation is one in which such problems are to be solved, and indeed in which the phenomenology of problem solving bears special weight given its special representation—the consciousness of discovery as target condition reinforced by the same natural addiction processes as those responsible for the life-consumptive character of heroin and cocaine. If we, as in the models of Sun and Tani, are built to minimize error between perceptual reality and projected ends, i.e., essentially addicted to problem solving, this begs the question: Why model interest in solutions to such problems as cosmology and ideal political economy? Why these ends? Why throw such computational power at these considerations—the lives and sufferings of a great many geniuses over the course of human evolution—if their solution were not essential to the object of the simulation in the first place?

The short answer is that one would not. "Post-human" does not imply wasting time, writing inelegant code representing unnecessary complexities and ultimately running senseless simulations. A simulation is best when the constituents of said simulation suit the aims of the simulators, and thus the constitution of constituents of any efficient simulation should serve as a source of information about the purpose of that simulation as well as the sort of intelligence behind it. It stands to reason, thus, that virtual residents of a simulation operating on such a principle should be able to look for the parts of the simulation that are at once most demanding and most compelling to discover clues to the focal purpose of their situation as a whole.

People don't last long when they make a habit of misspending critical resources, especially those which may forewarn them of otherwise surprising threats to existence. This truism points to our likely purpose in the whole of things. Problem solving. Solving problems is the most demanding aspect of our simulation. If we are in a simulation of problem solving, then we may look to the most difficult problems to solve for an estimate of the limits and aims of that simulation. Threats to global civilization are both our most pressing and most difficult problems to solve. Nuclear war over who runs the Ukraine is, for instance, a very pressing problem. Fukushima's lost coriums, for another example—this is a very big problem. These problems can stop progress to level 2, if not end complex life on Earth, permanently. Solutions to these really big problems facing our world today often appear intractable, and are accompanied by the deepest of emotions, what Heidegger called "angst" (Heidegger and Stambaugh 1996). Mortal fear of possible consequences to apparently insoluble problems doesn't mean that we shouldn't try to solve them, however. Quite the reverse. That said, we must be careful to attend only to those problems that are indeed most important, and this takes courage. Otherwise distracted, we may take a stand in the wrong place however politically expedient, fail to see our way forward to satisfactory ends, fail to recognize our full purpose, and either remain stuck at level 1 or end, completely.

Consider on this point Stephen Pinker's recent foray into political philosophy. In *The better angels of our nature: Why violence has declined* (2011), Pinker takes on the controversial and apparently pressing issue of an armed citizenry, arguing for a monopoly on violence to be assigned to the mechanisms—and by extension office-holders—of State, suggesting that under similar pogroms violence has declined. As friendly to non-violence as most philosophers are, it is difficult to not simply nod one's head. Of course, who needs weapons? Who wants more violence? However, that does not save Pinker's argument, as it—not unlike Birch's argument against Bostrom—selectively neglects direct evidence to the contrary. Facts. For example, it is clear that in the contemporary USA gun violence and violent crime go down as gun ownership goes up.[16] And after all, this is common sense. There may have been a time when the most vicious man with the biggest stick took what he wanted and left all else in ruin. Guns leveled that field, and an honest small and quiet man became able to defend his home, his daughter, his farm, his own. Firearms have thus long been understood to be a great equalizer against the brute thuggery that otherwise reigns when the meek are left no effective recourse. Pinker's reasoning also runs counter to established historical fact, as demonstrated in other nations and eras. Hitler monopolized arms, stripping weaponry from citizens and at once stripping rights of other forms. Occupants of the ghettos—like Palestinians in Gaza today—dug tunnels and smuggled arms and some seeded the same gangs that terrorized the Middle East as Israel was cut out for them, later, but this history lesson is beside the point. Stalin. Mao. Similar men made similar moves disarming the public and all demonstrated similar—in latter cases actually more murderous—results. And this all simply follows from the truth of experience mentioned above. If every Jew in Treblinka had

[16] See, for example, http://www.washingtontimes.com/news/2012/jun/18/gun-ownership-up-crime-down/. Accessed October 15, 2015.

had a gun, there could have been no "Holocaust" just as if every Palestinian had a gun today there would be no more illegal "settlements" and uprooted olive trees. With this, we may add to that old saying about armed agents of State arriving too late to stop violent criminals: "When seconds count, the police are only minutes away." Minutes are not long enough. "Never again" is more like it.

And this brings clearly to the fore the real problem driving civilian disarmament. Lasting peace is threatened less by arms than by empty bellies and broken promises, thus explaining also why the dependent and hungry should be forbidden knives while agents of the corporate State deserve weaponry enough to exterminate the living Earth accidentally. Some very powerful people have gotten very rich keeping us stuck in a level 1 situation. As controversial as this might seem, it is no less true. All of the attention diverted to gun ownership distracts from this truth and in so doing from the greatest threats facing humanity today, capabilities for violence belonging to State and corporate agents, and not just any State and corporate agents. In recorded history, State and corporate agents have never held such power. Even Hitler never threatened human existence, but we live under increasing threat of this very sort today. As I send this for initial review, for instance, Obama has allocated one trillion dollars for the improvement of US nuclear weaponry over the next 30 years.[17] This is a problem of political economy ill fit with the environment (i.e., one with a flourishing future in it), getting worse, not better.

It is difficult to imagine private citizens ordered according to an Aristotelian virtue-economics amassing, maintaining, and then upgrading 20,000 nuclear missiles. Rather, they have family and friends for whom to care. It is equally difficult to imagine those same citizens simultaneously executing each other with handguns. Either possibility implies self-extinction, yet either enjoys a positive potential only under machinations of the corporate State. If we accept that our purpose is the solution of the most pressing problems, then it is difficult to follow Pinker's reasoning to place the burden of non-violence on the non-corporate, non-State agent, and to rise in unified call to civilian disarmament. To do so would be to suboptimize problem solving for distractions, execute bad code, and make matters worse. That said, we can empathize with the effort. These issues are very confusing, and the current state of education often does not afford the facile solution to problems in applied political philosophy.

[17] An "obamanation." As reported in the Daily Mail: http://www.dailymail.co.uk/news/article-2765493/Projected-US-nuclear-weapons-spending-hits-1-TRILLION-just-five-years-Obama-s-Nobel-Peace-Prize.html. Accessed October 15, 2015.

Of course, there is some attraction to giving up and giving over responsibility for the way that this world turns out to the State through its agents and officeholders. But this is not really a solution to any problem at all. It is an example of what Sartre called "bad faith," and what Heidegger would have recognized as inauthenticity due a cowardly retreat to "fallenness." Here, we may recall the deception in Birch's presumed simulator, but locate it where it belongs—in ourselves. It is self-deception. Self-distraction, putting one's self in a position only to say "Well, I tried" rather than "Wow, we succeeded!" Such a life is lived on the gamble that personal death will come before the peace ends however expensively purchased as, of course, someone else is paying. Someone in the future, after all, may have to fight and die to get those guns back. Taking our simulated condition seriously, however, it is difficult to discern the purpose in the simulation of such civil cowardice. One might as well simulate very heavy moral cans getting kicked down the social-political road, i.e., war, peace, war. And by our prior reasoning, being an inefficient use of important resources, bad economics, and bad statesmanship, such a purpose is equally unlikely for our own simulation. Should we indeed be simulated, this is most likely not what our simulation is about. And if it is, then perhaps we deserve to be unplugged, after all. We should look elsewhere for our global purpose, else suffer from a self-fulfilling prophecy by way of bad faith regardless.

8 Simulated moral reality

> For when the truth squares up to the lie of millennia, we will have upheavals, a spasm of earthquakes, a removal of mountain and valley such as have never been dreamed of.
> -Nietzsche and Large (2007, p. 88)

> *Not all meaning is constructed.*
> -Heintzelman and King (2013, p. 97)

Time may be running out on our simulation and not for Bostrom's reasons, that we are too advanced to be efficiently simulated. Rather, we are failing, and why waste time on a broken system when a reboot might reveal more promising opportunities? Of course, we should anticipate that the plug will be pulled on our simulation when self-extinction is a lock. After all, the most useful simulation for a post-human civilization stands to be that of a civilization having achieved a post-human situation, offering the gift of (limited) prescience through the revelation of possible futures to which the post-human

simulator is also privy. Why continue in a simulation that cyclically kills itself off? What kind of future is that?

Ultimately, the value of information for us pre-post-humans is also as a window on the future. We should not expect this to change. In our own social-political space, however, we find an opposite account. Here, it is as if every lesson of historical statesmanship is rewritten and redacted such that the guide to the future is foggier, rather than more clear. Consider on this count a recent effort to remove from US high school history curricula the review of the role of civil disobedience in achieving historical social-political advances in access to economic opportunities, to political voice and moral standing.[18] Such a policy is contrary to John Kennedy's insight, for example, that to make non-violent change more difficult is to make violence more likely. Forbidding tools for self-defense and avenues to social-political self-determination to all but agents of the State is a dangerous distraction from the most serious threats confronting a level 1 civilization like our own. It doesn't make enlightened social engineering easier, and moreover it is psychologically unrealistic. People will disobey perceived injustices, and defend themselves against tyranny. To set out otherwise is to be going backward, with obvious risks to our collective existence. It is also to deceive ourselves, as if we are our own mad envatters.

Given such problems as gun-confiscation in the "cop-pupied" post-9/11 USA, coupled with education leaving students simultaneously ignorant of the non-violent exercise of political voice in the correction of systemic injustice, increasingly powerful roadblocks to political influence "from the margins" (cf. Habermas 1996) may lead us to imagine that global transition and coordination problems are presently insoluble, not simply difficult to solve. And, perhaps most troubling of all is the complimentary attitude that the employment of State-level violence—war—is inevitable if social change is to be directed, at all.[19] But what is our recourse? To ask where is the modern messiah to deliver us past self-extinction and to the peaceful "promised land" not for some "chosen people" alone but for everyone, always? Barring an answer, where is this incomputable potential in us?

If we take our simulated condition seriously, then there is good reason to suspect that help from G(g)od(s) is not forthcoming. After all, a supernatural *deus ex machina* runs contrary to any reasonable purpose for realistic simulation. Why create such a simulation, only to send in a rescue boat at the last minute to save it all from blowing up? Such a story suits the characterization of God as merciful, but invites charges of malfeasance and neglect—if not criminal abuse!—along the way. So much apparently senseless suffering, only to send in the clowns, prop up the simulation, and do it again ever bigger. Unless convinced that a deified simulator exists but deserves to be prosecuted for war crimes, we have good reason not to wait for a hero to save us from ourselves. Rather, we are simply stuck, at the edge of level 1, waiting on ourselves to save ourselves from ourselves.

Given the situation—tragedies most all man-made but for the climate—hopes are dim. Most certainly the same brand of humanity that has given us Nagasaki may fail in elevating itself above such banality in the future. Fortunately for us, however, it is exactly this sort of eventuality that Bostrom-scale simulations should excel at helping any level 1 civilization to avoid. Even with present technologies, we should be able to run countless simulations to get a sense of where critical resource allocations may lead us. It will take work to fit these simulations against measurable reality via theoretical ideal and so be able to judge good, better, and best roads ahead, but if we allow that this work can be done, then one thing comes very clear.

Predictive simulations afford a unique means for people to cooperate over the generations necessary to realize post-human goals. So far as our current situation goes, considering especially the rising distrust in leadership,[20] if perpetual armistice is to be replaced with something better, then realistic simulations may prove invaluable in informing public discourse. If such tools are to be a reality, then they will rise from the research bedrock under development, today. Moreover, we may take the motivations driving current efforts as evidence for similar efforts of more advanced societies who may have made for themselves a lasting peace, already.

If we allow for the existence of a post-human condition achieved by a civilization like our own through the exercise

[18] For instance, as reported in Salon.com: http://www.salon.com/2014/09/24/how_high_school_is_teaching_civil_disobedience/. Accessed October 15, 2015.

[19] People at the highest levels seem unable to imagine otherwise. Just the other day, I was speaking with one of the most influential scholars in the world about a ban on autonomous killing machines, one that he supports, and a few minutes later he refused to recognize the political voice of human beings supporting a second amendment right to arms. This is understandable, as human beings are autonomous killing machines. I asked who should be held responsible when these people become so frustrated by the loss of political voice to resort to violence. Should we blame those desperate enough to do desperate things, or those who shut them out of the decision space, exacerbating their desperation? And moreover, who is going to take their guns away? Autonomous killing machines? Without some way forward, the end is always murder on a grand scale, oppression, apartheid. The question of who or what carries the weapon is mere distraction. This is why simulations are so important.

[20] For instance, public favor for sitting Congress-critters hit 13 % 2013, the (then) all-time low: http://www.gallup.com/poll/166196/congress-job-approval-drops-time-low-2013.aspx. Accessed October 15, 2015.

of the potential for self-direction that psychologically realistic simulations may afford, then we must remain committed to Bostrom's conclusion, that we likely exist as a simulation. Once the technology is available to a civilization bent on a post-human condition, it should be used as well as possible for the highest purposes, resulting in countless simulations more than actual worlds. So, the likelihood is a simple one, proportional to the ratio of simulated to non-simulating civilizations. With enough simulations, then this likelihood may approach unity.

Even this certainty should not cause us to "go crazy" and embrace moral nihilism, however. Instead, accepting the fact of our simulated nature, our existence becomes more meaningful rather than less. Indeed, our existence may be more meaningful than that of our host. If this is a simulation, and if this simulation is intended to reveal solutions to global coordination problems through simulations, then these solutions—our solutions—could contribute to the survival of the world on which our host resides as well as countless others, simulated and actual alike, through Bostrom's Russian-doll-like "levels of reality." Finally, the meaning*ful*ness of a life dedicated to the constructive solution of these greatest possible problems is revealed, especially as a simulation, to be all the more worth living.

What better excuse for the horror show that is Gaza, Falluja, Dresden, the Donbas? What other excuse could a post-human offer for the suffering that some cause others due their roles in the grand simulation? What use is simulated self-annihilation for anyone other than a mad envatter? Or, are our simulators as desperate for solutions as we ourselves are? We left the possibility of a mad envatter behind as either nonsensical or self-incriminating. So that leaves desperate, and this only adds to the urgency with which we must fulfill our purpose, the purpose behind all of the used-up resources, cognitive, computational, energetic, material—human. There is no excuse for Dresden but one, that it forces on us the biggest questions and the hardest problems to solve. What is the meaning of life, and how do we order our world in order to best realize it? This is not a question for psychologically realistic simulations, alone. It is question for the philosophy that shapes them. One may object, stop the inquiry, but to do so would be to invite skepticism. Moral nihilism. Dresden is meaningful, or the stories that we tell ourselves about the meaning of life represent no lesson. They represent no error. There is no felt need for correction, only countless generations of successors to deceive.

On the other hand, if our purpose is the correction of error then we may do well to remember John Kultgen's call for "any acceptable means" to a stable, lasting and sustainable peace. Consider the tools for statesmanship to be derived from an ability to realistically simulate, first, the set of political systems described by Aristotle. As a measure against actual states of affairs and their proposed modifications, such an appropriation of Aristotelian political theory could establish a standard for calibration in an industry of realistic social simulations. With such a standard, we may then simulate different leadership strategies within increasingly realistic natural and political environments. We may well discover that currently accepted classifications no longer hold, e.g., some political systems are no longer democracies, or republics, or monarchies, and discourse over them and the officers who manage them should adopt a corrected terminology. In this way, simulation technologies may do more than "ground the social sciences in the cognitive sciences" per Sun's program. *They may normalize them*.

Realistic simulations may allow for the normalization of the social sciences in standard philosophical constructions. We may simulate a thousand generations into the future, and confirm the long-standing philosophical suspicion that rule in the optimization of the constituency toward "self-sufficiency" for Aristotle, "self-sovereignty" for Kant, or "genuine authenticity" for Heidegger results in an optimally adaptive social-political conformation. We might on this evidence decide not to wait for a thousand generations, and rather encourage such a policy now. Further simulations may illustrate how to transition to optimally adaptive conformations, through intermediate states and thereby we may manage our own self-development, in the open, as a civilization. Finally, we may instantiate a similar self-sufficiency in our simulations, and immerse ourselves in this community. We may, in a virtually real moral reality, directly consult with especially virtuous yet simulated subjects about our own potentially post-human futures. Some simulated subjects may achieve lasting recognition for the solutions that they represent for host civilizations. Others may be replayed over and again during especially critical periods in order to focus on a specific approach to leadership, for example, and still others may arise spontaneously in Sisyphean reminder that behind all great acts is simple repetition, i.e., life in an attractor basin. Spun out accordingly, we well realize the critical role for simulations in effecting necessary social transitions for any pre-post-human civilization going forward. It is a moral and ethical role, because the alternative is as it has always been. Violence. War.

9 Conclusion

> If you do away with yourself then you are doing the most admirable thing there is: it almost earns you the right to live…
> -Nietzsche and Large (1998, p. 61)

> What this means is that, by and large, a human agent's premiss-conclusion reasoning is the right way to reason when his conclusion-drawing cognitive equipment is in good working order and, on that occasion working in the right way, operating on good information in the absence of hostile externalities.
> -Magnani (2015, p. 15)

The worries that had prompted this research, that we may inhabit a simulation for the entertainment of evil demons, are laid to rest. The world is just as it must be, and we must do something about it. Rather than continue to live, predict, plan, and act as normal, we should accept it as our constitutive purpose to correct error. The remaining trouble is simply that we are the only ones present to fulfill this purpose. The future—the only real future facing us unless living under constant threat of self-annihilation can be counted as a "future" at all—exists in adapting technologies in the optimization of social and political systems for problem solving, sustainability, and ultimately the good life. To this end, the future of philosophy arises at least in part in the grounding of the cognitive and the social sciences in the physical sciences, in understanding the metaphysical in terms friendly to realistic simulation, at the very least so that emerging simulations can be measured against an enlightened and guiding account of the human condition. Through this industry, stable visions of possible futures and the paths that take us there may be discerned, proactive human self-direction past extinction-level threats may be facilitated, and post-human political potential may be realized. Again, what is the alternative?

As out of reach as it may have been for the ancients to bring diverse if not divergent people cooperatively together under one umbrella of excellence without the "noble lie" of a founding dissemblance, information technologies can help to facilitate such a system in the open, today. There remains now only the setting out and ordering of the world accordingly. There is no reason that such a setting out cannot be first a model, a simulation. In fact, a simulation is exactly the sort of setting out that one might expect necessary, given the scale of change and the capacity for perhaps a single uniquely disaffected person to effect it.

In the end, the probability of a simulated existence cannot be grounded in the possibility of a post-human existence accidentally interested in an ancestor simulation for which we have no evidence beyond speculation. Rather, the probability is proportional to the clearest and most distinct direct evidence imaginable, our own felt commitment to a lasting peace through similar means, today. Given that we are not the first in the universe to find ourselves in such a situation, and as a simulating civilization should create countless more simulations than exist as actual worlds, the sum of our collective commitment to a post-human condition is proportional to the probability that we live now in a simulation setup in a similar effort by someone else perhaps a very long time ago. For myself, this is a certainty.

Acknowledgments Special thanks to Peter Broedner for the patience to advise multiple drafts of this paper. He is largely responsible for its depth of analysis. Thanks also go to the Fall 2014 Ethics class at KAIST for testing some of the arguments entertained herein, especially those regarding the statesman and leadership. This work would not be possible without the consistent support of Ron Sun and Jun Tani. Thanks also to Luis Pereira and the anonymous reviewers of this journal for constructive comments on earlier drafts. Finally, this paper began as a conference talk the travel to which was funded in part by MBR 2012, and sets out a collective limit on self-abduction as currently under development for a monograph for SAPERE, so would not have been finished in this form without opportunities afforded by Lorenzo Magnani. Grazie mille.

References

Allen P (1982) 'Ought' from 'Is'? What Hare and Gewirth should have said. Am J Theol Philos 3:90–97

Alnajjar F, Yamashita Y, Tani J (2013) The hierarchical and functional connectivity of higher-order cognitive mechanisms: neurorobotic model to investigate the stability and flexibility of working memory. Front Neurorobot 7:1–13

Aristotle, Jowett B (2000) Politics. Dover Publications, Mineola

Backhouse RE, Bateman BW (2009) Keynes and capitalism. Hist Polit Econ 41:645–671

Birch J (2013) On the 'Simulation Argument' and selective skepticism. Erkenntnis 78:95–107

Bostrom N (2003) Are we living in a computer simulation? Philos Q 53:243–255

Bostrom N, Kulczycki M (2011) A patch for the simulation argument. Analysis 71:54–61

Chomsky N (1973) One man's view: Noam Chomsky interviewed by an anonymous interviewer. Bus Today. http://chomsky.info/197305__/. Accessed 15 Oct 2015

Cogburn J, Silcox M (2014) Against brain-in-a-vatism: on the value of virtual reality. Philos Technol 27:561–579

Davies D (1997) Why one shouldn't make an example of a brain in a vat. Analysis 57:51–59

Dostoyevsky F, McDuff D (2003) The brothers Karamazov: a novel in four parts and an epilogue. Penguin, London

Dyson F (2015) The Register, Science. http://www.theregister.co.uk/2015/10/11/freeman_dyson_interview/?page=1. Accessed 15 Oct 2015

Gok SE, Sayan E (2012) A philosophical assessment of computational models of consciousness. Cogn Syst Res 17–18:49–62

Habermas J (1996) Between facts and norms: contributions to a discourse theory of law and democracy. MIT Press, Cambridge

Heidegger M, Stambaugh J (1996) Being and time: a translation of Sein und Zeit. State University of New York Press, Albany

Heintzelman S, King L (2013) The origins of meaning: objective reality the unconscious mind and awareness. In: Hicks JA, Routledge C (eds) The experience of meaning in life: classical perspectives emerging themes and controversies. Springer, Dordrecht, pp 87–99

Huemer M (2000) Direct realism and the brain-in-a-vat argument. Philos Phenomenol Res 61:397–414

Huxley A (1958) Brave new world revisited. http://www.huxley.net/bnw-revisited/. Accessed 15 Oct 2015
Keynes JM (1920) The economic consequences of the peace. Harcourt, Brace and Howe, New York
Kultgen J H, Lenzi M, Annual Conference of the Concerned Philosophers for Peace (2006) Problems for democracy. Rodopi, Amsterdam
Magnani L (2009) Abductive cognition: the epistemological and eco-cognitive dimensions of hypothetical reasoning. Springer-Verlag, Berlin
Magnani L (2015) Naturalizing logic. J Appl Log 13:13–36
Murata S, Arie H, Ogata T, Sugano S, Tani J (2014) Learning to generate proactive and reactive behavior using a dynamic neural network model with time-varying variance prediction mechanism. Adv Robot 28:1189–1203
Nietzsche FW, Large D (1998) Twilight of the idols. Oxford University Press, Oxford, Or How to Philosophize with a Hammer
Nietzsche FW, Large D (2007) Ecce homo: how to become what you are. Oxford University Press, Oxford
Obama B (2014) Remarks by President Obama in Address to the United Nations General Assembly. Office of the Press Secretary. https://www.whitehouse.gov/the-press-office/2014/09/24/remarks-president-obama-address-united-nations-general-assembly. Accessed 15 Oct 2015
Pinker S (2011) The better angels of our nature: why violence has declined. Viking, New York
Pope Francis (2015) Udienza a bambini e ragazzi di Scuole italiane, partecipanti alla manifestazione promossa da "La Fabbrica della Pace" Official Vatican Network. http://www.news.va/it/news/253366. Accessed 15 Oct 2015
Putnam H (1982) Reason truth and history. Cambridge University Press, Cambridge
Rashdall H (1914) Is conscience an emotion?. Three Lectures on Recent Ethical Theories, Houghton Mifflin
Richetin J, Sengupta A, Perugini M, Adjali I, Hurling R, Greetham D, Spence M (2010) A micro-level simulation for the prediction of intention and behavior. Cogn Syst Res 11:181–193
Sun R (2001a) Cognitive science meets multi-agent systems: a prolegomenon. Philos Psychol 14:5–28
Sun R (2001b) Individual action and collective function: from sociology to multi-agent learning. Cogn Syst Res 2(1):1–3
Sun R (2006) Cognition and multi-agent interaction: from cognitive modeling to social simulation. Cambridge University Press, Cambridge
Sun R (2009) Motivational representations within a computational cognitive architecture. Cognit Comput 1:91–103
Sun R (2012) Grounding social sciences in cognitive sciences. MIT Press, Cambridge
Sun R (2013) Moral judgment, human motivation, and neural networks. Cogn Comput 5(4):566–579
Sun R, Coward LA, Zenzen MJ (2005) On levels of cognitive modeling. Philos Psychol 18:613–637
Tani J (1998) An interpretation of the 'self' from the dynamical systems perspective: a constructivist approach. J Conscious Stud 5:516–542
Tani J (2014) Self-organization and compositionality in cognitive brains: a neurorobotics study. Proc IEEE 102:586–605
Tani J (2015) Exploring robotic minds: actions, symbols, and consciousness as self-organizing dynamic phenomena. Oxford University Press, Oxford (in press)
Tani J, Ito M, Sugita Y (2004) Self-organization of distributedly represented multiple behavior schema in a mirror system. Neural Netw 17:1273–1289
Tolstoy L (1891) Ivan the fool. C L Webster and company, New York
Wells T (2013) Adam Smith on morality and self-interest. In: Luetge C (ed) Handbook of the philosophical foundations of business ethics. Springer, Dordrecht, pp 281–296
White JB (2006) Conscience: toward the mechanism of morality. Dissertation, University of Missouri-Columbia. http://hdl.handle.net/10355/4327. Accessed 15 Oct 2015
White J (2010) Understanding and augmenting human morality: an introduction to the ACTWith model of conscience. In: Magnani L (ed) Model-based reasoning in science and technology: abduction, logic and computational discovery. Springer, Berlin, pp 607–621
White J (2012) An information processing model of psychopathy and anti-social personality disorders integrating neural and psychological accounts towards the assay of social implications of psychopathic agents. In: Fruili AS, Veneto LD (eds) Psychology of morality. Nova Science Publishers, Hauppauge, pp 1–34
White J (2013) Manufacturing morality: a general theory of moral agency grounding computational implementations. In: Floares A (ed) Computational intelligence. Nova Publications, Hauppage, pp 163–210
White J (2014) Models of moral cognition. In: Magnani L (ed) Model-based reasoning in science and technology: theoretical and cognitive issues. Springer, Berlin, pp 363–391
Wilson EO (1998) Consilience: the unity of knowledge. Knopf, New York
Yamashita Y, Tani J (2012) Spontaneous prediction error generation in schizophrenia. PLoS One 7:e37843. doi:10.1371/journalpone0037843