frontiers in
PSYCHOLOGY

PERSPECTIVE ARTICLE
published: 17 March 2011
doi: 10.3389/fpsyg.2011.00044

Regarding reality: some consequences of two incapacities

Shimon Edelman*

Department of Psychology, Cornell University, Ithaca, NY, USA

Edited by:
Jay L. Garfield, Smith College, USA

Reviewed by:
Dan Lloyd, Trinity College, USA
Mark Siderits, Seoul National University, South Korea

****Correspondence:***
Shimon Edelman, Department of Psychology, Cornell University, Ithaca, NY 14853, USA.
e-mail: se37@cornell.edu

By what empirical means can a person determine whether he or she is presently awake or dreaming? Any conceivable test addressing this question, which is a special case of the classical metaphysical doubting of reality, must be statistical (for the same reason that empirical science is, as noted by Hume). Subjecting the experienced reality to any kind of statistical test (for instance, a test for bizarreness) requires, however, that a set of baseline measurements be available. In a dream, or in a simulation, any such baseline data would be vulnerable to tampering by the same processes that give rise to the experienced reality, making the outcome of a reality test impossible to trust. Moreover, standard cryptographic defenses against such tampering cannot be relied upon, because of the potentially unlimited reach of reality modification within a dream, which may range from the integrity of the verification keys to the declared outcome of the entire process. In the face of this double predicament, the rational course of action is to take reality at face value. The predicament also has some intriguing corollaries. In particular, even the most revealing insight that a person may gain into the ultimate nature of reality (for instance, by attaining enlightenment in the Buddhist sense) is ultimately unreliable, for the reasons just mentioned. At the same time, to adhere to this principle, one has to be aware of it, which may not be possible in various states of reduced or altered cognitive function such as dreaming or religious experience. Thus, a subjectively enlightened person may still lack the one truly important piece of the puzzle concerning his or her existence.

Keywords: simulation, statistical inference in cognition, dreaming, virtual reality, experimental metaphysics

A realist is simply one who knows no more recondite reality than that which is represented in a true representation.

– Peirce (1868)

Is this the real life? Is this just fantasy?
…
Ooh yeah, ooh yeah
Nothing really matters
Anyone can see
Nothing really matters – nothing really matters to me

– Freddy Mercury/Queen
(*Bohemian Rhapsody*, 1975)

1 THE SEEDS OF DOUBT

The suspicion that the world is not quite, or maybe even not at all, what it seems has a long history of being toyed with by those who are predisposed to metaphysical speculation. It also has a long history of being roundly rejected by practically minded people – a category that includes most off-duty metaphysicians[1]. The present paper is an attempt to understand both the perennial philosophical appeal of questioning reality and people's routine acceptance of reality at the face value, by considering metaphysical insights into this singularly important aspect of the human condition against the background of some recent developments in cognitive and computer sciences.

A metaphysical doubt of reality may arise from such a common human experience as waking from a dream, surrounded by fleeting memories of another world that hint at the possibility of a deeper reality hiding behind waking life itself. Among the countless expressions of this experience, the one by Tzu (1968, Chapter 2, p. 49), which dates back to the fourth century BCE, stands out: "Once Chuang Tzu dreamt he was a butterfly, a butterfly flitting and fluttering around, happy with himself, and doing as he pleased. He did not know he was Chuang Tzu. Suddenly he woke up and there he was, solid and unmistakable Chuang Tzu. But he did not know if he was Chuang Tzu who had dreamt he was a butterfly, or a butterfly dreaming he was Chuang Tzu."

In Western philosophy, a succinct statement of the case for doubting reality is found in *Discourse on Method* by Descartes (1637, IV): "When I considered that the very same thoughts (presentations) which we experience when awake may also be experienced when we are asleep, while there is at that time not one of them true, I supposed that all the objects (presentations) that had ever entered into my mind when awake, had in them no more truth than the illusions of my dreams." Between Chuang Tzu (or the butterfly that dreamed him) and Descartes, it would seem that philosophical inquiry has gotten to the bottom of the reality issue, and that no stone has remained unturned in the process.

Science, however, keeps expanding its own range of inquiry, and each newly added field comes with its share of stones under which there lurk implications for, or even answers to, philosophical questions (and, often enough, new questions). The question of

[1]I hasten to remark that this observation merely echoes a line from *Tlön, Uqbar, Orbis Tertius*: "Hume noted for all time that Berkeley's arguments did not admit the slightest refutation nor did they cause the slightest conviction." (Borges, 1941/1962)

the nature of experienced reality and its veracity is a case in point: along with many others, this question's import has been irrevocably transformed by the realization that a mind is made of computations, and that these computations construct a virtual reality.

2 THE MIND AS A VIRTUAL REALITY MACHINE

Every physical process – that is to say, every process – in the universe computes something. Indeed, if a ripe apple falling from a tree and the planet from which the tree grows did not compute their trajectory with respect to each other, they would be at a loss as to how fast and which way to move (Edelman, 2008b). While this observation implies that minds too are essentially and literally, not metaphorically, computational, not every process qualifies as a mind.

Succinctly put, a mind is a persistent bundle of computations over *representations* – state-space trajectories that reliably reflect, in a counterfactually consistent manner, the dynamics of physical processes that are external (that is, only weakly coupled) to the system that implements it. The arguments and evidence for this understanding of the nature of minds and its implications for cognitive science have been stated elsewhere and shall not be repeated here (for details and references to prior work, see Edelman, 2008a).

Whatever external reality a mind is immersed in, it can only access it indirectly: "For, since the things the mind contemplates are none of them, besides itself, present to the understanding, it is necessary that something else, as a sign or representation of the thing it considers, should be present to it: and these are ideas" (Locke, 1690, Book IV, Chapter XXI, p. 4). This hard constraint does not undermine the possibility of reliable knowledge of the outside world[2]. Indeed, the discipline of ecological psychology has been steadily adding to our understanding of the conditions for faithful representation (see, e.g., Shepard, 1984). Moreover, neurally mediated representations maintained by brains are generically veridical (Edelman, 1998, 1999). However, as noted most poignantly by Merker (2007), many of those representations are also decidedly virtual.

In human vision, for instance, the binocular visual field is synthesized (by a set of sophisticated stereopsis algorithms) from two monocular sources of data. My perception of the visual world is phenomenally fused or "cyclopean" despite having two disparate physical sources. Moreover, the phenomenal focus of my "first person perspective" (Metzinger, 2004) is located inside my skull, a couple of centimeters behind the bridge of my nose, where it is in fact perpetually dark. These simple observations demonstrate not only that the world as I see it is virtual, but also that the central aspect of my phenomenal self – the "I" that sees the world – is a construct.

Philosophical analysis (Metzinger, 2003) and neurobiological data (Merker, 2007) extend these insights to encompass the nature of the self in general. As Metzinger (2005) puts it, the mind/brain functions as a "total flight simulator": a virtual reality rig that, in addition to collecting and processing the information about the world that is needed to steer the body, also simulates the pilot – a virtual entity that is entrusted with the control of the body for as long as it is awake.

Insofar as it is charged with maintaining representations of the environment that can be used to support foresight, a person's mind also endeavors to represent other people, as a means of facilitating and furthering social interactions (Edelman, 2008a, Chapter 6). Depending on how familiar another person is to me, my simulation of him or her can be more or less detailed, in some cases affording quite accurate predictions concerning what the other person is likely to say or do in a given situation. Thus, autonomous agents who populate my virtual reality may be quite independent of me (in case they are representations of actual other people, living or dead) or merely aspects of myself (in case they are simulations that I maintain of other people).

Given that both the "experiencer" (the self) and the universe that it experiences are virtual constructs, it makes sense to drop the distinction between the two. This move, exemplified in Western thought by the writings of Mach (1886) and in the East by the teachings of Buddhist philosophers such as Vasubandhu (Scharfstein, 1998; Siderits, 2007), assigns primacy to the content of experience itself. The natural explanatory tool for this purpose is, as I already noted above, computation. Although explaining experience has been long considered the "hard problem" of philosophy of mind (Chalmers, 1995), computational accounts of experience are now being offered in terms of the dynamic pattern of information in certain representational systems (Spivey, 2006; Tononi, 2008; Fekete and Edelman, 2011). What does a computational theory of mind and experience have to say about the reality issue?

3 NESTED REALITIES

Computation has a curious property that is directly relevant to the reality question: it admits multiple equivalent realizations. A given computation can be instantiated in different forms or substrates, as long as the relevant aspects of their organization are identical. In other words, computation is an organizational invariant (Chalmers, 1994). As a consequence, carrying out a computation in some substrate $\mathcal{S}$ is indistinguishable from simulating it (that is, from performing it in a different substrate $\mathcal{S}'$ that is identical to $\mathcal{S}$ on a certain level of organization).

Thus, when an $\mathcal{S}$-brain that would compute a mind M (if allowed to play out its dynamics) is itself simulated down to the appropriate level of detail within $\mathcal{S}'$, the mind M' that arises as a result is indistinguishable from M. In particular, the phenomenal experience of being M' within $\mathcal{S}'$ is the same as that of M within $\mathcal{S}$[3]. To privilege the experience of one instantiation of a mind over another is the same as to claim that the trajectories of two cannonballs falling side by side are not really the same because one of the cannonballs is made of iron and the other of copper.

Consider now the special situation in which one computational substrate, while being fully capable of simulating a brain along with a sufficiently large chunk of its environment, is wholly contained within another: $\mathcal{S}' \subset \mathcal{S}$. Although the extra chunk of environment is necessary (because minds are not confined to individual brains; cf. Dretske, 1995; Edelman, 2008a, Chapter 10), it does not have to extend to all of $\mathcal{S}$, and so true containment

[2]Nor does it imply that all representations are internal in the sense of being confined to one's skull (O'Regan and Noë, 2001).

[3]Thus, a simulated hurricane would feel very real, and indeed can prove fatal, to a person who happens to reside in the same simulation.

is possible. To an observer situated outside $\mathcal{S}'$, a mind M' that is contained within $\mathcal{S}'$ has less "elbow room." Would this make any difference to M' itself?

A familiar example of the containment situation is a multitasking computer system, such as my notebook, which can run a number of applications in parallel, each inside a dedicated shell, all the while interacting with me via a text editor in a separate window. One of those applications can be a neuron simulator. Had someone taken a sufficiently detailed scan (Sandberg and Bostrom, 2008) of the brain of René Descartes back around 1640, and were my notebook many orders of magnitude faster (so as to simulate the temporal dynamics of a brain on a fine enough scale; cf. Fekete and Edelman, 2011), I would have been able to obtain that philosopher's running commentary on where I am going with the present argument.

Another example of reality containment is a human brain that is asleep and dreaming. During sleep, the thalamocortical pathway is partially inhibited and activity in several brain areas (notably, prefrontal, and primary visual cortices) is reduced, while others remain fully active (Hobson et al., 2000; Muzur et al., 2002). The resulting functional network, which is anatomically and physiologically fully contained within the network formed by the waking brain, has a peculiar dynamics, which apparently has many uses, notably in learning and memory consolidation (Karni et al., 1994; Dave and Margoliash, 2000; Lee and Wilson, 2002; Stickgold and Walker, 2004; Wagner et al., 2004; Stickgold, 2005; Ji and Wilson, 2007; Peyrache et al., 2009). The phenomenal experience that arises from this dynamics is that of the dream self, situated in a dream world.

Let us take a moment to appreciate the nature and the full extent of the phenomenal (experiential) predicament of an agent who doubts the reality of its experience. Traditionally, this doubt could only be assuaged through religion. Thus, Descartes professed a belief in a benevolent deity as a guarantor of reality being true, while his contemporary, the playwright Pedro Calderón de la Barca, made the protagonist in *La vida es sueño* declare that even though life may be a dream, one still has to act as if it is real in obeying God.

Nowadays, the case for doubt is, if anything, stronger, while many formerly popular varieties of belief have devolved into folly. On the one hand, the realization that the dynamics of a computational system determines all of its cognition, including its phenomenal experience (Spivey, 2006; Fekete and Edelman, 2011), implies that a brain that is embodied and situated in the world will *experience* the same reality as its simulacrum that is situated within a sufficiently faithful virtual copy of the observable universe. On the other hand, the realization that a cognitive agent capable of phenomenal experience can be simulated by any punk with a computer suggests that a belief in a supreme being that would not deceive its creatures is a tad too optimistic.

Phenomenology and faith aside, we may still wonder whether or not science has a say in this matter, and if yes, whether its lessons would differ from those of intuition. After all, we experience the sun as revolving around the earth, and it used to be an item of religious orthodoxy that it does, yet basing a space exploration program on that premise would doom it from the outset (as most of us would agree). Is there is a principled, empirically grounded, and theoretically viable means for an agent to decide whether or not it resides in what Bostrom (2003a) calls the "basement reality" rather than in a simulation or in someone else's dream?

4 ON THE POSSIBILITY OF EXPERIMENTAL METAPHYSICS

While probabilistic arguments from first principles regarding certain aspects of reality are being widely discussed by cosmologists and philosophers (see for instance the extensive literature on anthropic reasoning and self-locating belief; Bostrom, 2003b), there does not seem to be much in the way of systematic inquiry into the possibility of observing or actively testing an environment so as to yield clues regarding its location in a potential simulation hierarchy. One line of work that may eventually prove relevant to the self-location issue examines interreality – situations in which coupling exists between the present consensual reality, which I shall refer to as $\mathcal{R}$, and a virtual or *simulated* reality $\mathcal{R}_S$ contained within it.

For example, an interreality system in which a physical pendulum within $\mathcal{R}$ is coupled to a virtual pendulum within $\mathcal{R}_S$ behaves, not entirely surprisingly, like a real pair of coupled pendulums (Gintautas and Hübler, 2007). The dynamics of an interreality system can, of course, be much more complex: coupling a person to an appropriately engineered virtual environment has even been claimed to have therapeutic value in treating post-traumatic stress disorder (Riva et al., 2010).

Can these ideas be developed into a method for determining whether or not $\mathcal{R}$ is contained within some larger *simulating* reality ${}^{S}\mathcal{R}$? Gintautas and Hübler (2007) show that in a coupled real/virtual pendulum system, a parametric manipulation on one side of the divide can be felt on the other side, both in the real-to-virtual and in the virtual-to-real directions. As one would expect in this system, under certain conditions the effect is sharp or, in the standard jargon of dynamical systems theory, catastrophic.

This suggests that if our $\mathcal{R}$ is imperfectly contained within some ${}^{S}\mathcal{R}$, an event in the latter could in principle manifest itself to us as a state-space transition that is (i) catastrophic, and (ii) inexplicable, even with the full knowledge of the situation and the physics of $\mathcal{R}$ – in other words, an apparent *miracle*[4]. A symmetrical situation is one in which something that happens in $\mathcal{R}$ percolates up to ${}^{S}\mathcal{R}$ and causes a catastrophic event there. This latter scenario may be quite frequent in universes that are poorly engineered in that they fail to provide adequate protection from simulated events, which makes them susceptible to a *system crash*[5].

Unfortunately, even a confirmed miracle still leaves us with the need to distinguish among several possibilities. The miracle may be due to (1) a true top-down reality dysfunction, in which information seeps from ${}^{S}\mathcal{R}$ into $\mathcal{R}$, (2) a limitation of our knowledge of the physics of $\mathcal{R}$, or (3) an imperfection or inconsistency in the functioning of $\mathcal{R}$ itself. With regard to (3), the common assumption that the universe must obey a fixed set of globally consistent and

[4]As per Hume's *Enquiry*, "A miracle may be accurately defined, *a transgression of a law of nature by a particular volition of the Deity, or by the interposition of some invisible agent.*" (Hume, 1748, Book X, Part I)

[5]In my computer, protection from contained virtual universes is provided by the operating system (OS). Under a good OS, an embedded agent's attempts to overstep the bounds of the memory space allocated to it would either fail quietly or cause it to be terminated. In a poor OS, the entire system may be compromised. In my $\mathcal{R}$, such imperfect OS tend to have names that begin with "W."

immutable principles is pure wishful thinking, rooted in the medieval scholastic tradition of imagining a perfect creator. Moreover, the possibility of (2) could be exacerbated by intervention from without (that is, from ${}^{S}\mathcal{R}$) that tampers with the observer's knowledge of physics. Thus, for all we know, the above three possibilities are not mutually exclusive, which further complicates our predicament.

While convincing miracles are at best exceedingly rare in $\mathcal{R}$, they abound in the dream state. Dreaming, as noted earlier, is a kind of simulated reality $\mathcal{R}_S$ that is (imperfectly) contained within $\mathcal{R}$ by virtue of the particulars of the brain functional architecture and the dynamics of sleep. The containment in this case is imperfect in at least two senses. First, outside data can seep into a dream, affecting its contents, and possibly waking the dreamer up (Coenen, 2010)[6]. Second, events that transpire within a dream sometimes strike the dreamer as out of the ordinary or bizarre.

The feeling of bizarreness is a symptom of imperfect reality containment (of $\mathcal{R}_S$ within $\mathcal{R}$) because something can only appear extraordinary when compared to the ordinary, and so for a situation in $\mathcal{R}_S$ to appear bizarre, some record of the ordinary must be retained by $\mathcal{R}$ as it "shrinks" to become $\mathcal{R}_S$, or else seeps in later, just as real-time outside data do. Crucially, one would be justified in concluding that the present situation is bizarre only if the difference between it and a baseline (a memory of ordinary reality) is statistically significant[7]. In judging reality, therefore, the mind must rely on statistics just as it does in everything else (Edelman, 2008a) – an unsurprising finding, given that, as David Hume pointed out, "all knowledge resolves itself into probability" (Hume, 1740, Part IV, Section I).

5 USING STATISTICS TO TEST REALITY

Putting in place and maintaining a model of the statistical structure of the world, so as better to predict it, is an overarching computational task of the brain (Craik, 1943; Ingvar, 1985; Grush, 2004; Edelman, 2008a). The need to learn the statistics of the world is especially pressing during development, which is when powerful computational principles are brought to bear on seeking patterns in perceptual data, social cues, and motor behavior (Goldstein et al., 2010). Among such patterns, the suspected causal ones are particularly important. A situated cognitive system, whether or not it is itself endowed with a sense of agency, would do well to discern other agents at work. Causal inference (Gopnik et al., 2004) thus plays a key role in cognitive development.

In statistical testing of reality, causal inference may take the form of *anomaly or outlier detection* – a task that is intensively studied in computer science because of its many applications, for instance in credit card fraud, network intrusion detection, system fault diagnosis, and health monitoring (Hodge and Austin, 2004; Patcha and Park, 2007). Chandola et al. (2009) offer a useful survey of problems and methods in anomaly detection, although their taxonomy is confused[8].

Another relevant set of techniques aims at *intrusion detection* in situations in which information is exchanged between multiple agents and the challenge is to find a rogue one that does not belong to the network. For example, Li and Joshi (2009) describe a method for distributed detection of rogue nodes in an *ad hoc* mobile network, which is based on feeding gossip among nodes into a Dempster–Shafer evidence evaluator[9].

In an information-exchange network about which one may not assume that all bona fide agents have cryptographically verifiable identities, intrusion detection cannot be signature-based and thus must rely on anomaly detection (Patcha and Park, 2007). Anomaly detection is also the only way to stop an ongoing "zero-day" attack – one that uses a hitherto unknown method, for which a canned defense is by definition unavailable (this situation is analogous to an immune system encountering a radically novel antigen).

These last observations highlight the Achilles' heel of testing reality for bizarreness. All statistical methods for anomaly detection need a *baseline* – a stored body of data that captures the regular state of affairs to which new data are to be compared. Baseline data, in turn, need verification and protection from reality intrusion. It would appear that here too computer science comes to the rescue, in the form of cryptographic methods for safeguarding data integrity.

6 CRYPTOGRAPHY TO THE RESCUE?

If we assume, as we must, that a process in the "outer" reality ${}^{S}\mathcal{R}$ may have access to any part of the contained reality $\mathcal{R}$, then a cryptographic scheme for guarding $\mathcal{R}$ must be such that both the encryption algorithm and the encryption key (used by all but the simplest schemes) are tamper-resistant[10]. Gennaro et al. (2004) consider the problem of providing security against an adversary (think of Descartes' evil demon) who is allowed to apply arbitrary computationally feasible functions to the secret key. They prove that it is generally impossible to achieve this type of ATP (algorithmic tamper-proof) security. Although ATP security does become feasible when certain conditions are imposed on the task[11], these conditions make the solution irrelevant to our problem of reality authentication.

In an intriguing recent development, Armknecht et al. (2009) introduced physically unclonable functions (PUFs) – a cryptographic means for implementing a key that maps challenges to responses which are highly dependent on the physical properties of the device in which the PUF is embedded. For instance, the state of

[6]The dreamer's curious metaphysical predicament in such cases has been expressed by Ludwig Wittgenstein, in a passage written 2 days before his death on April 29, 1951: "I cannot seriously suppose that I am at this moment dreaming. Someone who, dreaming, says 'I am dreaming', even if he speaks audibly in doing so, is no more right than if he said in his dream 'it is raining', while it was in fact raining. Even if his dream were actually connected with the noise of the rain." (Wittgenstein, 1972, p. 90e)

[7]We must remember that a faulty decision device may raise the bizarreness flag when in fact there is nothing out of the ordinary. Furthermore, it may be subverted into doing so by an intervention from the containing reality.

[8]Chandola et al. (2009) distinguish among classification, nearest-neighbor, clustering, generative, information-theoretic (Kolmogorov), and spectral (dimensionality reduction) approaches, of which only the generative methods they call "statistical." They also classify anomalies as point, contextual, and collective.

[9]The Dempster–Shafer approach differs from Bayesian inference in that probability is replaced by an uncertainty interval bounded by belief and plausibility.

[10]They do not have to be absolutely tamper-proof, if a high probability of success suffices and can be guaranteed. Relaxing the requirements in this manner often leads to tractable solutions to hard problems in computer science.

[11]They show that security in this model can be achieved if and only if the system has: (1) a self-destructing capability, and (2) a publicly available hardwired data from a separate server that cannot be tampered with. However, ATP security against an adversary limited to differential fault analysis (flipping random bits) is possible without (1) and (2).

a static random-access memory (SRAM) upon power-up appears random to an outside observer but is in fact closely determined by the physical details of the device that implements the SRAM.

More to the point for biological agents such as myself, there is certainly enough idiosyncrasy in the dynamics of the human brain to support PUF-based cross-verification of some of the mind's processes by others. It may be that this computational mechanism for self-authentication is behind the perceived continuity of the self across periods of temporary self-alteration (sleep) or even near self-dissolution (deep anesthesia, the cognitive recovery from which is typically complicated and prolonged relative to waking up).

For a PUF to be useful for reality authentication, however, the adversary must not be allowed access to the innards of the physical device that implements it – a condition that, alas, cannot be taken for granted[12]. At the very least, though, the use of PUFs raises the ante in the reality game: the supreme being from ${}^{S}\mathcal{R}$ would have to tamper either with the dynamics of a very complex system such as an entire brain, or with the fundamental physical laws of $\mathcal{R}$, to effect even a simple intervention such as flipping one memory bit inside the mind of a PUF-using agent in $\mathcal{R}$.

7 CHANGE WE CAN BELIEVE IN

The potentially staggering computational complexity of making a change in a simulated reality that would go unnoticed by its denizens has implications for our predicament. The issue here is not merely a high computational cost of a clandestine intervention but rather its very feasibility: as it is well known in computer science, certain problems – those whose complexity grows exponentially with the relevant measure of problem "size" (say, the number of neurons in a brain that is to be prevented from noticing that it is being tampered with) – are essentially intractable (Garey and Johnson, 1979). This suggests that a perceived inexplicable change in, or bizarre quality of, one's reality R should be treated as a potentially revealing clue to its true nature, simply because either falsifying or masking such clues would be so hard.

Even if an agent within $\mathcal{R}$ is not cryptographically protected by a PUF from any but the most determined and pervasive tampering coming from ${}^{S}\mathcal{R}$, there is another reason for it to trust a perceived change: the computational complexity of truth maintenance. Specifically, it may be too costly or infeasible to track down all the beliefs (entries in a database) that need to be modified if one of them happens to change (this is known as the Frame Problem in AI and in epistemology; McCarthy and Hayes, 1969; Shanahan, 2009).

Suppose that the ${}^{S}\mathcal{R}$-agent in charge of the simulation of $\mathcal{R}$ is less than omnipotent, in that it lacks the resources needed for clandestine intervention. Suppose, further, that it wishes to avoid being discovered from within $\mathcal{R}$. Its only recourse in this case is to launch the simulation of $\mathcal{R}$ and then to leave it alone. The theological repercussions of this essentially computational argument (e.g., parallels to the Gnostic idea of a demiourgos, or a flawed creator; cf. Waldstein and Wisse, 1995) are mostly beyond the scope of the present paper. I cannot refrain from remarking, however, that the impossibility of upgrading the universe without the requisite tampering being noticed from within it offers an intriguing explanation for the rather poor shape that it is in.

8 LIKELY OBJECTIONS

The foregoing analysis of the doubter's predicament depends critically on the applicability of the concept of simulation to computational systems capable of having experiences – that is, to minds[13]. As Fekete and Edelman (2011) argue, experience must be an *intrinsic* property of a system's dynamics – one that does not require interpretation from without and is thus not merely a matter of choosing a particular description for the system in question. The notion of a simulated mind would thus seem to be self-contradictory, insofar as what really matters is the intrinsic dynamics of the simulator (say, my laptop's electrical circuits), rather than the dynamics of the virtual brain that it computes. This objection becomes especially poignant if the simulator is digital (that is, if its intrinsic dynamics is discrete).

I believe that this concern dissipates with the observation that a complex system may possess multiscale intrinsic dynamics (Bar-Yam, 2004; Shalizi, 2005; Halley and Winkler, 2008; cf. Edelman, 2008b, Section 5, *How a continuous state space gets its spots*). In such a system, different scales constitute distinct inherent levels of operation (rather than merely levels of external description). Accordingly, a simulated phenomenal mind can emerge on a certain level of a properly engineered multiscale system, whose lower levels would remain devoid of experience.

Having noted that, one still wonders whether or not a digital computer can be properly engineered so as to simulate a mind. To put things into perspective here, observe that a digital computer is a discrete dynamical system that is itself simulated by a continuous-voltage electronic circuit, in which the current is carried by discrete entities, whose quantum-mechanical properties (some discrete, others continuous) may or may not matter, depending on the implementation details. If it is indeed possible to simulate a mind in a digital system, intervening in it would be computationally more feasible than in an analog simulation (e.g., by suspending time and computing the requisite next state "offline"). This and other interesting issues will be treated elsewhere (Fekete and Edelman, in preparation).

9 SOME CONSEQUENCES

The present exploration of the reality question leads to four conclusions. The first two complement each other nicely, but seem to amount to little progress relative to the starting point of our discussion:

- *Epistemological Ceiling.* The incapacity in principle of ruling out pervasive clandestine reality tampering implies that an agent can never tell for sure whether or not its reality is fundamental (in Bostrom's terminology, "basement").
- *Phenomenal Indifference.* The incapacity in principle of distinguishing "real" from "simulated" phenomenal experience implies that subjectively it does not matter – in both cases, the agent feels the same.

[12]This indicates that the idea of personal "totems" for testing reality, which figures prominently in Christopher Nolan's 2010 film *Inception*, is not workable.

[13]I am indebted to T. Fekete for pointing this out to me in a private communication. Egan's (1994) novel *Permutation City* is a highly entertaining informal introduction to a host of related concerns.

In comparison, the second two conclusions, which ensue when certain computational facts (regarding what is feasible in the way of information manipulation) are combined with certain metaphysical assumptions (regarding the possible motivations of beings that are capable of simulating entire realities), do have practical repercussions:

- *Credible Change.* It makes sense to treat statistically significant perceived bizarreness of certain aspects of reality as evidence for it being manipulated, perhaps as a part of a broader simulation.
- *Ontological Indifference.* The absence of reliable evidence to this effect indicates either that the present reality is fundamental or that the agent that started off its simulation has subsequently withdrawn from intervening in it – two alternatives that for all practical purposes are one and the same.

Of the latter two conclusions, the Credible Change principle is clearly at work when a dreamer's realization of the bizarreness of the dream causes its disruption, such as waking up, or the onset of lucidity, which is a conscious realization of being in a dream (LaBerge, 1990).

In estimating bizarreness, care must be taken not to err on the side of sounding a false alarm: after all, waking life is at times stranger than a dream[14]. Everett Ruess, a young artist who in 1934 disappeared in the canyons of the Escalante River in the Utah wilderness (a surreal place, if there ever was any), wrote in one of his last letters home: "Often as I wander, there are dream-like tinges when life seems impossibly strange and unreal. I think it is, too, except that most people have so dulled their senses that they do not realize it." In the face of this predicament, and in the light of the principles listed above, the rational course of action would seem to be to take reality at the face value, and to cast doubt aside, just as Descartes did, for reasons of his own.

There is, of course, a venerable school of thought – the doctrine set in motion by the Buddha, or the Awakened One – that preaches the realization of unreality. Insofar as each mind's phenomenal world, and indeed the mind itself, is virtual in the cognitive-scientific sense (and therefore could be said to have no independent, unconditioned nature, or *svabhāva*), the *śūnyatā* or emptiness postulate of Mahāyāna Buddhism (Siderits, 2007, Chapter 9) happens to be literally true, as noted by Metzinger (2003; cf. Edelman, 2008a, Chapter 9). This, however, only shows that the emptiness principle is (somewhat paradoxically) the least consequential of the aspects of reality that can be profitably pondered. In the present paper, I tried to identify some of the more interesting questions to think about, and perhaps to broaden the discussion so as to draw on scientific methods in addition to metaphysical speculation.

What I believe I should offer in concluding this discussion is a word of caution for anyone who would forgo a thorough understanding of the nature of reality in favor of a Zen-like insight into it. The Phenomenal Indifference and the Epistemological Ceiling principles, each rooted in a different variety of fundamental incapacity, conspire to pull the rug from under any such insight. Thus, a smart agent should know better than taking too seriously *any* piece of experience or knowledge, including its own enlightenment[15]. Of course, the very same principles also imply that the agent's mind can be secretly tampered with, in which case it may feel smart, enlightened, and in touch with the deepest level of existence, all the while remaining a puppet of a game designer from a higher reality.

POSTSCRIPT

As pointed out by a reviewer, content externalism, as expressed for instance in Davidson's (2001) triangulation argument, would seem to rule out the possibility of the kind of sweeping metaphysical doubt of reality that this paper ultimately refuses to repudiate. According to Davidson, objectivity (and hence doubt) requires two creatures to create a "baseline" of truth against which an object, as the third vertex of the "triangle," can be evaluated. When combined with a standard brain-in-a-vat scenario, this notion arguably renders metaphysical doubt from within the vat, as it were, less than fully coherent.

Now, Davidson's argument can in turn be argued with (see, e.g., Bridges, 2006, or Glüer-Pagin, 2006, who writes, "Any 'norm' for truth and mistake determining these in relation to the reactions of fellow creatures would seem to determine them regardless of the actual presence or absence of those fellows"). We need not, however, engage in that debate here. Even if one accepts the triangulation argument at face value, it has little bearing on the metaphysical doubt entertained by a creature that suspects, not that it is in reality a solitary brain-in-a-vat, but rather that it, along with all of its fellows, is wholly contained in a strictly larger reality.

By Davidson's criterion, this doubt is coherent within the contained world, because it can be meaningfully shared among the creatures that populate it. Indeed, as the introduction to this paper clearly states, I did not invent the doubt: I inherited it from the community. I might add that I do not see how I can rule out the possibility that some members of that community, such as the reviewer in question, are virtual fronts for the game designer, who is using them to thwart the publication of this paper (perhaps because by drawing the attention of creatures like myself to their predicament, it interferes with the game's goals).

ACKNOWLEDGMENTS

Many thanks to Rick Dale, Tomer Fekete, Dan Lloyd, and an anonymous *Frontiers* reviewer for their comments on drafts of this paper. I am grateful to Jakub Limanowski, whose term paper written for a seminar on consciousness and free will that I taught in Spring 2009 at Cornell started me thinking about telling dreams apart from reality; to Christopher Nolan for his entertaining film *Inception* (2010), which prompted me to complete the present paper after sitting on it for too long; and to Borges (1940/1970), whose short story *The Circular Ruins* may have been an inspiration for Nolan (I know it was for me).

[14]Richard Linklater's film *Waking Life* makes this case very convincingly.

[15]Cf. the following passage, referred to as Buddha's Zen: "I discern the highest conception of emancipation as a golden brocade in a dream, and view the holy path of the illuminated ones as flowers appearing in one's eyes." (Reps, 1989, pp. 86–87)

REFERENCES

Armknecht, F., Maes, R., Sadeghi, A., Sunar, B., and Tuyls, P. (2009). "PUF-PRFs: a new tamper-resilient cryptographic primitive," in *Advances in Cryptology – EUROCRYPT 2009*, eds V. Immler and C. Wolf (Berlin: Springer), 96–102.

Bar-Yam, Y. (2004). A mathematical theory of strong emergence using multiscale variety. *Complexity* 9, 15–24.

Borges, J. L. (1941/1962). "Tlön, Uqbar, Orbis Tertius," in *Ficciones*, ed. A. Kerrigan (New York: Grove Press), 17–34. Translated by A. Bonner in collaboration with the author.

Borges, J. L. (1940/1970). *The Aleph and Other Stories, 1933–1969*. New York: E. P. Dutton. Translated by Norman Thomas di Giovanni in collaboration with the author.

Bostrom, N. (2003a). Are you living in a computer simulation? *Philos. Q.* 53, 243–255.

Bostrom, N. (2003b). The mysteries of self-locating belief and anthropic reasoning. *Harv. Rev. Philos.* 11, 59–74.

Bridges, J. (2006). Davidson's transcendental externalism. *Philos. Phenomenol. Res.* 73, 290–315.

Chalmers, D. J. (1994). On implementing a computation. *Minds Mach.* 4, 391–402.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: a survey. *ACM Comput. Surv.* 41, Article 15.

Coenen, A. (2010). Subconscious stimulus recognition and processing during sleep. *Psyche* 16.

Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press.

Dave, A. S., and Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290, 812–816.

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

Descartes, R. (1637). "Discourse on the method of rightly conducting the reason and seeking the truth in the sciences," in *Volume 37 of The Harvard Classics*, ed. C. W. Eliot (New York: P. F. Collier & Son), 30. Edition published in 1909–1914.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press. The Jean Nicod Lectures.

Edelman, S. (1998). Representation is representation of similarity. *Behav. Brain Sci.* 21, 449–498.

Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: MIT Press.

Edelman, S. (2008a). *Computing the Mind: How the Mind Really Works*. New York: Oxford University Press.

Edelman, S. (2008b). On the nature of minds, or: truth and consequences. *J. Exp. Theor. Artif. Intell.* 20, 181–196.

Egan, G. (1994). *Permutation City*. London: Orion.

Fekete, T., and Edelman, S. (2011). Towards a computational theory of experience. *Conscious. Cogn.* (in press).

Garey, M. R., and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: W. H. Freeman.

Gennaro, R., Lysyanskaya, A., Malkin, T., Micali, S., and Rabin, T. (2004). "Algorithmic tamper-proof (ATP) security: theoretical foundations for security against hardware tampering," in *Theory of Cryptography, Volume 2951 of LNCS* (Berlin: Springer-Verlag), 258–277.

Gintautas, V., and Hübler, A. W. (2007). Experimental evidence for mixed reality states in an interreality system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 75, 057201.

Glüer-Pagin, K. (2006). "Triangulation," in *The Oxford Handbook of Philosophy of Language*, eds E. Lepore and B. Smith (Oxford: Oxford University Press), 1006–1019.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J., Schwade, J., Onnis, L., and Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends Cogn. Sci.* 14, 249–258.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychol. Rev.* 111, 3–32.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442.

Halley, J. D., and Winkler, D. A. (2008). Classification of emergence and its relation to self-organization. *Complexity* 13, 10–15.

Hobson, J. A., Pace-Schott, E., and Stickgold, R. (2000). Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behav. Brain Sci.* 23, 793–842.

Hodge, V. J., and Austin, J. (2004). A survey of outlier detection methodologies. *Artif. Intell. Rev.* 22, 85–126.

Hume, D. (1740). *A Treatise of Human Nature*. Available at: http://www.gutenberg.org/etext/4705

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Available at: http://eserver.org/18th/hume-enquiry.html

Ingvar, D. H. (1985). Memory of the future: an essay on the temporal organization of conscious awareness. *Hum. Neurobiol.* 4, 127–136.

Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., and Sagi, D. (1994). Dependence on rem sleep of overnight improvement of a perceptual skill. *Science* 265, 679–682.

LaBerge, S. (1990). "Lucid dreaming: psychophysiological studies of consciousness during REM sleep," in *Sleep and Cognition*, eds R. R. Bootzen, J. F. Kihlstrom, and D. L. Schacter (Washington, DC: American Psychological Association), 109–126.

Lee, A. K., and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36, 1183–1194.

Li, W., and Joshi, A. (2009). "Outlier detection in ad hoc networks using Dempster-Shafer theory," in *10th International Conference on Mobile Data Management: Systems, Services and Middleware (*Taipei: IEEE Computer Society), 112–121.

Locke, J. (1690). *An Essay Concerning Human Understanding*. Available at: http://www.ilt.columbia.edu/publications/locke_understanding.html

Mach, E. (1886). *Contributions to the Analysis of the Sensations*. New York: Open Court.

McCarthy, J., and Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence," in *Machine Intelligence*, Vol. 4, eds D. Michie and B. Meltzer (Edinburgh: Edinburgh University Press), 463–502.

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Metzinger, T. (2004). The subjectivity of subjective experience: a representationalist analysis of the first-person perspective. *Networks* 3–4, 33–64.

Metzinger, T. (2005). Précis: Being No One. *Psyche* 11. Available at: http://psyche.cs.monash.edu.au/symposia/metzinger/precis.pdf

Muzur, A., Pace-Schott, E. F., and Hobson, J. A. (2002). The prefrontal cortex in sleep. *Trends Cogn. Sci.* 6, 475–481.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 883–917.

Patcha, A., and Park, J.-M. (2007). An overview of anomaly detection techniques: existing solutions and latest technological trends. *Comput. Netw.* 51, 3448–3470.

Peirce, C. S. (1868). Some consequences of four incapacities. *J. Speculative Philos.* 2, 140–157.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* 12, 919–929.

Reps, P. (1989). *Zen Flesh, Zen Bones*. New York: Anchor Books.

Riva, G., Raspelli, S., Algeri, D., Pallavicini, F., Gorini, A., Wiederhold, B. K., and Gaggioli, A. (2010). Interreality in practice: bridging virtual and real worlds in the treatment of posttraumatic stress disorders. *Cyberpsychol. Behav. Soc. Netw.* 13, 55–65.

Sandberg, A., and Bostrom, N. (2008). *Whole Brain Emulation: A Roadmap. Future of Humanity Institute 3, Oxford University*. Available at: http://www.fhi.ox.ac.uk/reports/2008-3.pdf

Scharfstein, B. (1998). *A Comparative History of World Philosophy: From the Upanishads to Kant*. Albany, NY: SUNY Press.

Shalizi, C. R. (2005). "Symbolic dynamics, coarse-graining, and levels of description in statistical physics and cognitive science," in *Proceedings of workshop on Symbol Grounding: Dynamical Systems Approaches to Language*, Potsdam. (accessed March 14–17, 2005).

Shanahan, M. (2009). "The frame problem," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Winter 2009 ed.).

Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychol. Rev.* 91, 417–447.

Siderits, M. (2007). *Buddhism as Philosophy*. Indianapolis, IN: Hackett.

Spivey, M. J. (2006). *The Continuity of Mind*. New York: Oxford University Press.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature* 437, 1272–1278.

Stickgold, R., and Walker, M. (2004). To sleep, perchance to gain creative insight? *Trends Cogn. Sci.* 8, 191–192.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242.

Tzu, C. (1968). *The Complete Works of Chuang Tzu*. New York: Columbia University Press. Translated by B. Watson.

Wagner, U., Gais, S., Haider, H., Verleger, R., and Born, J. (2004). Sleep inspires insight. *Nature* 427, 352–355.

Waldstein, M., and Wisse, F. (eds). (1995). *The Apocryphon of John: Synopsis of Nag Hammadi Codices II,1; III,1; and IV,1 with BG 8502,2*. Leiden: Brill Academic Publishers.

Wittgenstein, L. (1972). *On Certainty*. New York: Harper Torchbooks.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 September 2010; accepted: 03 March 2011; published online: 17 March 2011.

Citation: Edelman S (2011) Regarding reality: some consequences of two incapacities. Front. Psychology ***2****:44. doi: 10.3389/fpsyg.2011.00044*

This article was submitted to Frontiers in Theoretical and Philosophical Psychology, a specialty of Frontiers in Psychology.