Mind Mach (2006) 16:185–200
DOI 10.1007/s11023-006-9036-0

Quantity of experience: brain-duplication and degrees of consciousness

Nick Bostrom

Received: 6 December 2005 / Accepted: 24 July 2006 / Published online: 26 August 2006

Abstract If a brain is duplicated so that there are two brains in identical states, are there then two numerically distinct phenomenal experiences or only one? There are two, I argue, and given computationalism, this has implications for what it is to implement a computation. I then consider what happens when a computation is implemented in a system that either uses unreliable components or possesses varying degrees of parallelism. I show that in some of these cases there can be, in a deep and intriguing sense, a fractional (non-integer) number of qualitatively identical phenomenal experiences. This, in turn, has implications for what lessons one should draw from neural replacement scenarios such as Chalmers' ''Fading Qualia'' thought experiment.

Keywords Computation · Mind · Consciousness · Implementation · Duplication · Fading qualia · Chalmers · Searle · Program · Probabilistic · Deterministic

The duplication thesis

Suppose two brains are in the same conscious state. Are there two minds, two streams of conscious experiences? Or only one?

From a physical point of view, there is no puzzle. There are two numerically distinct lumps of matter that instantiate the same patterns and undergo qualitatively identical processes. The question is how we should ascribe mental properties to this material configuration. Even if we assume that the mental supervenes on the physical, we still need to determine whether the supervenience relation is such that two qualitatively identical physical systems ground a single experience or two numerically distinct (albeit subjectively indistinguishable) experiences.

N. Bostrom (✉)
Faculty of Philosophy, University of Oxford, 10 Merton Street, Oxford OX1 4JJ, UK
e-mail: nick.bostrom@philosophy.ox.ac.uk

The issue is not about personal identity. It is a separate question whether there would be one or two persons. One might hold, for example, that one person could have two subjectively indistinguishable experiences at the same time, or that two persons could literally share what is, numerically and not just qualitatively, one experience. These issues about personal identity will not be discussed here. My concern, rather, is about ''qualia identity.'' I will start by considering the numerical identity or non-identity of the phenomenal experiences that arise when brains exist in duplicates. The bulk of the paper will then examine some intriguing cases involving partial brain-duplication and the questions of degrees and of quantity of experience that these cases force us to confront.

Consider the case where two brains are in identical physical states. The brains have, let us assume, the same number of neurons, which are connected and activated in the same way, and the brains are descriptively identical all the way down to the level of individual molecules and atoms. Suppose, furthermore, that for each of these brains there would, if the other brain did not exist, supervene a particular phenomenal experience. Given the supervenience assumption, the phenomenal experience that would supervene on one of these brains would be qualitatively identical to the experience that would supervene on the other. But if both brains exist, are there two qualitatively identical but numerically distinct experiences (one for each brain) or is there only a single experience with a redundantly duplicated supervenience base?

A hardcore physicalist might be tempted to dismiss this question as being merely terminological. However, I believe that we can give content to the question by linking it to significant ethical and epistemological issues. Given such a linkage, the answer will not be an inconsequential terminological stipulation but will reflect substantial views on these associated issues. Let *Duplication* be the thesis that there would be two numerically distinct streams of experience when a conscious brain exists in duplicate. The content of *Duplication*, I suggest, might in part be constituted by its implications for epistemology and ethics.

Consider first the ethical significance of Duplication. It could matter ethically whether Duplication is true. Suppose that somebody is contemplating whether to make a copy of a brain that is in a state of severe pain. If the quantity of painful experience would not thereby be increased, it seems that there would be no moral objection to this. By contrast, if creating the copy will lead to an additional severely painful experience, there is a strong moral reason not to do it. In such cases, it would be an extremely important practical matter whether Duplication is true or false. We could not resolve it by appealing to an arbitrary verbal convention.

This ethical implication is a reason to accept Duplication and to reject its negation (*Unification*). Unification implies that we would not bring about pain if we created copy of a brain in a painful state, or changed an existing brain into a state that is qualitatively identical to a painful state of an already existing brain. At least on hedonistic grounds, there would be no moral reason not to create the copy. Yet it is *prima facie* implausible and farfetched to maintain that the wrongness of torturing somebody would be somehow ameliorated or annulled if there happens to exist somewhere an exact copy of that person's resulting brain-state.

We can bring this point out more forcefully if we consider the very real possibility that the universe is infinite. Recent cosmological data indicate that our universe is quite likely infinite and contains an infinite number of galaxies and planets.[1] Moreover, there are many local stochastic processes, each one of which has a non-zero probability of resulting in the creation of a human brain in any particular possible state.[2] Therefore, if the universe is indeed infinite then on our current best physical theories all possible human brain-states would, with probability one, be instantiated somewhere, independently of what we do. But we should surely reject the view that it follows from this that all ethics that is concerned with the experiential consequences of our actions is void because we cannot cause pain, pleasure, or indeed any experiences at all. It is much more plausible to hold that even if the universe is the way it now seems to be, we can still influence what experiences there are. Since this would be impossible on Unification, we should accept Duplication.

The choice between Duplication and Unification also has epistemological ramifications. I have argued elsewhere that in a ''Big World,'' in which all possible human experiences are in fact made, we can explain how our experiences can give us probabilistic evidence about the world by appealing to the fact that different theories (even when coupled with the Big World hypothesis) will predict that different sorts of experiences occur with different frequencies.[3] For example, the experience of observing a measurement of the temperature of the cosmic background radiation of 2.7 K will occur vastly more frequently than the experience of observing a measurement of, say, 3.1 K, if the actual temperature of the cosmic background radiation is 2.7 K rather than 3.1 K. If the temperature is 2.7 K then (practically) all observers who make veridical observations of the relevant sort will observe a measurement reading of 2.7 K, and only relatively rare observers, who are deceived in some way, will observe a reading of 3.1 K. (The opposite would hold if the temperature were 3.1 K.) Together with a plausible methodological postulate, which can be independently supported, this explains why our observation of 2.7 K gives us probabilistic evidence for the theory that the background radiation is approximately 2.7 K.[4] Yet if Unification were true, then experiences of observing 3.1 K would be just as frequent as experiences of observing 2.7 K. This is so because according to Unification there would, in a Big World, be precisely one of each maximally specific possible experience of an observation, and there is, presumably, equally many maximally specific possible ways of experiencing an observation of 3.1 K as there are of experiencing

[1] On the standard Big Bang model, assuming the simplest topology (i.e. that space is singly connected), there are three fundamental possibilities: the universe can be open, flat, or closed. Current data suggests a flat or open universe, although the final verdict is still pending. If the universe is either open or flat, then it is spatially infinite at every point in time and the Big Bang model entails that it contains an infinite number of galaxies, stars, and planets. See e.g. (Martin, 1995).

[2] See e.g. (Hawking & Israel, 1979), p. 19.

[3] (Bostrom, 2002a, b).

[4] (Bostrom, 2002a, b).

one of 2.7 K.[5] Thus Unification would undercut a natural account of why our experiential evidence enables us to learn about the world (even if the world is a Big World). This is another reason to accept Duplication.

If we accepted Unification, we would have to find some alternative way of blocking the unacceptable ethical and epistemological implications that threaten to follow from that view. These alternatives would, I think, be less natural and plausible than the straightforward accounts that are possible if brain-duplication leads to duplication of (qualitatively identical) experiences.

To the extent that we have direct intuitions about the ontological question, they seem to go against Unification. It would, to say the least, be odd to suppose that whether one's own brain produces phenomenal experience strongly depends on the happenings in other brains that may exist in faraway galaxies that are causally disconnected from the solar system, or that may have existed millions of years ago. If Unification were true, your brain may suddenly start to produce phenomenal experience at 10:32 pm tonight, having for the first time chanced into a state that happens not to be instantiated anywhere else in spacetime. And then, at 10:34 pm, it might just as suddenly cease to produce phenomenal experience as it enters a sequence of states that has already been instantiated somewhere else. It would also be very much an open question whether you would create a painful phenomenal experience when you poke your finger with a needle. Our direct intuitions about the matter clearly support Duplication.

In light of these considerations, one might expect Duplication to be entirely uncontroversial, but this is not quite the case. The only explicit opinion on the matter that I have been able to locate in the literature, that of Zuboff, was in favor of Unification. Zuboff argued for Unification as part of an attempt to support the radical conclusion that "in all conscious life [i.e. *all* conscious life *anywhere*] there is only one person—I—whose existence depends merely on the presence of a quality that is inherent in all experience—the quality of being mine."[6] Zuboff imagines two identical brains lying at opposite ends of an operating table and being fed identical sensory inputs. A small part of one brain is swapped for the corresponding part of the other brain, and the procedure is repeated until all the brain-matter has changed places. The argument seems to be that, if you are one of these brains at the outset of the experiment, there is no point at which you move to the other side of the operating table during the piecemeal process, yet at the end of it you are on the other side; hence you must have been on both sides all along.

It is an interesting question what happens to personal identity in Zuboff's scenario.[7] For present purposes, however, we need merely note that the scenario does not work as an argument against Duplication. According to Duplication, what

[5] If, for different temperatures, there are different numbers of possible maximally specific experiences of observing that temperature, it would not help. What is needed is that the frequency of experiences of a particular sort of observation strongly correlates with the veridicality of the observation. If the only reason for there being a greater frequency of experiences of observing 2.7 K than of observing 3.1 K were that there were more possible maximally specific experiences of the former kind, then this difference in frequency could not be the ground for our inferring that the actual temperature is probably 2.7 K, since the frequency would be the same whether the temperature is 2.7 K or 3.1 K. The frequency is only evidentially relevant if it correlates with the hypotheses under consideration.

[6] (Zuboff, 1991), p. 39. See also an earlier paper by the same author (Zuboff, 1978).

[7] Similar scenarios have of course been discussed in the earlier literature; see e.g. (Parfit, 1984).

happens in this case is simply that there are two qualitatively identical but numerically distinct streams of phenomenal experience. At any given time, there is, for each of the lumps of brain-matter, a phenomenal experience that supervenes on it. Whether we regard the situation as one in which ultimately two brains have changed places or as one in which two brains that remain on opposite sides of the table have exchanged all their matter, is of no consequence as far as Duplication is concerned.

For the reasons given, I will henceforth assume that we should accept Duplication. If we duplicate a brain, we create more phenomenal experience. But exactly when in the duplication processes does the new experience emerge?

One mind becoming two

From this point onward, it will serve clarity and convenience to assume a weak form of computationalism, implying that a sufficiently powerful digital computer, running a suitable (very complex) program, would in fact have phenomenal experiences. (We do not need to assume that this would be analytically or metaphysically necessary.) Given this simplifying assumption, we can imagine the case of interest as resulting from running the same mind-program on two different computers. We can simplify matters further by supposing that the simulated minds live in and interact with identical computer-simulated virtual realities. Under these conditions, two identical mind-and-virtual-reality programs, starting from the same initial conditions, will evolve in exact lockstep. The brain-simulation and the virtual-reality simulation are both parts of a more comprehensive program, and when this program is run on two identical (deterministic) computers, they will go through an identical sequence of state-transitions.[8]

Let us then imagine a computer constructed out of copper wires, running the mind-and-virtual-reality program. We can now consider a sequence of steps in which this computer is gradually modified in such a way that we end up with two separate computers running the same program. To start with, consider one of the copper wires (Fig. 1). In step 1, electrical signals are passing along this wire as the program is executed.

In step 2, an imaginary plane is placed along the axis of the wire, and it is assumed that no current passes across this plane. In step 3, a thin sheet of insulating material is inserted in the imaginary plane. In step 4, an insection is made through the insulator and the two parts of the wire are separated to a large spatial distance.

In this idealized model, the amplitude of signals traveling along the wire plays no computational role, but we can stipulate that in step 4 the power supply is adjusted so that each of the resulting computers has the same current and voltage as the original computer had before duplication. Using a similar sequence of steps, we can duplicate individual memory registers and logic gates. At the end of the process, we have two computers separately implementing the same mind-and-environment program.

[8] This might be the technologically most practicable way for an advanced civilization to create "brain-duplicates"; see e.g. (Bostrom, 2003). Here it mainly serves to facilitate exposition. This particular way of imagining the situation, however, is not necessary for the arguments that follow. One could transpose the examples that involve computers into examples involving brains-in-vats stimulated by mad scientists.

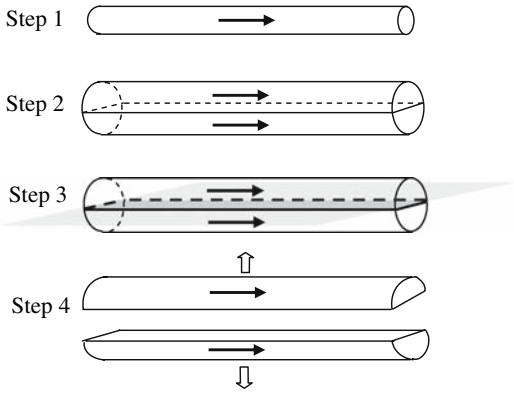

Fig. 1 Duplicating a wire

At step 1, there is one stream of phenomenal experience, and from Duplication it follows that after the completion of step 4, there are two streams of (qualitatively identical) experience. Where did the mind-duplication occur?

It is clear that inserting an *imaginary* plane has no effect on the system, so the change cannot take place in step 2. Step 4 consists in merely spatially segregating two conductors that are already insulated from one another and function as independent systems. The only plausible candidate is step 3, where the insulation was inserted.

What happened in step 3 is that a counterfactual dependence was eliminated. At step 2, although no current actually flows through the imaginary plane, it is still the case that if, say, the upper part of the wire were to be blocked then current would flow around the obstacle across the imaginary plane. Prior to step 3, the system is still acting as a single computer. The two parts of the system, separated only in our imagination, lack the capacity to perform differing and independent computations. After step 3, we have two computers that are counterfactually unlinked: if some part of one of these computers were to be provided with a different set of inputs than the corresponding part of the other computer, each part would proceed to independently compute on their respective inputs. Something like counterfactual dependencies seem to be crucially involved in individuating minds.

Klein has recently argued for an account of what it is to implement a computation that relies on dispositions rather than counterfactuals.[9] According to this account, too, it would be the case that the duplication occurs in step 3. The upper and the lower parts of the wire do not possess separate dispositions to block or conduct electric signals before the insulator is inserted. After the two halves of the system are insulated from one another, they each posses a full complement of parts, each of which has dispositions (to conduct, store, or gate electric signals) that are distinct from the dispositions of the corresponding parts of the other half of the system.

Partial duplication

Let us zoom in on step 3. There are at least two different ways in which we can subdivide this transition into a sequence of smaller steps. We can start by adding

[9] (Klein, 2004). For some other discussions of what it is to implement a computation, see also (Barnes, 1991; Chalmers, 1996; Maudlin, 1989; Wilson, 1994).

insulation to just one component of the computer and then repeat the procedure for one additional component at a time until all components are insulated from their counterparts such that we have in effect two separate computers (Case 1). Alternatively, we can start by inserting a very thin sheet of insulation through all the components of the computer and gradually increase the thickness of this sheet until it becomes a perfect insulator (Case 2). Let us examine these cases in turn.

Case 1

Consider first a small, simple part of the wiring diagram for some part of the computer (Fig. 2).

Suppose that we duplicate each of these basic components without altering the higher-level structure of the wiring (Fig. 3).

Components that have more than one input or output channel would be duplicated in a similar way. The resulting architecture is the same as before except on the scale of individual components. The redundant duplication and parallelizing of the basic computational elements does not increase the computer's capacity to execute complex programs. Since the outflows from each pair of parallel basic components converge before becoming the inflow to the next pair of components, there is no counterfactual independence between the ''upper'' and the ''lower'' half of the circuit: if, say, one of the upper loops were to be disconnected, the circuitry would still perform the same computation. Only the most basic computational fragments (such as a negation operation, or the readout of a memory bit) would be duplicated. But the computation performed by a single logic gate or memory cell is far too simple for phenomenal experience to supervene upon it. Phenomenal states supervene on computations involving large numbers of basic operations. The architecture in Fig. 3 therefore does not yet yield a duplication of phenomenal states.

We can now construct a sequence of cases in which the segments of insulated parallelism get consecutively larger.

In Fig. 4, the scope of one of the segments of the circuitry that support an independent computational process has been slightly increased so that it extends over two basic computational elements. Suppose that we continue to increase the scope until a significant part of the circuitry is provided with an independent parallel circuitry. Suppose, furthermore, that this parallelism is retained for a significant

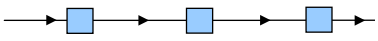

Fig. 2 The shaded boxes represent basic computational components such as a logic gate

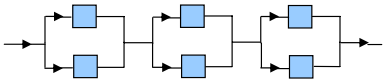

Fig. 3 Circuitry parallelized on the micro-scale

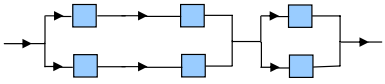

Fig. 4 Circuitry parallelized on an intermediary scale

period of time (after which the added parallel circuitry is removed). At some point, as we continue to increase its scope, the parallelism becomes comprehensive enough to form an independent supervenience base for a separate stream of phenomenal experience. In the limiting case, where the duplicated segment becomes a replica of the entire computer, we have two completely separate circuits each of which supports a complete stream of phenomenal experience.

In the limiting case, the entire stream of experience is duplicated. But even before we reach this point, the experience stream might still be duplicated in part. Partial duplication of this kind is not especially problematic. Once a sufficiently large chunk of the circuitry has been divided into two, and remains thus divided for a sufficiently long time (relative to the computer's clock speed), fragments of a separate phenomenal stream begin to emerge. No new criterion is needed to determine when this takes place. Duplication of phenomenal experience happens when a segment of added parallel circuitry that is insulated from the original circuitry is comprehensive enough that, if this segment had existed on its own, the computation being implemented by it would have generated phenomenal experience.

Case 2

In the second kind of transition, we start by inserting throughout the circuitry an extremely thin insulator—too thin to be able to block current. We then gradually increase its thickness until it perfectly insulates the two halves of the circuitry. For present purposes, we can model an imperfect insulator as a randomizing device, which, at any given time, lets a signal pass with probability P (Fig. 5). We assume that all the components of the entire system (logic gates, memory registers, and wires) have been connected with such randomization devices.

For $P = 1$, the situation is essentially that of step 2. For $P = 0$, the situation is equivalent to the one after step 3. As for the intermediary possibilities, we need to subdivide Case 2 into several different scenarios.

Case 2a

Consider first the case where randomization occurs only once or a small number of times. If the randomization occurs before the computation is begun, and the randomization devices then retain their on- or off-states throughout the computation, then the case reduces to Case 1. If sufficiently large parts of the circuitry have their upper and lower halves insulated by off-state devices, then numerically distinct phenomenal experiences supervene on these parts and thus come to exist in duplicate; otherwise there is no duplication. The same model applies if there are just a few randomization events during the program implementation. We look at the de facto insulated parts of the circuitry, for the period of time during which they are insulated, and the computations that these parts perform during that period will produce

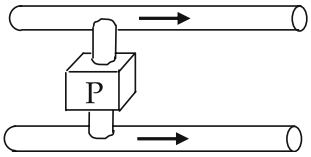

Fig. 5 Two wires imperfectly insulated from one another

a duplicate of precisely that phenomenal experience (if any) that would have supervened on that computation if it had been implemented all by itself.

Case 2b

If randomization events occur very frequently during the implementation of the program, and if P is relatively large, we would expect that many components would spend much of their time in their on-state, creating many cross-links between the upper and the lower half of the circuitry. This would prevent the two halves from acting independently and thus from supporting numerically distinct streams of phenomenal experience. However, it is possible that the randomization devices should just happen to be off most of the time. This would require a highly improbable coincidence. Yet given sufficiently many trials, we would expect such an atypical run to occur. In such a run, the independent randomization devices remain in their off-states throughout the program execution (despite a large P-value and frequent randomization events) producing a de facto insulation between the upper and the lower parts of the circuitry. Does this result in duplication?

Focusing on the de facto insulation of the two halves of the circuitry and ignoring the probabilistic dependency between them, one might be tempted to answer in the affirmative. Against this, however, one might claim that in order to implement a computation, a system must exhibit a certain degree of reliability. An ''implementation'' that completely lacked internal reliability and was merely the product of stochastic evolution would not really be an implementation at all but would be more accurately described as a sequence of chance patterns accidentally *mimicking* the implementation of a computation. If this view is correct, we face an intriguing implication: since reliability is a matter of degree, it would seem that duplication of phenomenal states would be so as well.

The sense in which duplication would be a matter of degree in this case is utterly different from that in which it would be a matter of degree in Case 1. In Case 1, the gradation is that smaller or larger parts of the original phenomenal stream are duplicated. But each such duplicated part would be completely duplicated—there would be a full, numerically distinct copy of the duplicated phenomenal experience. By contrast, the gradation arising in Case 2b does not concern which fragments of the original phenomenal stream are duplicated. Instead, it concerns the *degree* to which any given fragment is duplicated.

This notion of degree of duplication is puzzling. Are we to say that in some such intermediary cases there are e.g. 1.78 numerically distinct but qualitatively identical experiences? Or that it is indeterminate whether there is one or two? This challenge, however, is not limited to those who accept Duplication. The same sort of situation can arise where duplication is not an issue—such as in the case of an ordinary computer built from unreliable components.

Computing with unreliable components

Consider a computer made of highly unreliable components. For example, an AND-gate, which is supposed to output 1 if it gets input 1 from both of its input channels, and 0 otherwise, might follow this rule on any particular occasion with

95% probability; and in the cases where it does not follow this rule it may select an output at random, independently of its input. A computer built with such unreliable components will typically fail to implement an intended complex program. This is evident from the fact that it will, usually, not even mimic the target computation. It will instead tend to trace a deviating sequence of states. If some non-trivial calculation is performed, the computer will typically produce an incorrect result. But the interesting case is when such an unreliable computer happens by chance to exactly mimic the target computation. We need to distinguish several different ways in which this could occur.

Suppose we model each unreliable component as having a deterministic reliable core, which might on a particular occasion be disabled by a chance device.[10] When the chance device is in its on-state, and the core is thus disabled, we may say that the component has malfunctioned. The chance device has at any time a certain probability of being in its on-state. When it is in this state, it determines the output of the whole component in some stochastic fashion that is independent of the function that the component performs when the chance device is in its off-state. Two possible cases of this sort are:

Case 2b(i): None of the components malfunction on the present run.

Case 2b(ii): Some of the components malfunction but accidentally give the same output as they would have done if they had not malfunctioned.

Using Klein's dispositional framework, referred to earlier, we may judge that in Case 2b(i), the target computation was implemented, because the components had the requisite dispositional properties and these dispositions were appropriately activated during the implementation. In Case 2b(ii), by contrast, the requisite dispositions were disabled (having been switched off by the randomizing devices). Since the relevant dispositions were not activated, the target computation was not implemented, although a chance process occurred that mimicked the implementation of the computation.

Yet we can describe a third kind of variation of the 2b-case, in which the components do *not* possess deterministic dispositions that may or may not be disabled by an extraneous randomizing device, but where instead the dispositional properties themselves are probabilistic. That is, each component has a simple *probabilistic disposition* to produce a particular output given a particular input. For example, a probabilistic ''AND-gate'' might intrinsically have the following probabilistic dispositional property:

Input:	Output:
(0,0)	0 with probability .99; 1 with probability .01
(0,1)	0 with probability .99; 1 with probability .01
(1,0)	0 with probability .99; 1 with probability .01
(1,1)	0 with probability .01; 1 with probability .99

Other kinds of logic gates, memory cells, and even the wiring can have analogous probabilistic dispositions.

[10] I do not claim that all kinds of unreliability are best modeled in this way, but considering only ''unreliability'' that fits this model will serve the purposes of this paper.

Suppose an indeterministic computer is composed of such indeterministic components, ''implementing'' some program such that the implementation of that program on a computer with ordinary deterministic components would generate phenomenal experience. Consider a particular run on such an indeterministic computer where each component happened to respond to its input in the same way as it would have done if it had been of the ordinary non-probabilistic kind. We can now construct a continuum of cases, starting from the ordinary deterministic computer, progressing by gradually increasing the degree of indeterminacy of its components, until we reach the limiting case where its components are pure randomizing devices whose outputs are uncorrelated with their inputs. At the starting point of this continuum, the program is definitely implemented. Given what we know about the nature of the physical components used in real computers (which of course do implement programs), cases very close to the beginning of this continuum also implement the program. In the end case of the continuum, the program is not implemented. Let us consider a case located between these extremes.

> *Case 2b(iii)*: With a bit of luck, a computer with components whose computational dispositions are to some moderate extent indeterministic manages to complete a ''run'' of a program whose implementation on a deterministic computer would create phenomenal experience

By assumption, the system at least mimics the implementation of the program, but does it actually implement it? In this kind of intermediary case, we cannot explain the characteristic matter of degree by saying that larger or smaller fragments of phenomenal experience are generated depending on the degree to which the components are deterministic. Here, all the possible fragments of phenomenal experience are on a par: either all are generated or none. Furthermore, it is implausible to suppose that there is a sharp cut-off point, some specific degree of indeterminacy such that if the components are made infinitesimally more indeterministic, then a radical change in the associated phenomenology occurs—from there being all the phenomenology there would be in the deterministic case to there being no phenomenology at all. We are forced to recognize, it seems, that phenomenal experiences admit of degrees in a more fundamental sense: degrees that are not manifest in the descriptive qualitative character of the experience. The degree of experience, in this sense, would vary smoothly with the degree of determinacy in the components that implement the program upon which the experience supervenes.

It is not clear that an adequate term for this dimension of phenomenal experience exists in ordinary English. One might speak of it as the ''intensity'' of experience. Yet this is potentially misleading because intensity in this special sense would not be reflected in the quality of the experience in the way that, say, an intense pain is qualitatively different from a less intense pain. More important than terminology, however, is the question of how we should understand the nature of such variation in the ''intensity'' of experience. This problem does not seem to have been noticed before.

One approach would be to claim that the variation is one of *quantity* of experience. We would then say that there is a greater (numerical) quantity of experience of a given qualitative sort in the fully deterministic case than in cases of type 2b(iii). On this approach, depending on the degree of determinacy of the computational dispositions that are activated during the execution of the program, there would be a different *fractional* number of minds supervening on the execution of the program.

A given system might, for example, engender 0.85 numerically distinct but qualitatively identical minds or streams of phenomenal experience. In the limiting case of complete determinacy, it would engender one mind; and in the case of complete indeterminacy, no mind. A numerical quantity greater than 1 could be obtained by having multiple computers implement the same program. We would also get a fractional number of minds in cases where a computer's components are partially separated in a certain way. If each (partially) separated pair of components acts as a meta-component that has an indeterministic disposition to respond to the inputs to its component parts either separately or by pooling them, then we could again have an intermediary case analogous to 2b(iii).

''How can this be the case?'' one might ask. ''Either the experience occurs or it doesn't. How can there be a question of quantity, other than all or nothing?'' But the underlying reality, the system upon which the experience supervenes, does not change abruptly from a condition of implementing the relevant program to a condition of not implementing it. Instead, the supervenience base changes gradually from one condition to the other. It would be arbitrary and implausible to suppose that the phenomenology did not follow a similarly gradual trajectory. Moreover, given Duplication, it would in any case be wrong to suppose that the existence of a phenomenal experience is an all-or-nothing matter. Even apart from the possibility of fractional numbers of minds, there would still be the question of whether a particular type of mind exists in one, two, or more copies, implemented on physically independent systems.[11]

An alternative approach would be to claim that in the relevant sort of intermediary case, it is indeterminate whether a computation is implemented or not (or, in the case of the partially separated computer, whether it is implemented once or twice). But we could then regard the idea of fractional numbers of minds as a specification of our original concept of a mind, a specification that enables us to express determinate truths about some matters which the original concept was too blunt to capture. At the present time, the use of such a specification would be purely theoretical, but depending on how future technology develops, systems might one day be built where it becomes a matter of ethical or epistemological significance to determine the fractional number of minds that they implement.

How does this relate to the Fading Qualia thought experiment and other scenarios?

We should distinguish the possibility illustrated here from that envisioned by Chalmers in his well-known ''Fading Qualia'' thought experiment. Chalmers introduced this thought experiment as part of an attempt to argue for the principle of *organizational invariance*, which asserts that experience is invariant across systems with the same fine-grained functional organization. To this end, he considers a neural replacement scenario, in which the neurons of an organic human brain are

[11] These statements are consistent with an epistemicist account of vagueness (see e.g. Williamson, 1994). It might be true of any system either that it has associated phenomenal experience or that it does not. The point here is that systems that have associated phenomenal experience can have it in varying amounts or degrees of ''intensity,'' even when the duration and the qualitative character of the experience does not vary. Moreover, this particular quantity of degree does not come only in integer increments. *Formally*, this is no more mysterious than the fact that sticks come in different lengths and that length is a continuous variable (at least on the macroscopic scale).

replaced, one by one, by silicon processors with the same input/output function as their neuronal counterparts.[12] He tries to show that although it is logically possible that the resulting silicon brain would lack phenomenal experience, we should think it highly unlikely that it would in fact do so. After having argued that it would be implausible to suppose that there is a sharp cut-off at some point during the gradual replacement, at which the hybrid brain goes from having normal phenomenology to having no phenomenology, he then considers the alternative: that the qualia would fade gradually as more and more neurons are replaced. Suppose that Joe is at some intermediary stage in the replacement process, and that he claims to have a vivid experience of bright red and yellow. Chalmers considers what Joe's allegedly fading qualia might be like:

> By hypothesis... Joe is not having bright red and yellow experiences at all. Instead, perhaps he is experiencing tepid pink and murky brown. Perhaps he is having the faintest of red and yellow experiences. Perhaps his experiences have darkened almost to black. There are various conceivable ways in which red experiences might gradually transmute to no experience, and probably more ways that we cannot conceive.[13]

Chalmers then goes on to argue that it would be empirically implausible that Joe, a normal rational subject who is paying attention to what is going on, would fail to notice these dramatic changes in his qualia. But Chalmers thinks that Joe cannot notice any changes, since the functional organization of his brain, by assumption, remains unchanged. From this he concludes that Joe's qualia do not fade and that the silicon brain would have the same qualia as the original.

Searle discusses a similar thought experiment and proposes a different account of what happens during the replacement:

> ... as the silicon is progressively implanted into your dwindling brain, you find that the area of your conscious experience is shrinking, but that this shows no effect on your external behavior. You find, to your total amazement, that you are indeed losing control of your external behavior. You find, for example, that when the doctors test your vision, you hear them say, ''We are holding up a red object in front of you; please tell us what you see.'' You want to cry out, ''I can't see anything. I'm going totally blind.'' But you hear your voice saying in a way that is completely out of your control, ''I see a red object in front of me.''[14]

On both Chalmers' and Searle's accounts, the *quality* of Joe's experience changes as his neurons are replaced. Chalmers suggests that his bright red experience might turn into tepid pink, or darken, or become less vivid. Searle proposes even more dramatic changes to Joe's experience, including feelings of powerlessness and frustration as he discovers that he is going blind. By contrast, in the kind of intermediary case that I have sought to demonstrate, the quality of experience remains unchanged. Red does not become tepid pink, nor do the changes that are taking place trigger qualitatively new experiences such as of bewilderment or frustration. Nothing changes, except the *quantity* of experience. The difference in what experience there is, is of the same kind as the difference between a case where only one brain is

[12] Similar scenarios had been discussed earlier, e.g. (Cuda, 1985; Pylyshyn, 1980; Savitt, 1980).

[13] (Chalmers, 1995), p. 256.

[14] (Searle, 1992), pp. 66f.

having an experience and one in which two identical brains are having that same experience. I have argued that this kind of difference can come not only in integer increments but in continuous degrees, such that there can be a fractional quantity of particular qualitatively specified experience.

This possibility undercuts Chalmers' argument for the principle of organizational invariance by offering a more plausible account of how Joe's qualia could gradually fade when he undergoes the neural replacement process. If the fading took place in this way, Joe would not be strangely failing to notice any qualitative changes in his experiences, because there would be no such changes. Of course, this point does not show that the principle of organizational invariance is false; it merely undermines one argument in its favor. The principle might well be plausible other grounds.

It might be tempting to think that the possibility of fractional minds could be demonstrated by considering more mundane examples featuring gradations of consciousness, such as infant development, consciousness in animals at different levels of cognitive sophistication, the humanoid ancestors of homo sapiens, or the gradual loss of consciousness that occurs when we drift into dreamless sleep or anesthesia.[15] However, these cases are complicated. They certainly involve changes in the descriptive quality of experience, and these qualitative changes, too, can form a continuum. Starting with a fully awake normal human stream of consciousness, we could reach a state of unconsciousness by various possible sequences of small steps, in which our awareness gradually becomes more fragmented, the fragments become briefer, scarcer, and more diffuse until the stream completely dries out. These cases *might* also involve a gradual change in the quantity of particular qualitative experiences, in the sense relevant here; but it is not obvious that they do so. In the thought experiments presented in preceding sections, we carefully controlled for the confounding variable of qualitative change in order to focus on the fundamental question of quantitative change.

Conclusions

Suppose the implementation of a certain program gives rise to a phenomenal experience. I began by considering the question of whether two implementations of this program give rise to two qualitatively identical but numerically distinct experiences. The Duplication thesis answers this question in the affirmative. That Duplication is a substantial claim can be seen from the fact that it has important ethical and epistemological implications. I defended Duplication by arguing that its implications in these areas are much more plausible than the implications of its negation, Unification. Moreover, our direct ontological intuitions about the matter seem to support Duplication.

But how does duplication occur? I distinguished several different ways in which an implementation process could be gradually segregated into two independent processes, and argued that the phenomenological result at the intermediary stages depends on how the separation takes place. In some cases, only fragments of the supervening phenomenal experience will be duplicated, but each such fragment is at

[15] Split-brain patients might also come to mind as a candidate for mind-duplication. But the reason why we even consider the possibility of there being two minds in these cases is that it seems as if the two hemispheres have qualitatively *different* conscious experiences.

once fully duplicated. The matter of degree here resides in the number and size of the fragments existing in duplicate. In other cases, however, the entire supervening phenomenal experience is simultaneously duplicated, and the matter of degree resides in the total quantity of qualitatively identical experience. Intriguingly, this quantity can be a fractional number. This casts some new light on what it is to implement a computation.

The idea of a fractional quantity of qualia may be puzzling. By considering the case of a single computer built with unreliable elements, it can be shown that this possibility of fractional qualia would have to be confronted even if Duplication were rejected. Hence it is not a reason to reject Duplication.

One may still wonder about the notion of purely quantitative variation of qualia. A quale is supposed to be a subjective phenomenal appearance—what something feels like, ''from the inside.'' But in the central cases I described, the subject is not supposed to register any qualitative changes in her experience. What is it then that changes when the quantity of, say, a particular pain quale decreases from 1 unit to 0.3 units? If the pain feels just the same, in what sense is there less pain after the change? Here is my answer: There is less pain in precisely the same sense as there is less pain if now only *one* subject experiences a particular pain quale while before *two* subjects each experienced just such a pain quale. The nature of fractional quantitative change is the same as the nature of quantitative change when it occurs in more familiar integer increments. (If we wish to speak of ''subjects of experience,'' perhaps we should also say that such subjects can likewise come in fractional degrees, not only in integer increments.)

Finally, I considered the relation between the cases described in this paper and the Fading Qualia thought experiment, which Chalmers used as part of an argument for the principle of organizational invariance. The possibility of qualia fading quantitatively without any change in its descriptive, qualitative character undermines Chalmers' argument.[16]

References

Barnes, E. (1991). The causal history of computational activity: Maudlin and Olympia. *Journal of Philosophy*, *88*(6), 304–316

Bostrom, N. (2002a). *Anthropic bias: Observation selection effects in science and philosophy*. New York: Routledge

Bostrom, N. (2002b). Self-locating belief in big worlds: Cosmology's missing link to observation. *Journal of Philosophy*, *99*(12), 607–623

Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly*, *53*(211), 243–255

Chalmers, D. (1995). Absent qualia, fading qualia, dancing qualia. In: Metzinger, T. (Ed.), *Conscious experience*. Paderborn: Exetes Schoningh (in association with) Imprint Academic

Chalmers, D. (1996). Does a rock implement every finite-state automaton? *Synthese*, *108*, 309–333

Cuda, T. (1985). Against neural chauvinism. *Philosophical Studies*, *48*, 111–127

Hawking, S. W., & Israel W. (Eds.) (1979). *General relativity: An Einstein centenary survey*. Cambridge: Cambridge University Press

Klein, C. (2004). Maudlin on computation. *Working paper*

Martin, J. L. (1995). *General relativity*. London: Prentice Hall

[16] For their comments, I'm grateful to Heather Bradshaw, David Chalmers, Wei Dai, Adam Elga, Hal Finney, Guy Kahane, Colin Klein, Toby Ord, and Oliver Pooley.

Maudlin, T. (1989). Computation and consciousness. *Journal of Philosophy*, *86*(8), 407–432
Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press
Pylyshyn, Z. (1980). The 'causal power' of machines. *Behavioral and Brain Sciences*, *3*, 417–457
Savitt, S. (1980). Searle's demon and the brain simulator reply. *Behavioral and Brain Sciences*, *5*, 342–343
Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, Mass: MIT Press
Williamson, T. (1994). *Vagueness*. London: Routledge
Wilson, R. A. (1994). Wide computationalism. *Mind*, *103*(411), 351–372
Zuboff, A. (1978). Moment universals and personal identity. *Proceedings of the Aristotelian Society*, *52*, 141–155
Zuboff, A. (1991). One self: The logic of experience. *Inquiry*, *33*, 39–68