International Journal of Machine Consciousness
Vol. 4, No. 1 (2012) 141−158

DOI: 10.1142/S1793843012400070

World Scientific
www.worldscientific.com

# THE TERASEM MIND UPLOADING EXPERIMENT

MARTINE ROTHBLATT

*Cyber-Biology Group, Terasem Movement, Inc.*
*201 Oak Street, Melbourne Beach, FL 32951, USA*
*Martine4@gmail.com*

The Terasem Mind Uploading Experiment is a multi-decade test of the comparability of single person actual human consciousness as assessed by expert psychological review of their digitized interactions with same person purported human consciousness as assessed by expert psychological interviews of personality software that draws upon a database comprised of the original actual person's digitized interactions. The experiment is based upon a hypothesis that the paper analyzes for its conformance with scientific testability in accordance with the criteria set forth by Karl Popper. Strengths and weaknesses of both the hypothesis and the experiment are assessed in terms of other tests of digital consciousness, scientific rigor and good clinical practices. Recommendations for improvement include stronger parametrization of endpoint assessment and better attention to compliance with informed consent in the event there is emergence of software-based consciousness.

*Keywords*: Mind; mindfile; mindware; consciousness; artificial intelligence; mind uploading; Turing Test; personality capture; Popper; psychology; cyber-psychology; bioethics; medical ethics.

## 1. Techno-Immortality Feasibility

This chapter is motivated by my desire to assess the feasibility of a kind of techno-immortality. Specifically, is it possible that software written a few decades from now, and paired with a database of video interviews of and associated information about a predecessor person, will be able to faithfully mimic the workings of this predecessor's mind? An empirical answer can be obtained by tasking psychologists to determine whether they believe the new software-based mind appears to have a consciousness that is equivalent to that of its predecessor brain-based person. I have set up an experiment to see whether or not this is so. If it is, I believe the software-based mind is a techno-immortalized continuation of the predecessor's identity. While the software-based mind will realize it is not the original brain-based mind, just as each human adult realizes they are not their teenage mind, or even the precise mind of the previous day, this fact of personal consciousness flux does not undermine the continuity of unique identity.

"Mind" means a dynamic, idiosyncratic and self-inclusive representer and analyzer of, and volitive agent in, an environment, including the capability, when functioning normally, of communicating via linked systems in one or more characteristic patterns of impressions of its analyzed representations and volitions [Lloyd, 1989]. A good barometer of the magnificent complexity of even a simple mind is that it takes the preceding 39 words to non-tautologically define it. A few more or less alternative words could define "mind" perhaps as well, but not many fewer words, or else key concepts would be lost that are important to humanity's semantic association for "mind" [Gould, 1999].

Throughout history there has never been a mind without a brain. It is the brain that has billions of neurons and trillions of synapses to provide the patterns of electrochemical connectivity that, writ with extraordinary complexity, give rise to environmental representations, analyses and choices that are the hallmarks of a mind. The brain is to the mind as objects that are counted are to numbers. Some physical substrates, such as brains and abacus beads, necessarily entail non-physical phenomena, such as minds and math.

Now it is hypothesized that computer software can mimic the achievements of billions of neurons and trillions of synapses [Minsky, 2006]. Specifically, it is contended that computer software can provide patterns of representation, analysis and volition of such nuanced complexity as to give rise to human thought and feeling [Kurzweil, 1999]. Furthermore, it is posited that such software-based human thought and feeling could closely approximate the consciousness that is experienced by a biological person whose own sampled experiences provided the database for such software [Rothblatt, 2011]. Such contentions form the basis of the Terasem Mind Uploading Hypothesis and the long-term public experiment of its validity. The purpose of this chapter is to describe and critique the Terasem Mind Uploading Experiment. The motivation for the experiment is to see whether a kind of techno-immortality is possible, after a few additional decades of software development, through continuing a person's identity in a software-based mind.

## 2. The Terasem Mind Uploading Hypothesis

The Terasem Null Hypothesis is as follows:

> *Within a span of several decades worth of information technology growth at the Moore's Law or Kurzweil Rate, databases populated via the open public participation websites LifeNaut.com and/or CyBeRev.org with digital samples of participants' mannerisms, personality, recollections, feelings, beliefs, attitudes and values (hereinafter referred to as "mindfiles"), and used as reference databases by software designed to replicate and customize the functional characteristics of human minds (hereinafter referred to as "mindware"), will not give rise to software-based minds that are recognized by a panel of psychologists as equivalent to the matching brain-based minds of the participants, as determined by*

> *interviews with the software-based minds over a period of a year, and comparisons of impressions from such interviews with the mannerisms, personality, recollections, feelings, beliefs, attitudes and values reflected in the matched mindfiles of the original brain-based participants.*

In brief, the null hypothesis states that a panel of psychologists will not believe a software-based mind is a continuation of, or an analog of, a brain-based mind. To disprove the null hypothesis a panel of psychologists who observed both digital samples of experimental participants' mindfiles, and output from the purported consciousness of minds activated by mindware from those mindfiles, would conclude that, preponderantly, the purported consciousness did in fact appear to be equivalent to the consciousness of the subjects who created the mindfiles.

This hypothesis is a unique variant of the Turing Test [Turing, 1950]. It might be called a Self-Turing Test, although, unlike the Turing Test, the assessors of consciousness are not blinded in their interactions with the actual and potential sources of consciousness. While this does permit bias as to purported consciousness substrate to affect the assessors' decisions, it also permits a vastly more robust, and realistic, assessment of purported consciousness. Unlike with the Turing Test, the goal is not to trick the assessor into believing the artificial consciousness is indistinguishable from the biological consciousness. Instead, the goal for the Terasem variant of the Turing Test is to trust the expert psychological panel in their professional judgment as to whether the purported artificial consciousness is as good as, or equivalent to, or a continuation or analog of, the original biological consciousness. Because this experimental design is unblinded, unlike the blinded test proposal of Turing, it is especially important that this experiment have the additional rigor of a null hypothesis, unlike the positively stated claim of Turing's paper. However, what is common to both the Terasem and the Turing experimental designs is the replacement of the subjective concepts of "thinking" and "consciousness" with the empirical measure of whether others believe the interactions with the software-based mind to be "thinking" or "consciousness".

An important point to illuminate is why there should be an expectation that mindfiles will capture enough about an original person for the experimental design to be interesting. For example, it is argued that one could never record enough sensory experiences and actions to produce brain emulation [Gemmell *et al.*, 2004]. However, this is the right answer to the wrong question. The question is not whether one can replicate the 10 trillion synaptic strengths and yet greater number of connectivities of the human brain in a software matrix. This would be like trying to replicate human flight by building an airplane out of a trillion micro-widgets in the exact same configuration as found in an eagle or a sparrow. Instead, the goal here is to replicate the functionality of a specific human consciousness in software. There is no more reason to assume *a priori* that the only way to do this is to replicate a human brain than there is to assume *a priori* that the only way to fly is to replicate a bird. Instead, we reason that software emulation of a human mind via analysis of a set of mindfiles is

achievable because there are but a limited number of personality traits [Costa and McCrae, 1990], a finite set of universal facial expressions and emotions [Brown, 1991], a diminished repertoire of remembered thoughts from day-to-day [Ebbinghaus, 1885], and not more than a few gigabytes of remembered information [Landauer, 1986].

In general, dozens ($n$) of mannerism, personality and feeling types ($m$) yield many thousands of unique human combinations *via* $(n!)/(m!*(n-m)!)$. Once you add to these thousands of personality and worldview templates differential recollections, beliefs, attitudes and values (the gigabytes of remembered information) there are many billions of unique possible combinations of human psyches, one of which will be a best-match for digitally stored mindfiles on a predecessor biological person. Mindware best fits one of the "$m$" compound mannerisms, personality and feeling types to that analyzed from stored mindfiles, and then populates it with the recollections, beliefs, attitudes and values evident from the stored mindfiles. The Terasem Mind Uploading Experiment tests whether these combinations and correlations can be accomplished with software that need be nowhere near the complexity of synaptic connectivity of the human brain, and yet still appears as true to the original person as to persuade an expert panel of psychologists that the same personal identity is present.

Several parameters of the Terasem Null Hypothesis are open to experimental design. The specific multi-decade timeframe of information technology bi-annual (and decreasing) doubling should be set.[1] Moore's Law sets the doubling period for information technology on a chip as every two years since the mid-1950s, whereas Kurzweil sets the rate of doubling for information processing capability as a function of historical epoch, of which he observes our present integrated circuit age is the fifth such epoch [Kurzweil, 2005].

Hardware is only part of the problem, but growth rates in software efficiency are comparably impressive [Vinge, 2008].[2] To be clear, though, it is not expected that all of the code for a person's patterns of mindedness would be line-by-line coded. Instead, the mindware, or mind operating system, will be learning software [Bock, 1993]. It will designed to seek out and adopt, or "auto-tune" to, idiosyncratic data and patterns in each participant's mindfile in accordance with fundamental pre-programmed universal patterns of human thought and socio-economically specific cultural knowledge. Iterative internal quality assurance cycles will result in revisions until a stopping point is reached based on matches to pre-set parameters between the learned software-mind and all material elements of its biological precursor as reflected in the mindfile. Self-awareness functionality will be activated once cyber-consciousness health and safety checks are complete.

[1] Caution in setting a specific date can be learned from Hans Moravec's mistaken date of 2010 for software minds, given in 1988. *Wall Street Journal*, December 7, 1990, pp. 1, 9 (Carroll, P. "Good News: You Can Live Forever; Bad News: No Sex — It's the Far Edge of Robotics As One Scientist Asks What's a Body to Do?").

[2] Vernor Vinge estimates that software will keep up with Moore's Law progress on hardware such that by 2030 human equivalent intelligence in information technology will be achieved.

The number of psychologists on the endpoint determination panel, the duration of their interviews and observations of software-based minds and of mindfiles, and the statistical definition of "preponderance" for establishment of equivalence should all be parametrized. The number of mindfiles and purported conscious minds to include in the dataset should be defined, as well as how they should be selected (e.g., largest in terms of data size, file number, or use of different mindfile tools). Also, it is important to define the specific parameters of what kinds of outputs from purported conscious software-based minds should be compared to stored mindfiles, and how broad or enduring such a comparison should be. While a specific experimental design has not yet solidified these and other important open parameters, an experimental test of the hypothesis has begun and it has been useful to gaining experience with testing the hypothesis to commence the collection of mindfile data from the general public. Indeed, the current status of the Terasem Mind Uploading Experiment is primarily one of validating the utility of the mindfile database tools, and secondarily commencing the long-term population of the mindfile databases. Should it ultimately be determined that any particular database tool is of problematic value then such information will simply not be made available to mindware used to test the hypothesis.

Collecting mindfile data in advance of a full experimental design parametrization of the Terasem Mind Uploading Hypothesis does not bias the ultimate study. What is occurring now is analogous to fine-tuning a laboratory apparatus with test-runs so as to achieve reliable baseline values. Experimental bias would occur if the mindfile data being collected were altered based upon the endpoint assessments of the ultimate panel of psychologists. That is not occurring, as that endpoint is decades into the future — more than adequate time within which to fully parametrize the experiment.

## 3. Sexual Identity Research as a Trailblazer for Mind Uploading Research

As an example of this type of "state-of-mind" research, consider the question of whether it can be appropriately determined whether a purported transsexual is in fact someone suffering from a gender identity disorder appropriately treated with sex reassignment surgery. If not they may be instead suffering from another type of mental or endocrine disorder, or may simply be evidencing cultural diversity, and in these cases should not be made the subject of irreversible surgery [Gooren, 1995]. Resolving this question is usually tantamount to discovering the true state of a person's gender consciousness [Doctor, 1990]. In many regards it is not a very different quest from trying to discover the true state of a purported consciousness revived from a mindfile. In both cases one must judge if the consciousness being presented is a fake or is authentic. Does the consciousness being presented represent an authentic analog (albeit with different gender or substrate), or does it represent discontinuity (such as a different personality that has taken root in a new gender or substrate)?

After a few decades of trial and error the transsexual health field has settled upon what is called the "real life test" [Reid, 1995]. In this test a purported transsexual must meet with two psychologists regularly over a period of a year. Only if both psychologists affirm in writing, after a year's worth of therapy sessions, that the individual truly believes they are mentally of a gender associated with the other sexual phenotype, and that other confounding dysfunctions such as multiple personality disorder are not present, will such an individual be able to legally obtain a sex reassignment surgery [Rothblatt, 1995]. The lesson for consciousness purportedly revived from a mindfile is that it may well take up to a year for a panel of psychologists to gain adequate insight from which to reach a conclusion on the experiment's endpoint of consciousness identity.

## 4. The Terasem Mind Uploading Experimental Design

The Terasem mind-uploading experimental design consists of two independent sources of sampled human experiences — hereafter referred to as "mindfiles" — and two independent efforts at development of a mind operating system — hereafter referred to as "mindware" — an unlimited ability of unselected members of the general public to participate in the creation of mindfiles, and a multi-decade time horizon. The two independent sources of mindfiles are called CyBeRev.org and Lifenaut.com. Each are free, not-for-profit websites at which unselected members of the general public have a practically unlimited ability to create mindfiles that are used as databases for developing iterations of mindware. The mindfiles that may be created using a wide variety of autobiographical, human experience sampling tools are indicated in Table 1.

No assumption is made as to whether some of these tools are more insightful than others. Members of the public make use of self-determined and thus randomly varying amounts of the tools. Some participants visit both websites, while most restrict themselves to either Lifenaut.com or CyBeRev.org. The video, photo, audio and text upload tools all also have the ability for the participant to provide keywords, free-form descriptions, importance weightings from 1 to 10 and categorization as primarily a reflection of a mannerism, personality, recollection, feeling, belief, attitude or value. The item listed in Table 1 as "Bainbridge Inventories" are over 100,000 personality capture statements that require bi-modal responses, on dual scales of $(-5)$ through 0 to $(+5)$, as to agreement with the statement and with the importance of the statement to the responder [Bainbridge, 2003]. In the mobile app version of this mindfile option, one's response to each statement is simply plotted on an $x$-$y$ plane weighted the same as on the website. Bainbridge inventory-based personality capture alone may be an adequate basis upon which to test the Terasem Hypothesis [Bainbridge, 2006].

The two independent efforts at mindware development are performed by the staffs of Terasem Movement Inc. (TMI) and Terasem Movement Foundation (TMF), respectively. Each staff is simply given the guidance that they are responsible for

Table 1. Mindfile tools for personality capture, storage of digital reflections of consciousness at CyBeRev.org and Lifenaut.com public websites.

| Mindfile tool | Cyberev | Lifenaut |
|---|---|---|
| Free-Form Video | X | X |
| Free-Form Photos | X | X |
| Free-Form Audio | X | X |
| Free-Form Text Uploads | X | X |
| Structured Religious Views | X | |
| Structured Morality Views | X | |
| Structured Political Views | X | |
| Reactions to Stock Photos | X | |
| Biometrics | X | X |
| Favorite Rankings of Music Literature, Films, Things Places, People & Food | X | |
| Hated Rankings of Music Literature, Films, Things Places, People & Food | X | |
| Free-Form Journal Entries | X | X |
| Historical Contextualizing | X | |
| Bainbridge Inventories | X | |
| Chatbot Training | X | X |
| Avatar Training | | X |
| Personality Test | | X |
| Gratitude Test | | X |
| Free-Form Favoriting | | X |
| Life Geo-Tagging | | X |
| Social Network Maps | | X |
| Temporal Sequencing | | X |
| Website Linking | | X |
| Facebook Post Importing | X | X |

using their budget to obtain a mind operating system that will be able to replicate in software the human consciousness that gave rise to the mindfiles in their databases. They are told to expect that this will be a 20−30-year project. Each has independently chosen to build their mindware efforts upon the foundation of existing chatbot software. One group (TMF) uses a version of Jabberwacky[3] upgraded with natural language processing, while the other group (TMI) uses a homegrown analog of A.L.I.C.E. that relies on Princeton's WordNet for conceptual linking and a maximum entropy model toolkit for language decomposition.[4] Both chatbot systems incorporate database look-up into the mindfile of the person with whom a conversation is being attempted, as well as training algorithms that enable the participant to structure the conversational responses. Both teams expect more sophisticated artificial consciousness engines to incrementally replace these early artificial conversational

[3] Jabberwacky is a development of the ICogno company in the United Kingdom, http://www.jabberwacky.com/, retrieved September 29, 2011.

[4] http://wordnet.princeton.edu/, retrieved September 29, 2011.

entities over the next two decades, consistent with the timeframes projected by Kurzweil [2005]. Efforts at natural language processing (NLP) by Apache Software Foundation (OpenNLP), MIT Media Lab (ConceptNet) and Stanford University (CoreNLP) are early steps in this direction.

Participants in either system must complete an informed consent form. For example, at the LifeNaut site the participant must agree, among other things, that they are in a "Research Project", the purpose of which:

> *... is to test the Terasem Hypotheses that* (1) *a conscious analog of a person may be created by combining sufficiently detailed data about the person* ("*mindfile*") *organized at this website with future consciousness software* ("*mindware*"), *and* (2) *that such conscious analog can be downloaded into a biological or nanotechnological body to provide life experiences comparable to those of typically birthed humans. If even the first part of the two Terasem Hypotheses is shown to be true, the conscious analogs will be independent persons with rights and obligations dependent upon their capabilities. I hereby understand and accept that by creating a mindfile at this website I am participating in a test of the Terasem Hypotheses and that a result of this participation may be the creation of one or more new legally independent persons whose memory consists of the mindfiles I create, and whose consciousness arises from those mindfiles.*

Similarly, an extract from the CyBeRev informed consent provides that:

> *... my beingness stored herein at CyBeRev or any subsequent location be considered* "ex vivo *consciousness storage*", *which is the creation of digital files that digitally sample a person's mannerisms, personality, feelings, recollections, beliefs, attitudes and values. My purpose in creating this* ex vivo *consciousness storage is to preserve my individual consciousness so that it remains viable for possible uploading with consciousness software into a cellular regenerated or bionanotechnological body by future medicine and technology.*

As of August 2011, there are approximately 500 research participants on each of the two websites.

## 5. Popperian Assessment of the Terasem Experimental Design

Following Popper [1935], a theory is scientific if and only if it is falsifiable. The theory behind the Terasem experimental design is that software operating on a database of digital samples of a person's consciousness can regenerate an equivalent consciousness. The theory can be stated as a null hypothesis by stating that a panel of psychologists will not agree that a software-based digital representation of the consciousness of a biological person is functionally equivalent to that person's actual consciousness, as represented by such person's self-generated digital reflections of

consciousness stored at public websites. This is a falsifiable hypothesis, and hence a scientific basis upon which to create an experimental test.

It is possible that some of the software-based representations will persuade the panel of psychologists that their apparent consciousness is equivalent to that which the psychologists perceive in the biological original's mindfiles. If so, the hypothesis will have been falsified, and a scientific fact learned, to wit: psychologists can believe that software-based consciousness is equivalent to flesh consciousness. This does not mean that it has been proven as true that the software consciousness is equivalent to the biological consciousness. Such a truth is scientifically unattainable, for as Popper [1972] taught, "we cannot ever have sufficiently good arguments in the empirical sciences for claiming that we have actually reached the truth, we can have strong and reasonably good arguments for claiming that we may have made progress towards the truth. . ." What it does mean is that psychologists selected in a defined manner to judge consciousness based upon digital images generated by both biological brains and specific mindware-mindfile software combinations consider the two manifestations of consciousness to be equivalent.

Now, it is also quite possible that at the end of the Terasem Mind Uploading Experiment the null hypothesis will have been proven. In other words, panels of psychologists, in accordance with pre-set parametrization of all experimental independent variables, will not achieve the requisite statistical concordance in a judgment that the software-based minds are of equivalent consciousness to the brain-based original minds. As predicted by Popper, this will drive researchers to advance science further by creating larger mindfile databases and more advanced iterations of mindware, which can then be made subject to falsification in further experiments.

In this way science progresses not toward the truth of the software consciousness; instead, the progress is toward the ability of software to solve ever more interesting problems as exemplified by recreating the impressions of consciousness that are the hallmarks of human brains. The Nobel Prize winning biologist and serial author of neuroscience books Edelman [2004] opines that the human brain is the most complex object in the universe, with more neural interconnections than there are stars in thousands of galaxies. He does not believe the mind that arises from it can be replicated in software [Edelman, 2004]. Herein lies one of the most interesting problems science can address. Is it possible for software to replicate the functionality of consciousness that we associate with a brain-based mind? The Terasem Mind Uploading Experiment addresses this fascinating question in a concrete manner by focusing on comparability to the evidence of consciousness left by brain-based minds at mindfile-capturing websites.

Ultimately it is possible that digital representations of consciousness are produced via mindware that generally, repeatedly, and reproducibly, persuades panels of psychologists of their equivalence to human consciousness. This does not mean that the created cyber-consciousness is "truly" equivalent to human consciousness, because science cannot prove truth but only demonstrate falsehood. Instead, ever

more sophisticated versions of the Terasem experiment will be developed in efforts to falsify ever more challenging null hypotheses regarding software-based consciousness. Just as humans emerged from evolution not with a guarantee on survival, but only able to survive across many environmental conditions that defeated extinct species, Terasem-validated cyber-consciousness will not be able to claim scientifically proven unity with human consciousness. Instead, Terasem-validated cyber-consciousness can at best claim that it has falsified efforts to show it was materially distinguishable from various instantiations of brain-based consciousness. As Popper might have said, Terasem experimentally-validated cyber-consciousness is much more interesting than versions of software-based minds that proved the null hypothesis.

## 6. Limitations and Unique Aspects of the Terasem Experimental Design

A limitation of the Terasem experimental design is that it does not address head-on the question of software-based consciousness. For example, it may very well be easier to create a mind operating system that persuades experts of its consciousness as compared to the equivalence of its consciousness to that of a biological person. Ancillary to this weakness is that the comparator brain-based consciousness is only imminent in its mindfiles, which is clearly not as much of an apples-to-apples comparison as is the mindware-based purported consciousness. In other words, experts are being asked whether a software-based purported consciousness is equivalent to digital reflections of a brain-based actual consciousness. Hence, a weakness of the experimental design is that any disproof of the null hypothesis does not really say the software-based mind is equivalent to the brain-based mind, but only to digital reflections of the brain-based mind.

Another potential weakness is the Terasem experiment's reliance upon psychologists as arbiter's of consciousness. Psychologists may not in general be the most qualified persons to judge the existence of software-based consciousness. Furthermore, there are many specializations within the field of psychology.

A third limitation may be that should consciousness be in fact demonstrated, it is not clear that an ongoing experiment could be conducted with such consciousness under the aegis of the informed consent from biological participants. Research that is not ethical, or pursuant to good clinical practices, should not be conducted and, if nevertheless conducted, should not be cited or considered as contributing to the body of scientific knowledge. These three limitations are further discussed below.

### 6.1. *Why compare consciousness to that of digital reflections of a biological original instead of simply to a definition of consciousness?*

Minsky [2006] critiques those who decry the objectification of consciousness on the grounds that they are never precise about the definition of consciousness. He observes

that the word "consciousness" is a "suitcase word" that includes a plethora of subconcepts ranging from personality to perception, and from emotion to egotism. He notes that the difficulty of believing that consciousness can be established in software comes from thinking of it as an amorphous whole rather than as a set of interconnected programs.

The Terasem Mind Uploading Hypothesis aims to get around Minsky's criticism by substituting the amorphous concept of "is the software-based generator of digitized content conscious" for the specific, and measurable, concept of "is the software-based generator of digitized content at least equivalent to this other brain-based generator of digitized content?" This is an answerable question without ever having to plumb the depths of the possibly bottomless pit of consciousness. Let $C_B$ represent the consciousness of a brain-based being. Let $C_S$ represent the consciousness of a software-based being. Provided that it is agreed that $C_B = C_S$, it does not matter for the purposes of determining the consciousness of $C_S$ whether there is agreement on the constituents of $C_B$. Hence, the Terasem Mind Uploading Hypothesis, via its Turing Self Test, provides a more testable form of software consciousness than an assessment of software consciousness that lacks such a specific benchmark.

An unusual aspect of the Terasem Mind Uploading Experiment is that it compares purported software-based minds only indirectly with brain-based minds — via time spent reviewing the mindfiles of the brain-based minds, not via time spent actually interviewing, in real-time, the brain-based minds. Operationally this limitation is a useful part of the experimental design because many if not most of the brain-based contributors of mindfiles will no longer be legally alive[5] at the time of the experiment's multi-decade culmination. Even if they were alive, they may not be either well enough or willing to commit the time for a year-long series of interviews with a panel of psychologists.

Aside from operational convenience, this experimental structure has the advantage of delimiting the "fleshist" bias inherent in comparing a series of face-to-face, press-the-flesh interviews with a series of face-to-display interviews. Furthermore, if the mindfiles that will be reviewed by the assessment panel are good enough to give rise, with the help of mindware customization, to a persuasive appearance of software-based consciousness, then they should also be good enough to give the assessment panel an excellent feel for the biological original's consciousness.

[5] The term "legally alive" is used because under current law a person is judged to no longer be alive when there is irreversible cessation of brain activity. For example, in the United States, "An individual who has sustained either (1) irreversible cessation of circulatory and respiratory functions, or (2) irreversible cessation of all functions of the entire brain, including the brain stem, is dead". Somewhat different standards prevail in other countries. The US definition "is intentionally not entitled the Definition of Death Act. This is because it does not contain an exclusive definition of death". Uniform Law Commission, Determination of Death Act Summary, http://www.nccusl.org/ActSummary.aspx?title=Determination%20of%20Death%20Act, retrieved September 30, 2011. This leaves open the possibility for a distinction in the future to be drawn between "legal death" and "brain death". Such a distinction, if based on information theory, might provide that persons who have successfully copied their consciousness into software minds, to the satisfaction of appropriately certified psychologists, were not legally dead even if their biological brains irreversibly ceased functioning [Goertzl, 2010].

It is worthwhile to observe here that several studies have been done to assess the equivalence of "face-to-face" with remote, telecommunicated, psychiatry, also known as telepsychiatry. The balance of the studies indicate that telepsychiatry yields equivalent results such that "it is evident that telepsychiatry is one of the most widespread and accepted telemedicine applications" [Bashshur and Shannon, 2009]. The popularity of telepsychiatry implies that purported software-based consciousness can be compared remotely with digital reflections of brain-based consciousness. Since psychiatrists feel comfortable enough with telepsychiatry to diagnose and treat life-threatening mental illnesses they should also have comfort that they can adequately determine the presence, degree and homology of human consciousness.

It is also likely to be more socially useful to compare software consciousness to that of a brain-based original consciousness. Individuals who already have civil and political rights, which are all brain-based persons, have an ability to legally agitate for those rights to be extended to their software-based analogs. If it is shown that the brain-based and software-based consciousness are equivalent, then the brain-based original has achieved a significant extension of rights and privileges into the socio-economic space and time occupied by the software-based analog. However, if it is contended that a purported software-based consciousness exists without any reference to a brain-based master, then such software-based consciousness would have no legal rights as all such rights flow from biological birth and citizenship. Consequently, there is little motivation to establish with any legal certitude the consciousness of such *de novo* software-based minds.

### 6.2. *Why use psychologists as the arbiters of software-based consciousness?*

The endpoint for the Terasem experiment requires an assessment by psychologists because they are the profession's most versed in assessments of human consciousness. For example, in a court of law, when the question is one of somebody's state of mind (other than with respect to an element of a crime in the United States), to qualify as an expert witness one must have psychological certifications [Pozgar, 2011]. Another example is prior to a surgeon undertaking genital reassignment for a purported transsexual there are medical consensus standards calling for two mental health professionals to certify as to the gender consciousness of the patient [Standards of Care, 2001].

The general competence of psychologists on the topic of consciousness, at least as compared to other professions and to laymen, can still be challenged as not specific enough to truly ferret out cyber-consciousness from something masquerading as that. Conversely, absent specialized instruction in the reasons why consciousness can transcend substrate, it is likely that some psychologists could be "nativists" who categorically deny the possibility of cyber-consciousness. It would not be fair to have such psychologists on an endpoint assessment panel for an experiment in cyber-consciousness. Hence, it would be better stated for the Terasem Mind Uploading

Experiment that the endpoint assessment be performed by a panel of psychologists with certification in cyber-consciousness or machine intelligence. While such certifications do not exist today, they are likely to exist before the end of this decade.

A related criticism is that the requirement of an academic degree in psychology, or a medical degree in psychiatry, is wrongly limiting. As noted above, the certifications of "mental health professionals", who often are but need not be degreed psychologists, is deemed most appropriate for the determination (with significant juridico-medical implications) of true gender consciousness. Such mental health professionals other than degreed psychologists, with appropriate certification in cyberpsychology and consciousness, may provide a more useful set of persons from which to draw endpoint assessors.

### 6.3. *Why is the informed consent from a biological person valid for experimental testing upon a software-based consciousness?*

Since the Nuremberg Code was developed in 1947, there is general international agreement that informed consent is a prerequisite to, and a concomitant of, ethical medical research. The first principle of the Code states that:

> *The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, overreaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him/her to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonable to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment. The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.* [Mitscherlich and Mielke, 1949]

This principle has been expanded upon, and is current up through 2008, as expressed in the Helsinki Declaration of the World Medical Association. Paragraph 22 of the Declaration [2008] provides:

> *Participation by competent individuals as subjects in medical research must be voluntary. Although it may be appropriate to consult family members or community leaders, no competent individual may be enrolled in a research study unless he or she freely agrees.*

And Paragraph 24 of the Declaration [2008] expands upon this basic concept of truly voluntary (which implies non-fraudulent and thus honestly informed) consent:

> *In medical research involving competent human subjects, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail, and any other relevant aspects of the study. The potential subject must be informed of the right to refuse to participate in the study or to withdraw consent to participate at any time without reprisal. Special attention should be given to the specific information needs of individual potential subjects as well as to the methods used to deliver the information. After ensuring that the potential subject has understood the information, the physician or another appropriately qualified individual must then seek the potential subject's freely-given informed consent, preferably in writing. If the consent cannot be expressed in writing, the non-written consent must be formally documented and witnessed.*

Now it may be argued that the Terasem Mind Uploading Experiment complies with these provisions by virtue of the informed consent from a biological original as described earlier in this chapter.[6] However, a potential weakness of the experimental design is that the informed consent from a biological original is being held to apply to an ostensibly conscious, and humanly conscious, non-biological brain. As noted above in Paragraph 22 of the Helsinki Declaration, consent of family members (such as the biological original) is supportive, but not dispositive — it is the voluntary consent of the research subject, in this case the non-biological consciousness, which would determine the propriety of the experiment.

It might also be contended that modern rules regarding research consent apply only to humans. If they do not apply even to non-human animals, it may be argued, why would they apply to software? There is no evidence that the drafters of the Nuremberg Code or the Helsinki Declaration contemplated software-based research subjects. Yet, it seems disingenuous to escape from good clinical practices on the basis of an ontological difference between human subjects made of flesh and human subjects made of software. The very purpose of the experiment is to assess if a simulacra or analog of brain-based human consciousness can be created in software. How arbitrary it would be to then say "aha, we have a software-based form of human consciousness, but for purposes of medical ethics, it is not human". Surely this would be counter to the purposes of medical ethics. These were brilliant summarized by Harris [1985] as assuring the protection of lives that value such protection.

[6] Lifenaut.com Terms of Use Agreement, http://lifenaut.com/signup, retrieved September 27, 2011. CyBeRev Consent to *Ex Vivo* Consciousness Storage, https://www.cyberev.org/cmr/member_account/register/register.aspx, retrieved September 27, 2011.

At this point it may be observed that if the experiment is successful, then the software-based human consciousness is deemed equivalent to the brain-based human consciousness, informed consent from which was obtained originally. However, the potential flaw in the experimental design arises from the possibility that the null hypothesis is proven in such a way that the software-based human consciousness is not deemed equivalent to the original brain-based consciousness, but is nevertheless perceived by the expert psychological panel to be human consciousness. In this case, we have a human subject, albeit one based on software, participating in clinical research without having ever been asked for, much less tendered, informed consent.

The fact that an experiment is long underway is no exception to the obligation to have valid informed consent at all times. Paragraph 9 of the Nuremberg Code provides:

> *During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.*[7]

Indeed, as we are dealing here with the sensitivities of human psychology, the experiment should never bring the software-based human subject to an "impossible" mental state. Paragraph 4 of the Nuremberg Code [1947] provides:

> *The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.*

Very similar concepts are expressed in the contemporary Helsinki Declaration. Paragraph 28 notes that, "when a potential research subject who is deemed incompetent is able to give assent to decisions about participation in research, the physician must seek that assent in addition to the consent of the legally authorized representative." Hence, while the software-based analog being tested may be incompetent to give informed consent while under mindfile and mindware development, and even though consent is obtained from the biological original at the Lifenaut.com and CyBeRev.org websites, once it is determined that the software-based consciousness is both humanly competent and distinct from its brain-based consciousness parent, then, at that point, "the physician must seek" informed consent from the subject. This is not currently provided for in the Terasem Mind Uploading Experiment protocol.

Indeed, the Terasem protocol is further deficient because in such cases Paragraph 29 of the Helsinki Declaration insists that in:

> *Research involving subjects who are physically or mentally incapable of giving consent, for example, unconscious patients . . . the study may proceed without informed consent provided that the specific reasons for involving subjects with a condition that renders them unable to give*

[7] National Institutes of Health, Office of Human Subjects Research, Regulations and Ethical Guidelines, Directives for Human Experimentation, Nuremberg Code, http://ohsr.od.nih.gov/guidelines/nuremberg.html, retrieved September 27, 2011.

> *informed consent have been stated in the research protocol and the study has been approved by a research ethics committee.*

There is no research ethics committee in place at this time for the Terasem Mind Uploading Experiment. Rectification of this error is necessary in order to bring the studies within good clinical practices. Furthermore, current medical ethics research shows the scope of the research ethics committee's responsibility, if insufficiently broad, may need to be supplemented with a separate compliance function or office to ensure research integrity [Klitzman, 2011].

In summary, a weakness of the Terasem Human Mind Uploading Experiment is that it fails in some respects to comply with good clinical practices. Specifically, there is a failure to provide for informed consent from software-based human subjects that prove the null hypothesis both because such consent is never requested and because there is no research ethics committee that might approve the absence of such consent. However, it would not be difficult for these deficiencies to be rectified. For example, the purported software-based consciousness could be asked on a regular basis for informed consent. In this way, even if a divergence of person from the biological original was determined, there would be informed consent from the new software-based conscious person. To be more robustly compliant with good clinical practices such periodic re-obtaining of informed consent would best be the responsibility of an appropriately qualified individual other than those involved in the conduct of the experiment. The reason for this is succinctly summarized by Paragraph 26 of the Helsinki Declaration [2008]:

> *When seeking informed consent for participation in a research study the physician should be particularly cautious if the potential subject is in a dependent relationship with the physician or may consent under duress. In such situations the informed consent should be sought by an appropriately qualified individual who is completely independent of this relationship.*

The experimental protocol should also clearly provide for the caretaking and safekeeping of experimental subjects who refuse to give, or withdraw, their informed consent. Insofar as these subjects are software minds their needs will be different from, but perhaps no less challenging than, those of brain-based minds.

## 7. Summary of Assessment and Critique of the Terasem Mind Uploading Experiment

The experiment is a valid scientific challenge as it can be stated as a test of a well-stated null hypothesis that can be disproved but not proven true. Hence, an outcome that the software-based minds are equivalent to their brain-based forebearers will clearly establish that substrate borders between minds are not ontological boundaries. On the other hand, proof of the null hypothesis will simply drive scientific researchers further in their quest to develop mindfiles and mindware that can achieve

a continuity of mind across substrate ontology. Hence, this is a classic scientific endeavor fully in the footsteps of Popper and his many colleagues.

However, the experiment has the following deficiencies:

- Indefinite time period.
- Indefinite statement of assessor qualifications and number.
- Inadequate provision for regularly securing renewed informed consent as the software-based mind develops.
- No provision for an independent research review board to excuse informed consent for a software-based mind that diverges from its brain-based parent.
- No provision for an independent agent to assess the well-being and informed consent of the software-based mind that diverges from its brain-based parent.

All of these weaknesses can be readily addressed. It would be advisable for the deficiencies to be resolved as the experiment has great promise for illuminating the borders and boundaries of human consciousness.

## References

Bainbridge, W. [2003] "Massive Questionnaires for Personality Capture," *Social Science Computer Review* **21**(3), 267–280.

Bainbridge, W. [2006] "Strategies for personality transfer," *Journal of Personal Cybersconsciousness* **1**(4), 13–14, http://www.terasemjournal.org/PCJournal/PC0104/bainbridge01b.html.

Bashshur, R. and Shannon, G. [2009] *History of Telemedicine: Evolution, Context and Transformation* (Mary Ann Liebert Publishers, New Rochelle, NY), pp. 388–390.

Bock, P. [1993] *The Emergence of Artificial Cognition: An Introduction to Collective Learning* (World Scientific, New Jersey), p. 52.

Brown, D. [1991] *Human Universals* (McGraw Hill, New York, NY).

Costa, P. T. and McCrae, R. R. [1990] *Personality in Adulthood* (Guildford Press, New York).

Declaration [2008], *WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*, retrieved September 27, 2011, http://www.wma.net/en/30publications/10policies/b3/.

Doctor, R. [1990] *Transvestites and Transsexuals: Toward a Theory of Cross-Gender Behavior* (Plenum Press, New York), pp. 61–71.

Ebbinghaus, H. [1885] *On Memory*, translated by Ruger, H. and Bussenius, C. (1964) (Dover Press, New York).

Edelman, G. [2004] *Wider than the Sky: The Phenomenal Gift of Consciousness* (Penguin, London), pp. 16–138.

Gemmell, J., Bell, G. and Lueder, R. [2006] "Mylifebits: A personal database for everything," *Communications of the ACM* **49**, 88–95.

Goertzl, B. [2010] *A Cosmist Manifesto* (Humanity Plus Press, United States), pp. 213–221.

Gooren, L. [1995] "Biological aspects of transsexualism and their relevance to its legal aspects," in *Transsexualism, Medicine and the Law, Proc. XXIIIrd Colloquy on European Law* (Amsterdam, Holland) (Council of Europe Publishing, Strasbourg), pp. 117–136.

Gould, J. [1999] *The Animal Mind* (Scientific American Library, New York), p. 217.

Harris, J. [1985] *The Value of Life: An Introduction to Medical Ethics* (Routledge, London), p. 53.

Klitzman, R. [2011] Views and experiences of IRBs concerning research integrity, *Journal of Law Medicine and Ethics* **39**(3), 513.

Kurzweil, R. [1999] *The Age of Spiritual Machines* (Viking, New York).

Kurzweil, R [2005] *The Singularity is Near* (Viking, New York), pp. 110−140.

Landauer, T. [1986] "How much do people remember? Some estimates of the quantity of learned information in long-term memory," *Cognitive Science* **10**(4), 477−493.

Lloyd, D. [1989] *Simple Minds* (MIT Press, Cambridge, MA).

Minsky, M. [2006] *The Emotion Machine* (Simon and Schuster, New York), pp. 6, 109−111.

Mitscherlich, A. and Mielke, F. [1949] *Doctors of Infamy: The Story of the Nazi Medical Crimes* (Schuman, New York), Nuremberg Code, 1947, pp. xxiii−xxv.

Nuremberg Code [1947] National Institutes of Health, Office of Human Subjects Research, Regulations and Ethical Guidelines, Directives for Human Experimentation, Nuremberg Code, `http://ohsr.od.nih.gov/guidelines/nuremberg.html`.

Popper, K. [1935] *The Logic of Scientific Discovery* (Verlag van Julius, Springer, Vienna), republished in 2002 by Routledge, New York, p. 10.

Popper, K. [1972] *Objective Knowledge: An Evolutionary Approach* (Oxford University Press, Oxford), pp. 57−58.

Pozgar, G. [2011] *Legal Aspects of Health Care Administration* (Jones and Bartlett Learning, Sudbury, MA), p. 122.

Reid, R. [1995] "Psychiatric and psychological aspects of transsexualism," in *Transsexualism, Medicine and the Law, Proc. XXIIIrd Colloquy on European Law* (Amsterdam, Holland) (Council of Europe Publishing, Strasbourg), pp. 25−50.

Rothblatt, M. [1995] "An American perspective on transgender health law," in *Transsexualism, Medicine and the Law, Proc. XXIIIrd Colloquy on European Law* (Amsterdam, Holland) (Council of Europe Publishing, Strasbourg), pp. 189−202.

Rothblatt, M. [2011] "From mind loading to mind cloning: From genes to memes to bemes", in *Transhumanism and Its Critics*, (eds.) Hansell, G. and Grassie, W. (Metanexus, Philadelphia), pp. 114−118.

Standards of Care [2001], *Harry Benjamin International Gender Dysphoria Association's Standards of Care for Gender Identity Disorders*, Sixth Version, `http://www.wpath.org/documents2/socv6.pdf`.

Turing, A. [1950] "Computing machinery and intelligence," *Mind* **59** (1950), reprinted in *The Essential Turing*, (eds.) Copeland, B. (Clarendon Press, Oxford), pp. 433−450, 441.

Vinge, V. [2008] "Signs of the singularity," *IEEE Spectrum* **45**(6), 77−79.