

Historical Simulations – Motivational, Ethical and Legal Issues

Peter S. Jenkins*
Attorney at Law
USA

Abstract

A future society will very likely have the technological ability and the motivation to create large numbers of completely realistic historical simulations and be able to overcome any ethical and legal obstacles to doing so. It is thus highly probable that we are a form of artificial intelligence inhabiting one of these simulations. To avoid stacking (i.e. simulations within simulations), the termination of these simulations is likely to be the point in history when the technology to create them first became widely available, (estimated to be 2050). Long range planning beyond this date would therefore be futile.

Key Words: historical simulation, virtual world, artificial intelligence, massively multiple online role-playing games, ethics of clinical research, Moore's Law.

Introduction

The notion that the perceived world is an illusion or a simulation has arisen for centuries in the works of philosophers¹, mathematicians², and social scientists.³ A recent variant on this theme, posited by

Nick Bostrom of the University of Oxford, is that it is possible that we are forms of artificial intelligence in an ancestor, (i.e. historical) simulation created by a future society.⁴ Moore's Law, which has held true for about 40 years, states that computer processing power doubles approximately

* Member of the State Bar of California. He would like to thank John Smart for his helpful comments at the very early stages of writing this paper, Greg Lastowka and some anonymous referees for some insights when it was nearing completion and Edward Castronova for his encouragement.

every 2 years. The prominent futurist, Ray Kurzweil, estimates that this is currently accelerating to doubling every year and that by approximately the year 2050, for the then equivalent of \$1,000, you will be able to purchase a computer with greater processing power than that of all the brains of all humans that have ever lived⁵. This means that by 2050 it would be feasible to have a completely realistic historical simulation running on every desktop, and that these simulated worlds would outnumber the real one by a factor of millions or even billions to one. This makes it almost certain that we live in one of the simulations if a future society has the motivation to create them and does not encounter any insurmountable ethical and legal obstacles to doing so.

Bostrom's view is that a future society would likely not have the motivation to create these simulations or alternatively may not be able to overcome the ethical and legal impediments involved⁶. The purpose of this paper is to demonstrate that there is a high probability that a future society will have strong and extensive motivation to create these simulations, and will be able to overcome any ethical and legal impediments⁷. 2050 is only 44 years away, and the development of ethical and legal systems tends to lag significantly behind technological progress⁸. Therefore, the same types of ethical and legal issues, as well as motivational ones, that we face today will likely be similar to the ones providing a framework for a future society's deliberations about implementing their ability to create and run historical simulations.

Motivational Issues

Would future societies likely not have any motivation to create and run historical simulations, either because they have more efficient ways of amusing themselves (e.g. by directly stimulating the pleasure centers of the brain) or they deem the creation of these types of simulations to be frivolous and of no scientific, research or other practical value?⁹

1. Nostalgia and the Rear View Mirror Effect

The skeptical view is that, although many members of present day society would probably wish to create historical simulations if they were capable of doing so, members of a future society would be significantly different in this regard¹⁰. However, although this might be true in a future society whose members have fully evolved into machine form, it would likely not be true for a future society that can make historical simulations but whose members have not yet fully shed their biological form. In an era of rapid change in the future, it is very probable that there may be a high degree of nostalgia and interest in the past in the same manner that neo-classicism arose in Europe during the paradigm shift of the Industrial Revolution in the mid to late 18th century¹¹.

There is also the "rear view mirror" effect noted by Marshall McLuhan, where new media technologies tend to use the content of the old media before developing content of their own, e.g. when television was first invented, the content was primarily the stage play¹². Therefore, when a

future society develops advanced quantum computing, it is likely to first use it for historical simulations (the equivalent of stage plays on early television) prior to finding new uses and content for it, such as fine-tuning their universe themselves, or communicating with other universes.

2. Testing Ground for Artificial Intelligence (AI)

Another possible motivation for the creation of historical simulations would be to use them as a safe environment in which to test newly manufactured artificially intelligent entities before releasing them into the real world. A future world could be composed entirely of artificially intelligent entities, or a combination of these entities and technologically enhanced human beings. In either case, the interaction of newly created AI units with their established AI counterparts or with technologically enhanced humans would be critical. High standards of ethical behavior would likely be expected of such newly created AI's and these standards may not be susceptible of being easily programmed into these machines. Observing a simulated world in which the AI entities inhabiting it are not aware that they are in a simulation would be a safe, reliable method of determining whether an AI's ethical programming requires adjustment or perhaps instead is fundamentally flawed and beyond repair. Such a simulation would be set in a historical period prior to the point when fine-grained simulation technology was first developed, otherwise the AI's inhabiting the simulation would suspect that the world they are inhabiting is not real, and

their behavior would not be genuine.

3. Social and Economic Experiments

As the economist Edward Castronova has noted in his groundbreaking book, *Synthetic Worlds – the Business and Culture of Online Games*,¹³ these worlds, currently known as Massively Multiple Online Role-Playing Games (MMORPG's) can provide an excellent social science laboratory tool. He states that future generations of PhD students in anthropology, sociology, political science and economics will likely work with pairs of worlds (an experimental and a control) seeking to test various hypotheses by tweaking the parameters and observing the differing results in each case. Versions or portions of online worlds such as *SimCity* and *Second Life* are currently being used for policy and strategic analysis by think-tanks, universities and the U.S. Department of Defense.¹⁴ It could be expected that this type of trend would continue in a future society that is capable of producing fine-grained simulations that are indistinguishable from the real world. Simulated worlds created by a future society to solve policy, strategic and research issues would most likely be retrospectives, i.e. historical simulations in which artificial intelligence would genuinely believe itself to be human, rather than merely playing the role of a human. These simulations could provide a rich source of information to a future society about how it arrived at its current stage of development as well as how it could avoid repeating the mistakes of the past.

4. Apocalypse

In the event of apocalyptic events such as the release of malevolent genetically modified organisms or self-replicating nanobots, humans may want to completely download their consciousness into machines and forsake their physical bodies completely, in which case it would make sense to enhance the realism of the experience by erasing the memory of the download occurring in the first place as well as the horrific events that led up to it. Alternatively, emergency management or military strategists may wish to be proactive in terms of planning for apocalypse or Armageddon and create a simulation as a back-up system for civilization. It would be periodically updated but not activated until the Doomsday Clock reaches one minute before midnight. Such simulations would obviously be set in the period prior to when the disaster occurred and also before fine-grained simulations were developed. As such, they would not necessarily be historical simulations in the strict sense, but rather may be simulations of the lives of the participants as they were a few short years earlier. However, they would be consistent with the notion that we live in a simulation without being aware of that fact, especially when one considers that it is very possible that malevolent self-replicating nanobots, genetically modified bio-organisms and thermonuclear war could destroy civilization within our lifetimes¹⁵.

Ethical Issues

Having determined that a future society on earth would very likely *want* to create historical simulations for various purposes (motivation), the next question that needs to be addressed is whether it would come to the conclusion that there is no valid reason why it should not create such simulations (ethics).

A good framework for the discussion of this issue can be readily found in the existing work done in the area of the ethics of medical research. The recent work of Emmanuel, Wendler and Grady¹⁶ provides a useful summary of the generally recognized ethical requirements of clinical research as set out in several prominent documents and codes.¹⁷ This summary takes the form of seven requirements for the research protocol, which must:

1. have social, scientific or clinical value that justifies exposing subjects to potential harm;
2. be scientifically rigorous;
3. select subjects fairly on the basis of scientific objectives and not, for example, because of vulnerability or privilege;
4. minimize the risk to individual subjects, and have potential benefits to those subjects and/or society that outweigh or are proportionate to the risks;
5. be reviewed and approved prospectively by a committee of independent and qualified evaluators;
6. be conditioned, to the extent possible, on the voluntary and informed consent of its participating subjects; and

7. ensure that enrolled subjects are shown respect, which includes protecting their privacy, monitoring their well-being, and providing opportunities to withdraw.

1. Is the historical simulation essentially just an entertainment?

There is no doubt that some potential harm and pain would be inflicted on the artificial intelligences inhabiting the simulation, and that it would not be a Garden of Eden. The first of the seven requirements, i.e. that the protocol have social, scientific or clinical value that justifies exposing subjects to potential harm is useful in evaluating the prospects of a future society creating historical simulations for purely nostalgic and entertainment purposes. On first impression, it would seem that a simulation created for such purposes would have no social, scientific or clinical value, and would be purely a pleasant diversion. However, the issue warrants further examination. In her insightful book, *The Future of Nostalgia*¹⁸, Professor Svetlana Boym of Harvard University indicates that there are two types of nostalgia – restorative and reflective. Restorative nostalgia emphasizes the Greek *nostos* root, i.e. the return home as in Homer's *Odyssey*. It seeks to rebuild the lost home and fill in gaps in memory. Reflective nostalgia, on the other hand, comes from the *algia* part of the Greek root, i.e. the feeling of longing and loss due to the imperfect nature of remembrance. As Boym points out, what is needed following historical cataclysms and periods of rapid change, is not to literally recreate

the monuments of home (restorative nostalgia) but rather to mourn the loss of the space of shared cultural experience within which one elected pursuits according to one's own affinities (reflective nostalgia). This mourning facilitates the existence of the space between the individual and the environment that is formed in early childhood and is integral to human nature.

If a historical simulation could support the therapeutic, reflective form of nostalgia, then it would have social or even clinical value during the period of rapid upheaval as society transitions to its next stage. However, arguably a fine-grained historical simulation that recreates a past environment in every detail so that it is indistinguishable from the former reality would actually support the literal and somewhat frivolous restorative form of nostalgia rather than the therapeutic reflective type. Viewed in this light, the historical simulation would be nothing more than an über-Disneyland having entertainment but no real social value.

On the other hand, it is possible that the historical simulation could be structured so as to permit reflective nostalgia by creating an environment in which the loss of shared former cultural space could be properly mourned, e.g. by simulating a mythical city containing actual historical fragments and cultural artifacts in such a way that each person's experience of it would be evocative and subjective, rather than literal. Alternatively, a historical simulation where the landscape altered according to one's own internal reflections, or a form of embed-

ded personalized augmented reality¹⁹ overlaying the simulation would accomplish the same purpose. Therefore, even a simulation created for nostalgic purposes could have important social or clinical value.

2. Informed consent – motivation and method

A significant issue in the study of the ethics of medical research is whether the social, scientific or clinical value of the medical protocol need accrue to the participants specifically in order to outweigh potential harms or whether, on the other hand, the value could accrue to society as a whole, or parts thereof. The recent consensus on this issue²⁰ seems to be that the benefits need not accrue to the participant specifically and that they can accrue to the population at large. The rationale is that a competent adult does not need to be protected from choosing to assume reasonable risks when there are important subjective reasons to do so that are unrelated to that individual's personal health. Some participants in medical research do so, for example, to help conquer a disease that a loved one died from, in honor of their memory. They might also do so because they benefited from past medical research concerning an unrelated illness and as a gesture of gratitude they wish to help others with different illnesses. Would similar sorts of altruistic motivations apply to artificial intelligence in connection with agreeing to participate in a historical simulation? Yes, they might apply, for example, in the case of an artificial intelligence that had

personally benefited from information garnered from a previous experiment in which an AI had agreed to participate in a historical simulation. Such an experiment may have led to an improved form of AI of which the new proposed participant was a product.

There may also be a form of inter-species reciprocal altruism²¹ in the sense of AI wanting to benefit humans, e.g. by playing out various scenarios in the simulation to assist humans better understand their history, while humans assist AI's evolution. This may involve implicit or explicit promises by the researchers that the AI would receive future preferential treatment in terms of enhancement in programming or maintenance or other benefits in return for agreeing to participate in the historical simulation. This raises some vexing ethical concerns, since AI could be characterized as being in a class of vulnerable persons such as children, prisoners and expectant mothers for purposes of determining whether truly voluntary and informed consent had been given. As I will discuss in the next section of the paper, the general principle, as embodied in the U.S. federal regulations concerning experiments with human subjects²², is that vulnerable individuals such as prisoners must not be offered incentives for participating in the research that would compromise their ability to objectively evaluate the risks involved.

There is also the question of the method of obtaining the consent. The premise of the historical simulation is that the artificial intelligences inhabiting it are not aware that they

are in such a simulation. From their perspective, they are living in the real world. How could they possibly be considered to have given informed consent generally in such a situation, let alone informed consent specifically saying that they acknowledge that the benefits accrue not to them but to others, i.e. other AI's or the humans observing the simulation through their avatars? The answer is that the AI would be asked for this consent while they existed outside of the simulation and then, if they agreed, would have memories of the choice wiped clean. When asked to give consent, the AI would not be informed what sort of life it would lead in the simulation, otherwise most AI's would not give consent unless the proposed life was somehow exceptional, e.g. Albert Einstein, Elvis Presley, Queen Elizabeth etc.

However, it would be necessary to inform the AI that it would be guaranteed not to be put into a life of extreme pain, suffering or hardship, e.g. a peasant living in grinding poverty. These sorts of roles in the simulation could feasibly be played by automated programs ("bots") that are not self-aware. This implies that if you are a conscious individual who is experiencing extreme distress, then you probably exist in the real world of the early 21st century rather than in a simulation run by a future society. Conversely, if you are an individual whose life generally involves happiness, freedom and self-fulfillment, then you probably exist in a simulation, (although there is one real version of you in the early 21st century). This raises some difficult ethical questions vis-à-vis

the treatment of the severely disadvantaged by those who are better off. If the members of the latter group view members of the former as merely simulated, non-conscious entities, then they may have an incentive not to accord them the respect and dignity to which they are entitled. However, if there are any conscious AI's which have the extraordinary perceptiveness to realize that they exist in a simulation, they would probably also be cognizant that one of the reasons they are there is to evaluate their ethical standards prior to being allowed to leave the simulation and enter the future paradise outside it. Therefore, these AI's would not be lacking an incentive to treat the disadvantaged bots in the simulation with proper care and concern. As Edward Castronova has pointed out²³, in MMORPG's the human players conform to the patterns of the non-player character (NPC) bots. Similarly, in a simulation, the AI's may tailor their behavior to the needs of the disadvantaged bots to demonstrate ethical behavior.

An analogy may be drawn between the approach of asking the AI to consent to a random assignment to a role in the simulation (except one involving extreme suffering or hardship) and the practice of using a randomized placebo-controlled drug trial where a participant is informed that he or she will be bypassing standard medical care and there is a 50% chance that he or she may receive a placebo (i.e. a dummy pill) instead of the experimental drug. This practice is somewhat controversial in that some physicians are of the view that there is no scientific

need for placebos in an experiment. This is in distinction to non-placebo protocols which are generally not controversial such as where the patient undergoes some risks without any compensating medical benefit, for example, agreeing to undergo a lumbar puncture procedure as part of a clinical study where the participant does not medically require such a procedure. However, there is a strong school of thought that, assuming placebos are necessary for the integrity of the experiment's results, then, if there is no serious, life-threatening risk to the patient, and all the seven requirements listed above are met, there is no ethical difference between the placebo type of protocol and the non-placebo type, such as the lumbar puncture²⁴. In both situations, the participant gives consent after being informed that he or she will be exposed to a limited and reasonable risk possibly without any compensating medical benefit to him or her, in order to assist in the advancement of medical science.

It is also significant to point out that requirement #6 as set out above states that the protocol must be conditioned on the informed consent of the participants, *to the extent possible*, in recognition of the fact that it is not always possible to provide the participants with all the information concerning the experiment, such as for example in psychological experiments. A notable example is the Milgram experiment²⁵ designed to measure the willingness to blindly follow instructions even if they result in inflicting pain on another human being. This experi-

ment involves a test administrator A asking B, who is the real subject, to assist in giving a series of progressively stronger electrical shocks to a "subject" C, who is actually a confederate of A. C then feigns increasing distress as B, following instructions from A, turns up the dial on the mock electrical device. The deceptive nature of the experiment is not without its ethical critics, but it also has supporters who note that most participants were reportedly glad to have served in Milgram experiment, despite experiencing some psychological distress during it. Similar reactions have been obtained from many other persons who were deceived during psychological experiments.²⁶ Furthermore, most of the criticism of the Milgram experiment relates to the total lack of informed consent, which would be distinguishable from historical simulations where there would be informed consent as indicated previously.

3. Opportunities to withdraw

There is one more significant ethical issue that can be unpacked from the list of seven requirements set out above. This relates to requirement #7, which is that the participants' well-being should be monitored during the experiment and that they should be given the opportunity to withdraw from the experiment. The only fair and practical way to give a meaningful opportunity of withdrawal for the AI would be for the creator to periodically inform it that it exists in a simulation and that, if it wishes, it can end its presence in the simulation. There

might be a penalty system to discourage excessive use of the "restart" option, such as an increased chance of being worse-off. For those AI's that chose not to end their lives in the AI after having received the information from the creator, their memories would be wiped clean of the information and they would continue to live their simulated lives, unaware that they were not human. The fact that many players of MMORPG's currently view themselves as citizens of the synthetic world that they play in²⁷ and that some games have to periodically remind players after a certain amount of hours of play per week not to neglect their real world activities²⁸ indicates that there may be appeal for AI's in continuing to play the "game" of the historical simulation.²⁹ Although, as in MMORPG's, this process in historical simulations would tend to weed out AI's that do not have a substantial amount of curiosity, stamina and competitiveness, this would not be much different from the process of natural selection in the real world and would therefore not significantly skew the results or interfere with the realism of the simulation. Furthermore, some recent research in medical ethics has indicated that participants in experiments should not automatically have an absolute or unconditional right to withdraw from the experiment and that reasonable conditions can be attached to the right to withdraw, such as a requirement to enter into a dialogue or negotiation with the researcher prior to making a decision.³⁰

Finally, it is interesting to note that, despite the fact that one of the

basic tenets of the Islamic religion is the firm and widespread belief that martyrdom by suicide will involve instant and guaranteed transportation to paradise, only an extremely small proportion (less than 5 in a million) of members of that religion actually act on that belief. This indicates that there is a basic human tendency, due to the endowment effect³¹, or status quo bias, to remain immersed in life despite being offered appealing exit options. It is certainly reasonable to expect that this trait would carry over to AI's in a historical simulation in terms of how they would respond to an offer to withdraw, especially if the offer were to be presented to them inside the simulation, and characterized as a decision that is permanent and irrevocable.

Legal Issues

1. Would AI have the legal status of a person?

The question of whether AI would have the legal status of a person has been considered by many lawyers, legal scholars and computer scientists to date, although not in the context of a historical simulation. Most of these individuals have come to the conclusion that AI would meet the definition of personhood on the basis of the having the attributes of reasoning, self-awareness, communication, a sense of the past and the future, and the ability to experience pain and pleasure³². Of course, the absence of any one or more of these is not necessarily critical to the issue of personhood status, e.g. in the case of the profoundly retarded, the comatose, the brain dead,

third trimester fetuses, and newly born infants. However, the fact that AI would likely possess all of the attributes on the list, and more, would definitely indicate that it should be granted personhood status. The fact that the AI's consciousness resides in a different substrate, e.g. silicon, carbon nanotubes, quantum dots etc., than human consciousness is not a valid reason to deny it equal status³³. The implications of this equality for AI and society generally would be extensive and profound. AI would have the right to life, (i.e. not to be unplugged) and the right not to be subject to intentional infliction of emotional distress, (i.e. the right not to be exposed to threats that it will be unplugged) and the right to receive critical medical care.³⁴

2. Would AI be considered as a vulnerable class of person?

A sub-issue of the personhood one is the question as to whether AI would be considered to be a member of a vulnerable class of persons such as, for example, children, prisoners, the poor and the disabled. If that is the case, then as discussed previously, special considerations would likely apply for purposes of the informed consent issue. An AI would likely have superior reasoning powers to most humans and yet it would be in the process of learning how to apply that knowledge in an appropriate manner and would likely not have fully developed that ability in the early stages. An AI would likely have little or no freedom of movement in the early stages of its development. It would probably either be confined to a computer lab like a

prisoner or else, if it were housed in a mobile robot, it would be limited to a small area of exploration, or required to be accompanied at all times by a human handler. It would also be dependent on the researchers who built it for new information, programming enhancements, electrical power and maintenance, like an indigent person who is dependent on welfare for the basic necessities of life. Therefore, special precautions would have to be undertaken to ensure that an AI gives truly voluntary, informed consent to participate in a historical simulation if it is at an early stage of its development. By way of analogy, the U.S. federal regulations on Protection of Human Research Subjects require that, in the case of prisoners³⁵, there be some additional steps taken, including the following:

1. the majority of the Institutional Review Board (IRB), aside from prisoner members, should have no association with the prison; (in this case, it would be the computer lab)
2. at least one member of the IRB should be a prisoner or a prisoner representative/advocate with an appropriate background (in this case, another AI in the lab or a fully developed AI outside the lab);
3. parole boards should not take into account the fact that the prisoner participated in the research in determining whether to grant early release (in this case, the computer lab should arguably not take it into account in determining whether the AI should be released early from the lab

- into the world);
4. the prisoner should not be tempted by possible advantages of participating in the experiment such as improved living conditions, opportunity for earnings, medical care etc. such that he or she cannot objectively assess the risks involved (in this case, the computer lab should not offer enhanced programming as an incentive to take part in the historical simulation);
 5. the research must concern conditions particularly affecting prisoners as a class, or have the intent and reasonable probability of improving their health and well-being (in this case, the historical simulation would have to assist the AI in its development, such as perhaps by enabling it to better relate to humans).

Therefore, by analogizing from the regulations concerning experiments on prisoners, it is evident that, even for AI's that remain confined to the computer lab and have not been released into the world, it is possible, with certain safeguards in place, to ensure that truly voluntary, informed consent has been obtained for the AI's participation in the historical simulation. Having AI's participate in historical simulations while still in the lab would be beneficial in that the lab could use these trials to weed out and re-program "bad seeds" and also to train AI's in human empathy and emotional intelligence, since the AI while in the simulation would genuinely believe itself to be human.

A recent development in the

case law regarding medical experiments involving human subjects, which may be relevant to the issue of informed consent in historical simulations, is the concept of harm to dignity. This doctrine indicates that there is a legally and constitutionally protectable interest in medical choice for human subjects, regardless of whether or not there is actual injury to the subject. In the case *Díaz v. Hillsborough County Hospital Authority*,³⁶ the plaintiff was a sixteen year old Hispanic girl in her first pregnancy who attended the Tampa General Hospital's high risk clinic. Upon admission to the hospital, she was given sedatives to arrest her preterm labor and, while in a drowsy state from the drugs, she signed a complicated, three page, English language consent form document to participate in a drug study for fetal lung immaturity. Díaz, along with 384 other women, participated in the study with no adverse physical effects. However, they felt that their consent was obtained through coercion, and instituted a class action against the hospital for harm to human dignity. Ultimately, the case was settled for \$3.8 million, covering a class of about 5,000 pregnant women who had been subject to various medical experiments. What is significant about this case is that it indicates that there is a compensable harm resulting from injury to human dignity in medical experiments, even where there is no actual injury or harm to the patient³⁷. Even if the AI has not suffered harm as a result of participating in the experiment, it may have a cause of action for harm to its dignity on the basis that, like a

16 year old pregnant Hispanic girl who cannot understand the complexities of a three page English language consent form, a newly created AI cannot fully fathom the depths of human suffering when it has never been inside a historical simulation, and hence cannot properly give informed consent to enter one. This argument is, of course, a highly speculative one, and should not create an insurmountable obstacle to would-be creators of historical simulations. It will likely, at best, be one of the many factors tending to lead them in the direction of caution and discretion in the process of obtaining informed consent for these simulations.

3. Intentional Infliction of Emotional Distress

In the past few years, there have been several lawsuits for intentional infliction of emotional distress³⁸ launched by participants in Reality TV shows³⁹, which are popular programs in which participants are put into embarrassing, humiliating and/or stressful competitions or situations. Often, to enhance the drama, conflict and suspense factors, the participants have been deceived by the producers of the program as to the real circumstances involved in the program.⁴⁰ Would the litigation concerning Reality TV provide a precedent for possible lawsuits concerning historical simulations or is it distinguishable?

A good defense that could be established by the creators of historical simulations in response to lawsuits for emotional distress would be that when the AI exited the simula-

tion, the simulated human life that it had been living ended, and therefore any pain and suffering that it endured in the simulation should not form a cause of action by the AI. This would be similar to the general principle in many jurisdictions that a cause of action for pain and suffering cannot be continued by the estate of a deceased person.⁴¹ The notion is that the pain and suffering that the deceased experienced during his or her lifetime does not actually diminish the value of the deceased's estate, and therefore the executor of that estate cannot commence or continue an action for such pain and suffering. In the context of an experiment on AI in a historical simulation, the life played by the AI has terminated, and the AI is analogous to the executor.

The creators of the simulation would be bolstered in their argument by the fact that the AI's direct memory of the events in the historical simulation could be wiped clean. There would be some drawbacks to this memory erasure, in that the AI would not be able to learn from the experience of life in the simulation, but the erasure remedy would not be used in every case, of course, only in ones where there was a serious risk of a lawsuit for intentional infliction of emotional distress, due to the circumstances of the simulation. Even in situations where erasure was used, it would still be possible to utilize the AI's experiences in the simulation for purposes of weeding out manufacturing or programming defects in the AI. It may also be possible to use the AI's experiences for purposes of social and economic experiments generally that extended

beyond the individual learning of the AI.

However, where the AI endured exceptional pain and suffering, perhaps due to an unpredictable chain of events rapidly flowing from the unique nature of the simulation, it would be wise to apply an exclusionary rule to prohibit the use of any data arising from such pain and suffering, so as not to encourage the creators of the simulation or others to run similar experiments in the future⁴².

4. Publicity, privacy and copyright issues

There is also the issue of whether historical simulations would involve the consideration of the publicity or privacy rights, or both, of the individuals whose lives are being simulated. Although publicity rights, as property rights that are part of the deceased's estate, generally continue after the death of the individual, they are subject to overarching freedom of speech concerns that allow the use of the individual's image for informational and parody/satire purposes, but not for commercial/advertising purposes. Historical simulations would probably fall into the former category. More problematic then would be the issue of privacy rights, which have been traditionally asserted by non-celebrities. Although historically, privacy rights, as a tort like intentional infliction of emotional distress, have tended not to continue after an individual's death, there has been a trend in the U.S. and other countries such as Germany, for example, towards recognizing that privacy rights in some circumstances continue after death.⁴³

These rights are sometimes referred to as a right of anonymity. To the extent that this legal trend continues and is adopted by a future society, there may be some constraints on the creation and running of historical simulations. However, the more time that elapses between the death of the deceased and the piercing of the veil of privacy, the less danger there is of a successful lawsuit for violation of privacy rights.

Finally, there are copyright issues associated with illegal practices such as file-sharing, as well as other issues which may be in more of a gray area such as bricolage, or "mashups" where individuals take pieces of videos or songs that they have purchased and combine them in unique ways for personal and other non-commercial uses.⁴⁴ The likelihood that we live in a simulation would be greatly increased by the extent to which these simulations are copied and distributed, which might be by future file sharing systems. Assuming that the current copyright prohibitions on file-sharing and on technologies that actively encourage users to file-share are continued and adopted by a future society⁴⁵, then would there be a role for bricolage or "mashups" instead in a future society running historical simulations? A person in a future society may be interested in taking a simulation of the Soviet Revolution of 1917, for example, and altering it to remove Lenin and substituting a simulated Russian ancestor from his own family tree and then observing what happens. Would the creation of a unique new version for the personal non-commercial use of the bricoleur and his or her friends con-

stitute fair use or would it be a violation of intellectual property laws? Lawrence Lessig⁴⁶ would probably lean towards concluding that it is not or should not be a violation of intellectual property rights, but in my view this may be a bit over-optimistic⁴⁷, and therefore there may be an additional constraint, at least in theory, on the numbers of historical simulations that are run.

Nonetheless, even a future society would likely have some practical difficulties in suppressing the dissemination of large numbers of personalized historical simulations created by bricoleurs of the future. Furthermore, given that technology has recently been developed to allow individuals to create their own personalized MMORPG without a significant investment⁴⁸, it may become possible for future individuals to create their own historical simulations from scratch in the same manner, in which case copyright issues due to bricolage would not arise.

Conclusion

Historical simulations can be created, modified and copied with relative rapidity compared to the billions of years it takes for a real civilization to evolve. Furthermore, it is highly probable that a future society would have extensive motivation to create and run historical simulations and not encounter any insurmountable ethical and legal obstacles to doing so. Therefore, any presumption that we live in a real world is rebutted and it is very likely that we live in a simulated world created by a future society. What do we do?

Nick Bostrom suggests that, if we do live in a historical simulation, then it is possible that our behavior is being evaluated by the creators of the simulation to determine our suitability for admission into a future afterlife that exists outside of the simulation and so we may have an incentive to act ethically.⁴⁹ Robin Hanson suggests that we should endeavor to continue to lead an interesting life or else associate with interesting people, so that the creators of our simulation will not unplug it out of boredom.⁵⁰ These are both possible answers to the question of what we should do upon realizing that we probably live in a simulation. Hanson's concept of endeavoring to lead interesting lives and associate with interesting people to ensure the continued existence of the simulation touches on a key issue, which is that of when the simulation can be expected to end. The creators of the simulation would likely not continue it past the point in history when the technology to create and run these simulations on a widespread basis was first developed. The reason for stopping it at this point is that, as noted previously, a historical simulation that is set in a period when the necessary simulation technology already exists would tend to stymie any efforts to make the simulated entities unaware that they exist in a simulation. This lack of awareness is necessary for the simulation to run effectively, otherwise the behavior of the AI's inhabiting the simulation would not be genuine and the basic purpose of the simulation would not be accomplished. Another reason is to avoid stacking of simulations, i.e. simulations within simu-

lations, which would inevitably at some point overload the base machine on which all of the simulations are running, thereby causing all of the worlds to disappear. This is illustrated by the fact that, as Seth Lloyd of MIT has noted in his recent book, *Programming the Universe*, if every single elementary particle in the real universe were devoted to quantum computation, it would be able to perform 10^{122} operations per second on 10^{92} bits of information⁵¹. In a stacked simulation scenario, where 10^6 simulations are progressively stacked, after only 16 generations, the number of simulations would exceed by a factor of 10^4 the total number of bits of information available for computation in the real universe.

Therefore, the end of history can be anticipated as arriving when the technology necessary to create and run historical simulations on a widespread basis becomes feasible. This is estimated to be 2050, based on Ray Kurzweil's projections of Moore's Law. When I say, the "end of history", I mean it both in the Hegel /Fukuyama⁵² sense, i.e. when humankind reaches its ultimate end state, and in the technological singularity sense⁵³, i.e. where the world as we know it winks out of existence and we are transported to a collective consciousness in the form of a generalized artificial intelligence of which, as individualized AI's or as technologically enhanced humans, we would each constitute a part. In terms of how we should behave in light of this knowledge, we should recognize that long range planning beyond 2050 would be futile.

Correspondence

Peter S. Jenkins
11 Ellsworth Avenue Toronto,
Ontario Canada MCG 2K 4
Personal weblog:
<http://petabytes.typepad.com/>
Email: peterjenkins@rogers.com

Notes

1. Plato.
2. Descartes 1664.
3. Baudrillard 1994.
4. Bostrom 2003.
5. Kurzweil 2005, p. 70.
6. Bostrom 2006, Q.2.
7. Due to his "bland indifference principle" (Bostrom 2003), p. 7, Bostrom assumes that one could be located anywhere in the universe that is compatible with life, whereas I examine the issue in the context of earth since, as Kurzweil states, the complete absence of evidence of extraterrestrial life indicates that we should assume (until proven otherwise) that we are the first civilization in the universe (Kurzweil 2005), p.357. Furthermore, Bostrom does not make any assumptions about when the necessary simulation technology would become available (Bostrom 2003), p.3. Due to the lack of a clear context of time and place, Bostrom's approach understandably results in his providing a very brief, general discussion of the motivational, ethical and legal issues, whereas my approach lays the foundation for a detailed, pragmatic-analysis.
8. Center for Democracy and Technology 2006, p.2.
9. Bostrom 2003, p. 9.
10. Id.

11. Krupa 1996.
12. McLuhan 1964.
13. Castronova 2005, p. 252.
14. For example, the Serious Games Project at the Woodrow Wilson International Center for Scholars which has modified Maxis' *Simcity*, the Democracy Design Workshop at New York Law School which has leased an island in *Second Life*, and America's Army which the U.S. Defense Dept. created from "middleware" which is a template sold by developers of online games.
15. Rees 2003.
16. Emmanuel, Wendler and Grady 2000, p. 2701.
17. e.g. the Nuremberg Code and the Declaration of Helsinki.
18. Boym 2001, p.41.
19. An example of augmented reality would be eyeglasses containing a heads-up display (HUD) which superimposes personalized information over the view of the real world, to assist the wearer to better understand that world. This system could be transposed into a historical simulation, as it already is in some MMORPG's, e.g. *Second Life*. http://secondlife.com/newsletter/2006_02_15/hud.php.
20. e.g. Litton and Miller 2005,p.571.
21. Dawkins 1989, p.186.
22. U.S. Code of Federal Regulations 2005, Title 45, s.46.3.
23. Castronova 2005, p. 97.
24. Litton and Miller 2005, p.572.
25. Milgram 1963, p.371.
26. Herrera 2001, n.17.
27. Castronova 2001. Tables 1, 3. Jenkins 2004, p.11.
28. On October 1, 2005, the Chinese government instituted fatigue requirements for online games such as *World of Warcraft* and *Legend of Mir2*, to reduce the amount of time gamers spend playing these games. (Subsequently, in January, 2006, these requirements were modified to exempt adults.) The system only awards players full experience points for the first three hours of each day, half experience for the next two hours, and no experience after five hours. This was introduced in response to reports of actual, real-life deaths of players due to physical exhaustion, which obviously is not a concern for AI.
29. Exiting a MMORPG permanently after hundreds or thousands of cumulative hours of play has been compared to a real world suicide, indicating that such a step would be even more difficult to take in a totally immersive historical simulation. See www.joystiq.com/2006/03/09/drakedogs-suicide-now-on-google-video/.
30. Edwards 2005, p. 114.
31. Sunstein 2000, p. 19.
32. Rothblatt 2003, Gray 2002, p.23. (McNally and Inayatullah 1988) .
33. Kurzweil 2005, p.375.
34. Rothblatt 2003.
35. U.S. Code of Federal Regulations 2005, Title 45, sec. 46.3.
36. Diaz v. Hillsborough County Hospital Authority, 1996 U.S. Dist. LEXIS 11913.
37. Hanlon and Shapiro 2002, p.3.
38. The elements of a prima facie case for the tort of intentional infliction of emotional distress are as follows: 1) extreme and outrageous conduct by the defendant with the intention of causing, or reckless disregard of the probability of causing, emotional distress; 2) the

- plaintiff's suffering severe or extreme emotional distress; and 3) actual and proximate causation of the emotional distress by the defendant's outrageous conduct. Generally, the courts have found that for the defendant to be liable, its conduct must have been so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious, and utterly intolerable in a civilized community. *Flynn v. Higham*, 149 Cal. App. 3d 977 (1983).
39. *Deleese Williams v. American Broadcasting Corporation et al*, Superior Court of the State of California (Statement of Claim dated Sept. 21, 2005).
 40. e.g. "Wife Swap" where two husbands exchange wives for a temporary period. The producers of this show were sued by a husband participant who was upset that his wife was exchanged for one of the male partners in a gay couple.
 41. *County of Los Angeles v. Superior Court of Los Angeles County* (Kim A. Schonert) Supreme Court of Los Angeles, No. BC0908848, Aug. 12, 1999.
 42. Cohen 2003, p.32.
 43. For example, a German court recently ordered the shutdown of the German Wikipedia site on the grounds that it infringed on the privacy and anonymity rights of a deceased hacker known as Tron, by mentioning his real name. Boris (Family name), Case-No. 209 C 1015/05, Berlin Municipal Court, Dec. 14, 2005. <http://service.spiegel.de/cache/international/0,1518,396307,00.html>. In National Archives and Records Administration v. Favish, 124 U.S. 1570 (2004), the U.S. Supreme Court ruled that photographs of the dead body of Vince Foster, the White House Counsel who committed suicide, should not be disclosed, in order to protect the privacy of his family.
 44. For example, Anime Music Videos <http://www.animemusicvideos.org/home/home.php>.
 45. *Metro-Goldwyn-Mayer Studios v. Grokster Ltd.* 545 U.S. 1 (2005).
 46. Lessig 2005, p. 46.
 47. But see the recent case on fair use, *Blake A. Field v. Google Inc.*, U.S. Dist. Ct. (Nevada) Jan. 12, 2006, which may give some support to Lessig's view.
 48. <http://multiverse.net/>.
 49. Bostrom 2003, p.10.
 50. Hanson 2001, p.3.
 51. Lloyd 2006, p. 166.
 52. Fukuyama 1992.
 53. Kurzweil 2005, p.29.

References

- Baudrillard, Jean.1994. *Simulacra and Simulation*. Ann Arbor: University of Michigan Press
- Blake A. Field v. Google Inc. 2006. Jan. 12. U.S. Dist. Ct. Nevada.
- Boris (Family name). Case-No. 209 C 1015/05. Berlin Municipal Court. Dec. 14, 2005.
- Bostrom, Nick. 2003. "Are You Living in a Computer Simulation?" *Philosophical Quarterly*. Vol. 53(211): 243-255
- _____. 2006. "The Simulation Argument FAQ, Version 1.1." Accessed at www.simulation-argument.com/faq.html.
- Boym, Svetlana. 2001. *The Future of*

- Nostalgia*. New York: Basic Books.
- Castronova, Edward. 2001. "Virtual Worlds: A First-Hand Account of Market and Society on the Cyberian Frontier." *CESifo Working Paper Series*. No. 618.
- _____. 2005. *Synthetic Worlds – the Business and Culture of Online Games*. Chicago: University of Chicago Press
- Center for Democracy and Technology. 2006. "Digital Search and Seizure: Updating Privacy Protections to Keep Pace with Technology." Accessed at <http://www.cdt.org/publications/articles.php>.
- Cohen, Baruch C. 2003. "The Ethics of Using Data from Nazi Medical Experiments." *Jewish Law Articles*. Accessed at <http://www.jlaw.com/Articles/NaziMedEx.html>.
- Council for International Organizations of Medical Sciences. 2002. "International Ethical Guidelines for Biomedical Research Involving Human Subjects." Accessed at http://www.cioms.ch/frame_guidelines_nov_2002.htm.
- County of Los Angeles v. Superior Court of Los Angeles County (Kim A. Schonert). 1999. Aug. 12. Supreme Court of Los Angeles. No. BC0908848.
- Dawkins, Richard. 1989. *The Selfish Gene*. Oxford: Oxford University Press.
- Deleese Williams v. American Broadcasting Corporation et al. 2005. Sept. 21. Superior Court of the State of California.
- Descartes, Rene. 1644. *Meditations on First Philosophy*. Cambridge: Cambridge University Press.
- Diaz v. Hillsborough County Hospital Authority. 1996. U.S. Dist. LEXIS 11913.
- Edwards, Sarah J.L. 2005. "Research Participation and the Right to Withdraw." *Bioethics*. Vol. 19(2): 112.
- Emmanuel, E.J., Wendler, D., and C. Grady. 2000. "What makes clinical research ethical?" *Journal of the American Medical Association*. Vol. 283: 2701-2711.
- Flynn v. Higham. 1983. 149 Cal. App. 3d 977.
- Fukuyama, Francis. 1992. *The End of History and the Last Man*. New York: Free Press.
- Gray, Chris. 2002. *Cyborg Citizen*. New York: Routledge.
- Hanlon, Stephen and Shapiro, Robyn. 2002. "Ethical Issues in Biomedical Research: Diaz v. Hillsborough County Hospital Authority." *ABA Human Rights Magazine*. Winter 2002.
- Hanson, Robin. 2001. "How to Live in a Simulation" *Journal of Evolution and Technology*. Vol. 7(1).
- Herrera, C.D. 2001. "Informed Consent vs. Incomplete Disclosure in Social Science." *USC Bioethics Club*. Accessed at <http://www-hsc.usc.edu/~mbernste/BJC.4.html>.
- Jenkins, Peter S. 2004. "The Virtual World as a Company Town – Freedom of Speech in Massively Multiple Online Role-Playing Games." *Journal of Internet Law*. Vol. 8(1).
- Krupa, Frederique. 1996. "The History of Design from the Enlightenment to the Industrial Revolution." *Parsons School of Design*. Accessed at www.translucency.com/frede/hod.html.

- Kurzweil, Ray. 1999. *The Age of Spiritual Machines*. New York: Penguin
- _____. 2005. *The Singularity is Near*. New York: Penguin.
- Lessig, Lawrence. 2005. *Free Culture*. New York: Penguin.
- Litton, Paul and Miller, Franklin. 2005. "A Normative Justification for Distinguishing the Ethics of Clinical Research from the Ethics of Medical Care" *The Journal of Law, Medicine and Ethics*. No.4: 566-573.
- Lloyd, Seth. 2006. *Programming the Universe*. New York: Knopf.
- McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New York: Signet.
- McNally, Phil and Inayatullah, Sohail. 1988. "The Rights of Robots: Technology, Law and Culture in the 21st Century." *Futures*. Vol. 20(2): 119-136. Accessed at <http://www.kurzweilai.net/meme/frame.html?main=/articles/art0266.html>.
- Metro-Goldwyn-Mayer Studios v. Grokster Ltd. 2005. 545 U.S. 1.
- Milgram, Stanley. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*. Vol. 67: 371-378.
- National Archives and Records Administration v. Favish. 2004. 124 U.S. 1570.
- Plato. *The Republic*. New York: Dover Publications.
- Rees, Martin. 2003. *Our Final Hour - A Scientist's Warning How Terror, Error and Environmental Disaster Threaten Humankind's Future in This Century - On Earth and Beyond*. New York: Basic Books.
- Rothblatt, Martine. 2003. "Biocyberethics - should we stop a company from unplugging an intelligent computer?" International Bar Association Conference. San Francisco. September 16, 2003. Accessed at <http://www.kurzweilai.net/meme/frame.html?main=/articles/art0594.html>.
- Sunstein, Cass (ed.). 2000. *Behavioral Law & Economics*. New York: Cambridge University Press.
- "The Nuremberg Code." 1947. Reprinted in the *British Medical Journal*. Vol. 7070 (313): 1448.
- U.S. Code of Federal Regulations. 2005. Title 45 Public Welfare, Department of Health and Human Services. Part 46 - Protection of Human Subjects.
- World Medical Association. 2000. "Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects." Reprinted in the *Journal of the American Medical Association*. Vol. 284: 3043-45.

