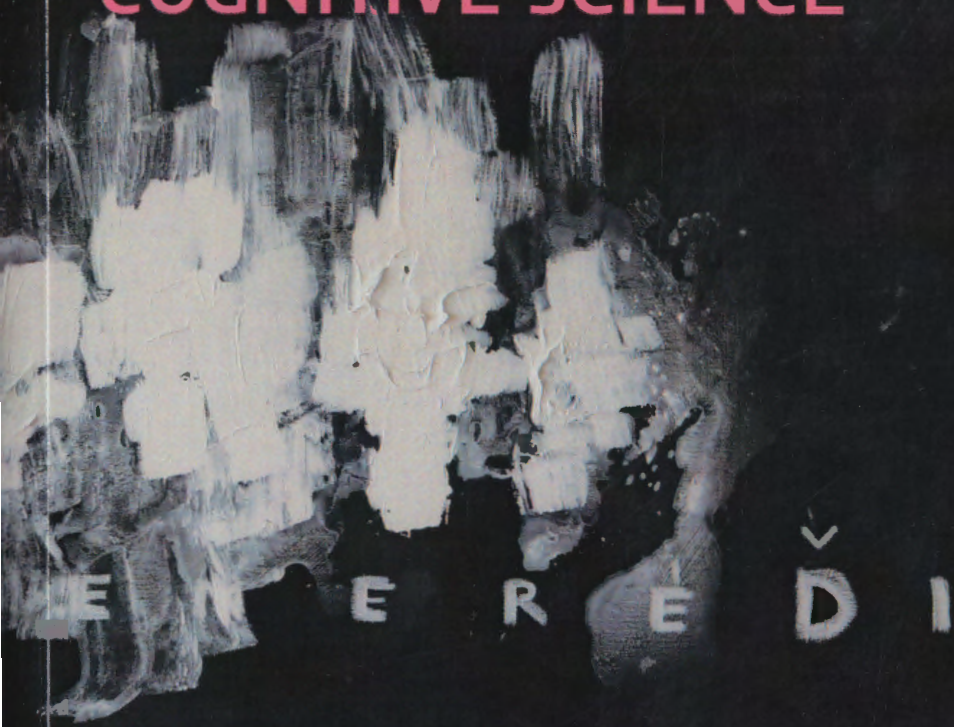


# CANADIAN UNDERGRADUATE JOURNAL OF COGNITIVE SCIENCE



E R E D I

$$+1) (\log R)^{m+1-j} + O(e^{-\delta \sqrt{\log R}})$$

**2013** *edition*

# CANADIAN UNDERGRADUATE JOURNAL OF COGNITIVE SCIENCE

2013 *edition*

by the Cognitive Science Student Society  
at Simon Fraser University

Simon Fraser University

Burnaby, British Columbia



# CANADIAN UNDERGRADUATE JOURNAL OF COGNITIVE SCIENCE

2013 *edition*

## EDITORS

Casey Haber  
*BA, Cognitive Science*  
Freya Olson  
*BA, Cognitive Science*

## REVIEW BOARD

Maite Taboada  
*Department of Linguistics, SFU*  
Keir Moulton  
*Department of Linguistics, SFU*  
John Alderete  
*Department of Linguistics, SFU*  
Tim Racine  
*Department of Psychology, SFU*  
Kathleen Slaney  
*Department of Psychology, SFU*  
Joanna Solomon  
*M.Sc., Applied Psychology*  
David Anderson  
*Department of Philosophy, SFU*  
Michael Picard  
*Cognitive Science Program, SFU*

## LAYOUT & DESIGN

Satt J. Quintanar

## CONTRIBUTING ARTISTS

**Cover art**  
*Green-Tao Theorem*  
with *Endre Szemerédi*,  
by Oliver Sin

**Inside art**  
Stefano Menicagli  
Eleanor Qu  
Oliver Sin  
Amanda Smart  
Xin Zhang

## SPECIAL THANKS TO

Maite Taboada  
Shamina Seranatne

The Canadian Undergraduate Journal of Cognitive Science is an electronic and print journal published by the Cognitive Science Student Society (CS<sup>3</sup>) at Simon Fraser University. Our aim is to provide a forum for students to share work among peers and gain valuable experience in the process of getting an academic paper published. As a publication, CUJCS provides a unique reference for students, showcases research by other undergraduate students, improves the contact and exchange of ideas between Canadian students and cognitive scientists alike, and illustrates the interdisciplinary work that is the hallmark of cognitive science. Although preference is given to Canadian students, contributions from students elsewhere are strongly encouraged.

## Submissions

Undergraduate students interested in contributing a paper to CUJCS should contact the editors for up-to-date submission guidelines: [cujcs-inquiries@sfu.ca](mailto:cujcs-inquiries@sfu.ca)

## Copyright Information

In keeping with the editorial policy of other cognitive science journals, we are of the opinion that papers submitted to our journal will usually be samples of works-in-progress. As such, acceptance in CUJCS should not impede publication of a later version in traditional journals. The copyright of all contributions remains with their authors. By submitting to CUJCS, authors acknowledge that submissions reflect original work, and that proper credit has been given to outside sources.

ISSN 1913-0651



simon fraser  
student society

CS<sup>3</sup>  
cognitive science student society  
simon fraser university

# Contents

- 5 **A case for mandating the disclosure of incidental findings when seeking informed consent**  
Victoria H. Saigle  
*McGill University*
- 19 **Transhumanism and personal identity**  
Arielle Friedman  
*University of British Columbia*
- 27 **Effect of cognitive dissonance on attitude change**  
Wook Yang  
*McGill University*
- 37 **Dating profile age analysis**  
Arielle Friedman, Rushil Singh & Ao Ao Feng  
*University of British Columbia*
- 47 **Sound changes of the /h/-phoneme from old english to modern english**  
Derrick Buchanan  
*Carleton University*
- 57 **Speaking, remembering and feeling the self**  
Mauro Silva  
*The University of Texas at Austin*

# **A case for mandating the disclosure of incidental findings when seeking informed consent:**

## **Ethical issues surrounding IFs in fMRI research on "healthy" populations**

Victoria H. Saigle

*McGill University, Montreal, QC, Canada*

Currently, there is no standardization across consent documents concerning the disclosure of incidental findings (IF) discovered during research on human participants. Cases in which protocols governing this information are absent or in which participants are given the option to be informed about IFs are unethical, as volunteers are unable to truly evaluate the potential risks that could arise from their participation prior to giving their consent. Using cases in which IFs are found in "healthy" research populations using functional magnetic resonance imaging (fMRI), ethical problems that arise from the current system are explored. IFs could pose a significant risk to an individual's health and it should be mandated that they be disclosed to all research participants.

*The ability to search for the truth implies also a duty; not to conceal any part of what one has found to be true.*  
— Albert Einstein (Milstein, 2008)

When conducting research, scientists are required to abide by a set of ethical guidelines that have been created to protect the rights of individuals participating in their experiments. One of the basic tenets of ethical human research mandates that informed consent—that is, agreement to participate in research after all



potential risks and benefits that may arise from taking part in the experiment have been disclosed (National Institute of Health [NIH], 1979)—must be obtained from each individual before a researcher can include them in his or her study. Since its implementation, however, informed consent has raised many ethical concerns ranging from the manner in which to obtain consent to the extent to which participants fully comprehend the consent document (Paris, 2010). With the growing use of neuroimaging technologies in research, additional questions regarding what information to include in consent documents have also gained attention from ethicists. The incorporation of functional magnetic resonance imaging (fMRI) in scientific research has greatly expanded our understanding of many brain functions as well as the brain's underlying physical structure (Iles, 2002). However, there are many reported instances in which fMRI scans have also uncovered findings in subjects that are not relevant to the question being investigated. Such unexpected discoveries are known as incidental findings (IF), where "incidental finding" is defined as "[...]a finding concerning an individual research participant that has potential health or reproductive importance [that] is discovered in the course of conducting research but is beyond the aims of the study" (Wolf, 2008). IFs in neuroimaging can range in severity from relatively benign, such as sinusitis, to much more serious, such as brain tumours or neurovascular abnormalities. Unfortunately, many consent documents do not provide clear information regarding how researchers will handle IFs despite their occurrence in approximately 34–47% of seemingly "healthy" participants (Iles, 2002; Shoemaker, 2011). Due to the fact that IFs discovered using fMRI are found in the brain these IFs could potentially impact an individual's decision-making skills. As such, an examination of discussions surrounding their disclosure when seeking informed consent seems an apt way to highlight ethical issues as they exist in current research practice. In order to examine the issues presented, I will be considering cases involving healthy and competent adult volunteers, as rules governing the ability to give consent for minors and incompetent adults are more complex.

A thorough understanding of the nature of informed consent and the responsibilities that bind researchers to their participants is vital in order to understand the importance of always disclosing IFs to said participants. Following the crimes against humanity committed during World War II, the Nuremberg Code was created to ensure that individuals participating in scientific research are treated ethically (NIH, 1949). The code mandates that the rights of all research subjects be maintained by requiring their voluntary, uncoerced consent to participate in research before any experiment can begin (NIH, 1949). The Nuremberg Code states that all subjects have the right to withdraw from the experiment if they do

not feel they are able to continue. Furthermore, scientists must be prepared to stop the experiment if they believe that continuing experimentation on the participant would lead to any lasting injury (NIH, 1949). The ability to give informed consent entails that the individual be mentally competent enough to make an educated decision with regard to their participation in an experiment. This decision can only be made following consideration of all the potential benefits and consequences that could result from their involvement (NIH, 1949). In order to provide individuals the ability to truly give informed consent, researchers are required to disclose the research procedure, purpose of their experiment, anticipated risks and benefits, and alternative procedures (if applicable), as well as to inform the participants of their ability to withdraw from the study at any time (NIH, 1949).

A separate set of ethical guidelines, The Belmont Report, was created in part due to public outrage surrounding studies like the Tuskegee syphilis experiments (Beauchamp, 2008). In the Tuskegee syphilis experiments, researchers enrolled poor, African American men infected with syphilis and told them that they would be receiving treatment, while in reality very little help was provided to these individuals at all. In addition, this experiment continued after the discovery of a cure for syphilis (penicillin) as its true purpose was to track the progression of the disease up to and including death (Shavers, 2000). Clearly, the creation of this report was necessary, for previous "federal policies [that] relied on the discretion and good judgment of investigators to determine the conditions under which research should be conducted" (Beauchamp, 2008) had failed to protect research participants. The Belmont Report mandates that "[...] information about risks should never be withheld for the purpose of eliciting the cooperation of subjects" (NIH, 1979) and that "[...] in all cases of research involving incomplete disclosure, such research is justified only if it is clear that (1) incomplete disclosure is truly necessary to accomplish the goals of the research, (2) there are no undisclosed risks to subjects that are more than minimal, and (3) there is an adequate plan for debriefing subjects, when appropriate, and for dissemination of research results to them" (NIH, 1979). In summary, in order to conduct ethical research, participants must have all relevant information required to make an informed decision and, if certain information is to be withheld, it must not only be absolutely necessary to accomplish the goals of the research (providing the goal entails minimal harm to participants) but there must also be a plan to debrief the subjects at a later date.

In order for participants to give informed consent, researchers must provide full disclosure of all potential risks and benefits that could arise from participation in the study; however, given the nature of IFs, neither the researcher nor the participant can fully anticipate said risks and benefits. Empirical studies exploring



exploring the severity of fMRI-related IFs in “healthy” populations frequently classify IFs into five categories according to the urgency with which the individual needs to see a healthcare practitioner: (1) no abnormal findings, (2) no referral necessary, (3) routine referral needed (findings should be reported but do not require immediate medical attention), (4) urgent referral needed (within one week) and (5) immediate referral needed (within one to several days) (Shoemaker, 2011; Katzman, 1999). Various studies place the incidence rate of fMRI IFs between 34–47% of the total population of “healthy” participants in any given group (Illes, 2002; Shoemaker, 2011; Katzman, 1999). According to one study, in instances where IFs are discovered, 9.8% require routine referral and 4% require urgent referral (Milstein, 2008). While, on average, more IFs are found in elderly subjects, one study reported that, when incidental findings are found in young adults, around 75% of the cases require urgent care (Illes, 2004).

As previously stated, the researcher must continually evaluate the condition of their participant and stop the study should the research be deemed a risk to the participant’s health (NIH, 1949). While this normally applies to the research itself causing harm, participation in research could indirectly cause harm to the subject by exacerbating medical conditions caused by an undiagnosed IF. In cases where IFs require urgent or immediate referral, any increased amount of time between identifying a problem and seeking treatment could be incredibly detrimental to an individual’s health. If a researcher fails to examine fMRI scans for IFs or if they cannot communicate the issue to the participant because the individual indicated they did not wish to be informed about IFs, the participant’s continued involvement in the study without being notified that they require medical attention could worsen their condition. Additionally, if participants do not fully understand the purpose of disclosing IFs, they may neglect to mention symptoms caused by an IF to a healthcare practitioner (ex. headaches that are indicative of an underlying brain tumour) because the participants believe that the researcher would inform them if something were truly wrong. Furthermore, in cases where IFs are not properly investigated, the researcher may not be able to adequately inform the participant about the risks and benefits from continuing in the study, particularly in cases where the IF needs to be addressed. This contradicts the bioethical principle of beneficence, which is defined as treating people “[...] in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being” (NIH, 1949). By choosing not to examine IFs, the researchers are limiting the extent to which they are informed about their subjects’ health. In doing so, researchers are effectively ignoring their responsibility to ensure the participants’ overall well-being. While

this may seem like a strong claim to make, the current lack of guidelines governing IF disclosure means that this choice is possible and, therefore, the possibility for unethical research exists.

In order to provide some guidance concerning the ethical management of IFs, some consent documents provide participants with the option to indicate whether or not they wish to be informed about any unexpected results. While this method seemingly adheres to the principle of autonomy (which, loosely defined, is the right of individuals to self-govern their choices (Beauchamp & Childress, 2009)), I believe that including the option to opt out of receiving information related to IFs in a consent document is unethical and completely undermines the idea of informed consent. By providing people with this choice and respecting their decision, researchers could claim that they are respecting the participant’s autonomous decision regarding their healthcare preferences. However, this choice, rather than providing individuals with the information they need to make the more important choice of whether or not to take part in research, veils an opportunity for researchers to supersede ethical practice and merely gives participants the illusion that they are in control over their own healthcare needs. Regardless of their choice, studies looking at the efficacy of informed consent have consistently demonstrated that participants do not fully understand the information included in consent documents, regardless of how they are presented (Paris, 2010). In addition to this, people may feel pressured to say that they do not want to be informed of IFs due to outside forces (such as their inability to qualify for insurance premiums, pre-existing stressors in their daily lives, worries about increased medical costs, or their desire to please the researcher). By allowing participants to choose not to be informed of potentially life-threatening findings, researchers may go against their own ethical principles because they are not ensuring that all participants will receive sufficient levels of medical attention pertaining to their individual conditions. Finally, providing patients the opportunity to choose not to be told about IFs completely undermines the idea of informed consent given that it is supposed to be an ongoing process; while consent is only formally required at the beginning of each experiment, the participant reserves the right to change their mind at any time without penalty. By removing the ability of the subjects to know what could, during the course of the study, be going on within their own bodies, they are unable to truly evaluate both the risks to their health throughout the experiment and the consequences of participating.

Participants in scientific research should always be informed of IFs. I believe that developing a protocol to deal with how to disseminate this information will alleviate many worries that subjects may have regarding potential financial



or personal implications of discovering an IF. In order to protect the patient's autonomy, discovery of the IFs should only be reported to the patient and not to any third party source such as a primary care physician or an insurance company. Giving the information directly to a third party eliminates the research participants' control over their own personal information and compromises their privacy (Wolf, 2008). A clear, standardized method of dealing with such personal information, such as mandating the presentation of IFs to all research participants, may remove some external pressures influencing the participant's decision.

Several studies have made it abundantly clear that participants do not necessarily understand the nature of informed consent, nor do they always fully comprehend to what they are agreeing when they sign the consent document (Paris, 2010). Although there is a large degree of variability between consent documents even within the same type of study (Palmour, 2011), it is widely accepted that all information included should be presented in a clear and organized fashion that uses as little jargon as possible (NIH, 1949). Ideally, the level of language in which the material is presented should match the education level of each individual (Paris, 2010). In fact, studies have found no correlation between level of comprehension and age, race, gender, social status, nationality or source of employment – the only factor that has been shown to have a strong effect is the participant's level of education (Paris, 2010). Unfortunately, studies of overall comprehension and retention of information presented in the consent document indicate that only 59% of individuals could identify the purpose of the research and that only 40% said that they had read the consent from carefully, despite the fact that roughly 85% of subjects believe that they understood all or most of the information (Cassileth, 1980). An equally worrying point regarding overall comprehension is that roughly 30% of volunteers believed that if they were handed a consent document, they had to sign it (Cassileth, 1980). This lack of understanding indicates that even if researchers provide the choice to receive information regarding IFs, there is an elevated risk that subjects will not understand the significance or potential repercussions of these discoveries. Furthermore, the fact that IFs found during fMRI research are discovered in the brain may mean the capacity of the individual to understand the consequences outlined in the consent document may be increasingly compromised. Another potential misunderstanding with regard to the optional disclosure of IFs stems from the spurious belief that clauses regarding IFs function to protect the researcher rather than the participants themselves (Cassileth, 1980). This misinterpretation could shift the participant's motivation for their decision from being participant-centred (ex. "what is best for my health?") to a more researcher-centred focus (ex. "is it

necessary to burden the researcher with extra work?"). As mentioned previously, offering optional disclosure of information regarding IFs cannot be said to be in line with respecting a patient's autonomy if they do not fully understand to what it is they are agreeing, as in cases in which they misinterpret the purpose for this choice.

The hallmark of obtaining informed consent from a human subject for participating in research is that it must be voluntary—that is to say, it must be given in the absence of any coercion or outside influence (NIH, 1947). However, it is likely that in many cases, participants feel pressured to participate by the researcher or by other external pressures. Incentives such as receiving better grades, earlier paroles, or cash may influence volunteers to engage in research (Ingelfinger, 1972). Additionally, participants entrust their well-being and welfare to researchers and therefore researchers could unintentionally manipulate a subject's perception of the experiment due to the patient's trust and vulnerability (Wolf, 2008). Unfortunately, a lack of resources available to researchers to manage IFs may leave them ill-equipped to explain or identify potential problems. For example, many studies do not have a dedicated radiologist to analyze scans for IFs (Wolf, 2008). Given that the process of presenting information when seeking informed consent is not standardized (Cassileth, 1980), researchers could also neglect to indicate the risks associated with IF discovery. For example, a study of Canadian consent documents for MRI and fMRI research identified thirteen separate strategies for dealing with IFs including a few documents in which the discovery of IFs were described as a benefit (Palmour, 2011). Alternately, IFs could hypothetically be used to determine the eligibility of "healthy" subjects for inclusion in a study, regardless of whether or not the participants agree to be informed of the IFs found. Though it could be considered an unintended consequence, such actions could violate the ethical principle of non-maleficence, the 'do-no-harm' principle (NIH, 1979), because researchers could potentially cause future harm to the participant for their own gain.

Another ethical issue regarding the optional disclosure of IFs on consent documents pertains to the just treatment of each participant. For example, let us assume that there are two identical subjects taking part in a study about brain activation when listening to music. During the fMRI scan, researchers notice that each person has a tumour in the same location. Both participants require immediate care; however, one participant agreed to IF disclosure while the other did not. In cases where participants are able to choose whether or not to receive the IF results, some individuals will receive the required medical attention while others will not within the same study. This surely goes against the ethical principle



of justice. According to the Belmont Report, justice entails that everyone must be treated equally. As such, resources are allocated based on the following five criteria: “(1) to each person an equal share, (2) to each person according to individual need, (3) to each person according to individual effort, (4) to each person according to societal contribution, and (5) to each person according to merit” (NIH, 1979). Based on this definition of justice, researchers’ ability to withhold information that could drastically affect someone’s quality of life based on their decision to not be informed of IFs cannot be said to be ethical. This is especially true considering that, even if researchers were aware of all possible risks, as it has already been pointed out, many people do not adequately understand the consent documents. Maintaining the just treatment of all the patients can only be assured by providing each participant equal access to the same level of care. Inaction on the part of the researcher to alert their participants of potential health risks is not indicative of a responsible study. This is not to say that researchers should be responsible for prescribing treatment, for that would fall outside the realm of their qualifications; however, they should be held accountable for ensuring that participants in their studies are adequately informed about any health issues discovered. Not only does investigating IFs help ensure ethical practice, but it may further be to the researcher’s benefit since any undiagnosed conditions present in a “healthy” population could skew the final results of the study.

Some studies have already looked at the feasibility of having research MRI scans analyzed for IFs by a radiologist. Starting in 2004, the institutional review board (IRB) of the University of New Mexico Health Sciences Centre hired a radiologist to ensure that all MRI scans taken for research purposes were scanned for the presence of any IFs (Shoemaker, 2011). As part of this initiative, they also developed the Health Insurance Portability and Accountability Act that decreed all research participants must receive a copy of their MRI and the radiology report so that they can present it to their healthcare provider, if they so choose (Shoemaker, 2011). Out of the 4,447 participants involved in a study examining the efficacy of the IRB’s ruling, 34% had IFs and 19.5% of the total population required some form of referral (Shoemaker, 2011). A follow-up study revealed that the response to receiving information about IFs was overwhelmingly positive; 92% of the population said that they were happy to have received the information and 58% said that they would be willing to participate in studies again specifically because they received the radiology review (Shoemaker, 2011). The cost-effectiveness of the Health Insurance Portability and Accountability Act was also examined. The researchers calculated that the total cost of providing these scans was around \$60,000 a year; although, with an increased number of

participants and by automating the reporting and scanning system, the cost of each scan could be reduced significantly in the future. This study estimated the cost to be \$24 per scan when averaged over 2,500 scans a year (Shoemaker, 2011). Although this follow-up study was limited in the sense that the researchers did not speak to people who declined to participate in the study, the overall positive feedback from those who were informed about IFs indicates that many people believe that knowledge about potential problems matters more than the potential psychological distress they could suffer from this knowledge. The fact that 10% of all participants with IFs from this study sought out additional medical attention based on the information provided in their radiology reports (Shoemaker, 2011) also suggests that providing subjects with this knowledge will have overall long-term benefits to their health.

In Alan Milstein’s paper entitled “Research Malpractice and the Issue of Incidental Findings,” he poses the question “[...]what does medicine owe to the person who volunteers his or her body for medicine’s progress?” (Milstein, 2008). This is a very pertinent question to ask to all those opposed to the idea of disclosing incidental findings to all participants. Certainly, the idea of mandating that all IFs be disclosed to subjects in research raises many logistical issues regarding the resources allocated to each project including money, qualified personnel to properly examine fMRI scans for IFs, time constraints, and the potential for psychological distress in individuals who obtain a false-positive result. However, should these issues really stand in the way of disclosing information that could potentially save someone’s life? Identifying an unknown medical condition in the early stages can be just as beneficial as—if not more beneficial than—identifying one after the condition has progressed significantly. By making the disclosure of IFs mandatory in all research, future experiments would be designed keeping in mind any potential complications related to IFs. Furthermore, mandating that all researchers be mindful of IFs decreases the chances that data obtained from “healthy” patients (who may be abnormal as indicated by any IFs) will skew the experimental results. A continuing refusal to address the ethical issues raised by the current system of IF disclosure completely ignores the responsibility of researchers to look after the patient’s best interests. Not only is this aspect mandated in existing ethical codes, but it should be highly safeguarded because many of the advancements in medicine could not have been made without the altruistic actions of volunteers.

When discussing the inclusion of IFs in informed consent, it is important that the ethical implications of not disclosing IFs takes precedence over logistical constraints (such as those mentioned in the preceding paragraphs). As outlined



in the Nuremberg Code and the Belmont Report, obtaining informed consent to participate from all volunteers is absolutely crucial, but this can only occur if all pertinent information is presented to the potential subjects (NIH, 1949; NIH, 1979). Consent documents that omit IFs from being disclosed in the study and those that provide the option to choose whether to be notified of IFs are unethical, as neither guarantees that all participants have the required information to make an informed choice. This is especially true considering that studies show only 40% of participants believe that they have read the consent document carefully and roughly 59% do not understand or remember the purpose of the experiment (Cassileth, 1980). Furthermore, the option of choice in the consent process increases the likelihood that a participant's decision is influenced by outside pressures. By mandating that all participants be informed of any discovered IFs and thereby providing them with the choice to seek further medical attention, the autonomy and personal privacy of each subject is maintained. Demanding that this information be provided in all studies would also ensure that justice is maintained within these experiments, for each person would receive the same amount of medical treatment and individual attention based on their specific needs throughout the course of the experiment. As highlighted above, this idea has already been positively received during studies in which all MRI scans were assessed for IFs by a radiologist (Shoemaker, 2011). Finally, if the inclusion of information about IFs became standard in every experiment, logistical issues such as finding a radiologist to decipher the MRIs and the extra time required will become more commonplace and seem less daunting when designing experiments. The researcher should not be responsible for the treatment required as a result of an IF, however, it is the very least that the scientific community can do to look for potential health issues in participants. Including full disclosure of IFs as part of consent forms will increase the likelihood that more careful research is conducted, that each researcher is abiding by the ethical principles to which they are bound, and that the overall health of each participant is addressed. Many scientific advances cannot be made without the voluntary, altruistic involvement of individuals in research; it is the very least we can do to ensure their continued good health.

## REFERENCES

- Beauchamp, T. L., & Childress, J. F. (2009). Respect for autonomy. In *Principles of Biomedical Ethics*, (6th ed., chap 4, pp. 99-148). New York: Oxford University Press.
- Beauchamp, T. L. (2008). The Belmont Report. In *The Oxford Textbook of Clinical Research Ethics*. (chap 14, pp. 149-155). New York: Oxford University Press.

- Cassileth, B. R., Zupkis, R. V., Suttonsmith, K., & March, V. (1980). Informed consent—why are its goals imperfectly realized. *New England Journal of Medicine*, 302(16), 896-900.
- Illes, J., Rosen, A. C., Huang, L., Goldstein, R. A., Raffin, T. A., Swan, G., & Atlas, S. W. (2004). *Ethical consideration of incidental findings on adult brain MRI in research. Neurology*, 62(6), 888-890.
- Illes, J., Desmond, J. E., Huang, L. F., Raffin, T. A., & Atlas, S. W. (2002). Ethical and practical considerations in managing incidental findings in functional magnetic resonance imaging. *Brain and Cognition*, 50(3), 358-365.
- Ingelfin, F. J. (1972). Informed (but uneducated) consent. *New England Journal of Medicine*, 287(9), 465-466.
- Katzman, G. L., Dagher, A. P., & Patronas, N. J. (1999). Incidental findings on brain magnetic resonance imaging from 1000 asymptomatic volunteers. *Journal of the American Medical Association*, 282(1), 36-39.
- Milstein, A. C. (2008). Research malpractice and the issue of incidental findings. *Journal of Law Medicine & Ethics*, 36(2), 356-360.
- National Institute of Health. (1979). *The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research*. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Washington, DC: US GOP.
- National Institute of Health. (1949). *Regulations and Ethical Guidelines—Nuremberg Code*. Trials of war criminals before the Nuremberg Military Tribunals under Control Council Law No. 10 (Vol. 2, pp. 181-182), Washington, DC: US GPO.
- Paris, A., Brandt, C., Cornu, C., Maisson, P., Thalamas, C., & Cracowski, J. L. (2010). Informed consent document improvement does not increase patients' comprehension in biomedical research. *British Journal of Clinical Pharmacology*, 69(3), 231-237.
- Palmour, N., et al. (2011). Informed consent for MRI and fMRI research: Analysis of a sample of Canadian consent documents. *Bmc Medical Ethics*, 12.
- Shavers, V. L., Lynch, C. F., & Burmeister, L. F. (2000). Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. *J Natl Med Assoc*, 92(12), 563-572.
- Shoemaker, J. M., et al. (2011). A practical approach to incidental findings in neuroimaging research. *Neurology*, 77(24), 2123-2127.
- Wolf, S. M., et al. (2008). Managing incidental findings in human subjects research: Analysis and recommendations. *Journal of Law Medicine & Ethics*, 36(2), 219-248.





*your body is a beautiful vessel*

**Amanda Smart**



## Transhumanism and personal identity

Arielle Friedman

*University of British Columbia, Vancouver, BC, Canada*

**Transhumanism aims to integrate technological enhancements with human biology, and contemplates the challenge of uploading a human consciousness. Some philosophers believe that replacing organic features for digital ones will have a detrimental effect on personal identity, while others claim that such changes will have little or no effect. Patternism and reductionism are discussed as prominent transhumanist theories of mind that deal with the dramatic changes that one's sense of self may undertake through these processes. Animalism is similar, but emphasizes issues concerning the biological importance of self. Ethical considerations surrounding transhumanism are becoming ever more salient, as technology brings us closer to a transhumanist world.**

Transhumanism is a movement which claims that in the future humans can, will and ought to merge with technology in various ways, one of which is digital enhancements that would transform humans into cyborgs. Putting aside technological, and for the moment, ethical questions, this paper investigates the metaphysical question of whether personal identity would be preserved through these alterations. To answer this, we will have to look at the implications of transhumanism on theories of personal identity.

What makes up personal identity? This depends on which theory of the mind we ascribe to. Dualists argue that we have an immaterial soul that constitutes our basic essence, which is compatible with transhumanism, as it implies that we can change the physical composition of our bodies and retain our identities. Materialists argue that it is the physical composition of our bodies and brains that make up who we are. In this respect, transhumanism is doomed—if our bodies are



the seat of our identity, who would willingly give theirs up in exchange for a digital death? Fortunately for transhumanists we now know that the atoms and molecules that make up our bodies are constantly being replaced, i.e., we are not made up of the same matter throughout our lives. Yet, despite our lack of physical continuity, we tend to feel that our identities remain consistent over time. So, if our identity doesn't reside in our atoms and molecules, then what defines who we are?

Animalists maintain a philosophical position that humankind fundamentally identifies as animals—our bodies and biological processes wholly constitute our identity. Unlike materialists who locate identity within the physical particles that make up our bodies, animalists propose that even as those particles change, our identities remain consistent so long as we remain the same biological animals that we were before. Although this view might seem antithetical to transhumanism, animalists argue that the theory is consistent with certain forms of transhumanist modifications. However, this view is still not popular with transhumanists, who tend to support one of two general philosophies of identity: patternism and reductionism.

Patternists claim that our identities come from our mental and psychological patterns. Reductionists claim that a person's existence is nothing more than a body, a set of thoughts, and a sequence of experiences and that identity is nothing more than a linguistic construct we use to group these elements. This paper will discuss the prospects for transhumanism through the lenses of animalism, patternism and reductionism. We will focus on the transhumanist procedures of uploading, the process of transforming one's mind into a computer program, and enhancement, i.e., the process of morphing parts of one's brain and body into a digitized form.

Most people agree that humans are part of the animal kingdom, but this does not necessarily imply animalism. A dualist, for instance, might claim that people are composed of a human animal as well as an immaterial soul. Patternists might claim that humans are animals, but that identity is comprised of psychological patterns that tend to occur in the context of that physical animal. For an animalist, identity begins and ends with the animal organism (Olson, 2003). When the animal dies, the person dies. According to Eric Olson, animalism precludes the possibility of transhumanism. If you were to have your cerebrum removed and placed in the head of another organism, you would not become the other organism; there would be two organisms that would have had their parts shuffled. The original (now brain-damaged) organism would be more you than the animal now carrying your cerebrum. In holding with this view, we must accept that by becoming machines we are losing the basic constitutive element of our identity—our animal bodies. Liao (2010) has a different view of animalism, which he calls the “organism view”.

Liao believes that even as organisms that continuously regulate and coordinate metabolic and life processes, we could still successfully merge with technology. Imagine a scenario where an individual's brain and body are replaced particle by particle with a digital circuitry, without any interruption in conscious experience. At what point do we consider this individual to be different. According to Liao, we do not. It was previously established that an organism's status is defined by its ability to regulation and coordination of life processes, which are conditions that have not been breached. Therefore, we have not altered the status of the organism. This process is analogous to the gradual cell replacement that biological organisms experience in their lifetimes. The organism view, however, only takes us halfway to the transhumanist dream. Unlike enhancement, uploading requires that we forgo our physical bodies entirely and transform ourselves into data within a digital system. Uploading locates identity fundamentally with our mental experience; to be uploaded, we must be the data, not the physical system which stores this data. This understanding of identity contradicts even the organism view, which ties identity so closely to the body and its metabolic processes. To consider the possibility of uploading we need to explore other theories of identity.

Patternism is the view that our identity is composed of mental, psychological or physical patterns. The meaning of “patterns” may vary from memories to psychological continuity to the mechanisms of electrical neural activity. Some patternists liken our minds to computer software which operates on the physical “hardware” of our brains. According to patternism, both uploading and enhancement will preserve personal identity (Liao, 2010). Even though brain and body in their biological form would be lost, the pattern that makes up ones identity would continue. Susan Schneider (2009) showcases a critical weakness in patternism by pointing to the example of Jake Sullivan from Robert Sawyer's novel *Mindscan*. Jake is terminally ill and elects to have his mind, memories and personality—his pattern—uploaded to an android who would continue Jake's life. Jake wakes up from the procedure to find that he is still his original sick self and has signed his life away to an android. The robot may share Jake's pattern, but the robot is not Jake. Only Jake is Jake. This illustrates the reduplication problem, which is to say that an identity can only be located within one individual, and while an exact copy may appear identical to that person, the copy will not be that person. This demonstrates that psychological continuity is a necessary but not sufficient for personal identity. This poses a fatal challenge to patternism. Schneider suggests an additional condition that, in cooperation with psychological continuity, will be sufficient to support patternism. She suggests that spatial-temporal continuity is an essential feature to identity survival. As much as our bodies may change, our



minds seem to move through the spatial world in a continuous “line” through time. We can’t just transport our consciousness to another part of the world, as uploading would involve. If a procedure causes our minds to “jump” to a physical substrate separated by time or space from the original, we should be suspicious as to whether this procedure would preserve our identity.

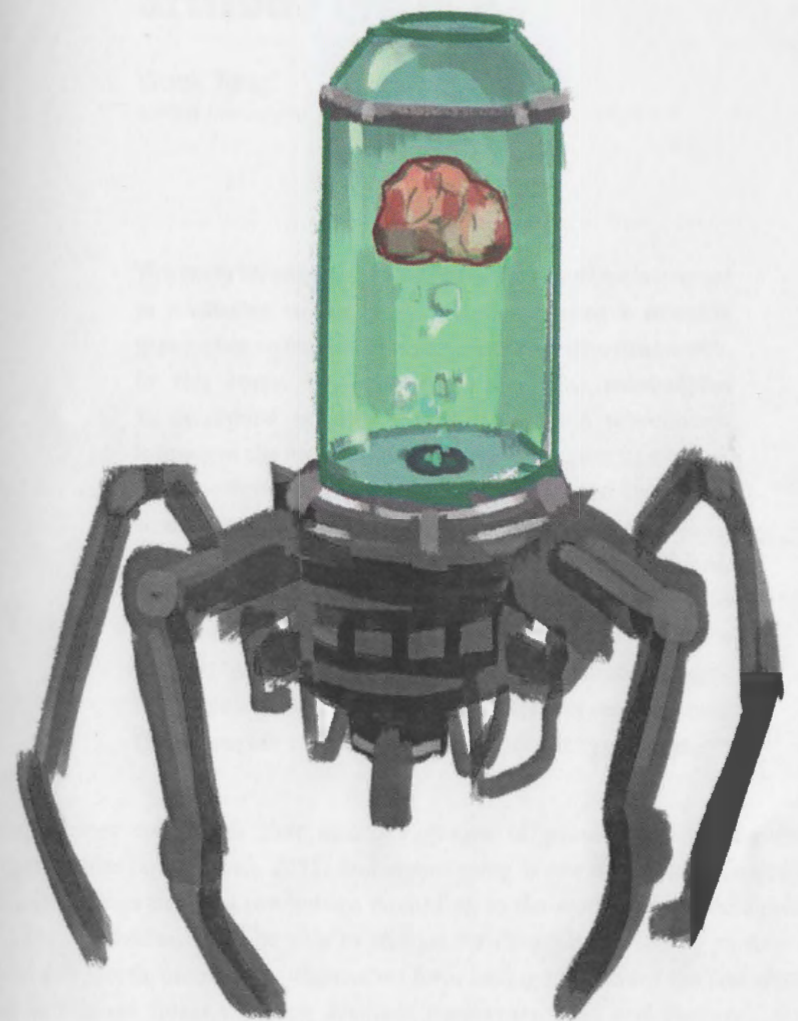
Derek Parfit (1995) argues that according to a reductionist viewpoint, questions about uploading and enhancements reduce to questions about language. Parfit considers reductionism to be the only viable alternative to dualism. According to his version of reductionism, our identities are nothing more than linguistic devices that we use to identify a body and set of experiences over time. He likens personal identity to a nation which is constituted by the people within it and their culture. The “nation” itself is merely a catch-all; it has no meaning beyond the particulars that make it up. In much the same way, identity doesn’t exist beyond the physical and mental events of a person’s life. If we notice our body or psychological pattern changing, then we already know everything we need to know. Further questions about identity relate only to language. According to Parfit, after undergoing uploading or enhancements it would make more sense to ask, “Should I still call myself the same person?” than to ask, “Am I still the same person?” Different answers to the former question represent different frameworks through which to view the world. Parfit goes on to argue that when one concept is constituted by other concepts, we must base our decisions on the constituent concepts (or as he puts it, “arguing from below”). Parfit claims that it is these constituent concepts that provide us with the most direct and relevant information about the world. When making the decision to alter our bodies or minds, the question should be will the outcome be positive overall and will the resulting individual contain what is fundamentally “me”.

Transhumanism as a growing movement will become more prominent as our lives and bodies continue to overlap and interact with technology. We may be given the options of enhancing or uploading even within our own lifetimes. If this happens, we need to make our decision based on careful philosophical consideration of what this would mean for our identities. Olson, Liao, Schneider and Parfit all provide invaluable perspectives for our deliberation. Olson and Liao, in discussing possible consequences of animalism. Schneider, in problematizing the easy answers of patternism, and Parfit, in providing a radically different lens through which to view who we are.

## REFERENCES

- Parfit, D. (1995). The unimportance of identity. In H. Harris (Ed.), *Identity: Essays based on Herbert Spencer lectures given in the University of Oxford* (pp. 13–45). Oxford: Clarendon Press.
- Liao, S. M. (2010). Twinning, inorganic replacement, and the organism view. *Ratio*, 23(1), 59–72. doi:10.1111/j.1467-9329.2009.00450.x
- Olson, E. T. (2003). An argument for animalism. In Martin, R. and Barresi, J., (Eds.), *Personal identity* (pp. 318–334). Oxford: Blackwell. Retrieved from <http://eprints.whiterose.ac.uk/734/1/olsonet5.pdf>
- Schneider, S. (2009). Mindscan: Transcending and enhancing the human brain. In S. Schneider (Ed.), *Science fiction and philosophy: From time travel to superintelligence* (pp. 241–256). Oxford: Blackwell. Retrieved from <http://www.sas.upenn.edu/~sls/documents/Schneider-Mindscanproofs.pdf>



*brainbot***Xin Zhang**



## Effect of cognitive dissonance on attitude change

Wook Yang

*McGill University, Montreal, QC, Canada*

This study investigates the role of self-concept maintenance in mediating stereotyping behaviors toward a minority group after an individual experiences cognitive dissonance. In this study, cognitive dissonance was manipulated to determine whether this threat to the self-concept influences the degree of stereotypical behavior as a means to self-affirm. Statistical analysis determined that there were no significant differences on the degree of stereotyping between the dissonance and non-dissonance conditions. In addition, there were two marginally significant main effects for the effect of cognitive dissonance on attitude change, and degree of stereotyping on attitude change, but there was no significant interaction between the two. Therefore, our results failed to support our hypotheses.

Stereotypes are beliefs that associate groups of people with certain traits (Baumeister & Bushman, 2011) and stereotyping is one of the most frequently studied topics in social psychology. According to the work of Fein and Spencer (1997), individuals may be able to reclaim for themselves a feeling of mastery and self-worth, often saving themselves from having to confront the real sources of self-image threat by using available stereotypes. Fein and Spencer's study demonstrated that facilitating self-affirmation—i.e., providing people with information that restores their positive sense of self-integrity—can reduce the likelihood that individuals will derogate members of stereotyped groups. Aronson (1969) argued that a dissonance-provoking act should cause more consistency-restoring self-justification because the act would be more consistent with a positive than a negative self-concept. In contrast, Steele and his colleagues (1993) assumed



that following a self-image threat, people are not concerned with consistency, but with restoring a general image of self-integrity (Steele, Spencer & Lynch, 1993).

Cognitive dissonance theory maintains that inconsistencies among a person's thoughts, sentiments, and actions create an aversive emotional state that leads to efforts to restore consistency (Gilovich, Keltner, Chen & Nisbett, 2013). In the current study, cognitive dissonance was conceived as a threat to the self-concept and stereotyping behaviors as a self-affirmation process. The purpose of the study was to investigate whether threats to the self-concept will also cause an increase in stereotyping behavior and whether this would, in turn, have an influence on total attitude change within a certain topic such as tuition increase. The first hypothesis was that the participants in the cognitive dissonance condition would show a higher level of stereotyping behavior as compared to the participants in non-cognitive dissonance condition. The second hypothesis was that the participants in the cognitive dissonance condition who showed a high level of stereotyping behavior would be less likely to show a total attitude change regarding a certain topic such as tuition increase compared to the participants who showed a low level of stereotyping behavior.

## Method

### Participants

Thirty-seven undergraduate students and recent graduates from Quebec universities participated in this experiment: 33 women and 4 men aged from 19 to 25 ( $M = 21.60$ ,  $SD = 1.36$ ). There were 12 participants in cognitive dissonance condition and 19 participants in non-cognitive dissonance condition ( $N = 31$ ). The imbalanced sample size in each condition is due to the high attrition rate. Out of 37 participants, 4 participants showed a neutral attitude towards tuition increase and 2 participants did not fill out the questionnaire properly, so their data was not included in subsequent analyses. Participants were recruited through McGill University's subject participant pool in order to be eligible for extra credits for their psychology courses.

### Procedure

The experiment consisted of two parts. First, all participants completed the same pre-screening questionnaire. The pre-screening questionnaire elicited demographic information (age, gender, and education level) and participants' initial self-esteem using the ten items from the Rosenberg Self-esteem Scale (Rosenberg, 1965). Although the pre-screening questionnaire included demographic and self-esteem measure questions, the main interest was to find out what the students'

attitude towards tuition increase was. The questionnaire items included: "Students should be mostly responsible for paying their own education," "A small increase in tuition will not prevent anyone from attending university if they want to," "I oppose any increase in tuition," "Education should be free," "The participants rated their attitude using 13-point Likert scale ranging from -6 (*extremely disagree*) to 6 (*extremely agree*).

Two to three days after filling out the online pre-screening questionnaire, the participants were contacted by e-mail with customized link to either cognitive dissonance and stereotype implied condition, cognitive dissonance and non-stereotype implied condition, non-cognitive dissonance and stereotype implied condition, or non-cognitive dissonance and non-stereotype implied condition. To examine the effects of cognitive dissonance, depending on their condition, participants were either asked to write counter-attitudinal sentences on tuition increase or attitude-consistent sentences on tuition increase. To examine the effects of stereotyping behavior, after the sentence writing task, participants were asked to read a short story about a 31-year-old struggling actor Tom (described below). Then, the participants were presented with a list of adjectives for them to rate Tom's personality. Finally, to measure attitude change, participants were asked to answer the same questions about their attitude towards tuition increase that were presented in the pre-screening questionnaire.

### Manipulations

The manipulation of story character's apparent sexual orientation that was described by the article of Fein and Spencer (1997) was used. All participants read about a character called Tom, a 31-year-old struggling actor living in the East Village in New York City. The information summarized Tom's ambitions and career struggles and listed some of the many odd jobs that Tom had taken to pay the rent while he pursued his dream. The excerpt continued by detailing a recent event in Tom's life concerning landing "a fairly large part in a serious and rather controversial play directed by a young director." Participants read that Tom was excited about the play and, in particular, about working with this young director. The director's name was not mentioned, but gender pronouns indicated that the director was a man. The participants read that after first week of rehearsals, Tom approached the director and asked him whether he wanted to get "a drink or something" with him after that night's rehearsal so that they could talk about his role in some more depth. The information about Tom was identical across conditions with the following exceptions. In the first sentence, the participants in the non-stereotype implied condition read that Tom "had been living with his



girlfriend, Lisa, in a small apartment" for several years. Lisa's name was mentioned three more times subsequently in the story, and there was one additional reference to his "girlfriend." For the stereotype implied condition, in the first sentence the word "girlfriend" was replaced with "partner" and all references to "Lisa" were dropped from the story. Neither the partner's name nor the partner's gender was specified, and there were no subsequent references to this partner (Fein & Spencer, 1997). After reading the story, participants used a 7-point scale ranging from 0 (*not at all*) to 6 (*extremely*) to rate Tom's personality. The stereotype-relevant traits included: sensitive; assertive; aggressive; considerate; feminine; strong; creative; and passive. Also, stereotype-irrelevant filters were included: intelligent; funny; boring (Fein, Cross, & Spencer, 1995).

### Measures

The internal consistency of social context attitude measure of the pre-screening measure indicated that the 4 items questionnaire ( $\alpha = .45$ ) had low internal consistency. However, the internal consistency of the same social context attitude measure of the part 2 indicated that the 4 items questionnaire ( $\alpha = .63$ ) showed increase in internal consistency.

### Scoring

Attitude change on tuition increase was measured by taking the raw scores from the post-test and subtracting the raw scores of the pre-screening questionnaire. This resulted in a possible range of -24 to +24, with scores around 0 indicating no overall change in attitude about the tuition increase.

### Results

With the data of 12 participants in cognitive dissonance condition and 19 participants in non-cognitive dissonance condition ( $N = 31$ ), the results showed that the mean for participants' degree of stereotyping after going through cognitive dissonance was lower ( $M = 26.67$ ,  $SD = 2.67$ ) than the mean for participants' degree of stereotyping after not going through cognitive dissonance ( $M = 27.42$ ,  $SD = 5.21$ ). However, independent samples t-test indicated that this mean difference was not statistically significant,  $t(29) = -.46$ ,  $p = .647$ . Therefore, the first hypothesis was not supported.

The second hypothesis was also not supported since there was no statistically significant interaction between the cognitive dissonance and the degree of stereotyping,  $F(1,26) = .46$ ,  $p = .504$ . However, the results showed that there was a marginally significant main effect of cognitive dissonance on total attitude change,  $F(1,26) = 3.36$ ,  $p = .078$ . The participants in non-cognitive dissonance

condition showed mostly no total attitude change ( $M = .00$ ,  $SD = 4.49$ ) whereas the participants in cognitive dissonance condition showed a high level of total attitude change ( $M = -5.82$ ,  $SD = 7.96$ ).

Also, a marginally statistically significant main effect of degree of stereotyping on total attitude change was found,  $F(1,26) = 3.71$ ,  $p = .065$ . The participants were divided into low degree of stereotyping and high degree of stereotyping by median split on their responses on judgment task. The participants who showed a low degree of stereotyping behavior showed a higher level of total attitude change ( $M = -4.35$ ,  $SD = 6.86$ ) than the participants who showed a high degree of stereotyping behavior ( $M = .77$ ,  $SD = 4.89$ ).

### Discussion

The purpose of this study was to investigate whether threats to the self-concept would cause increases in stereotyping behavior and whether this would have an influence on total attitude change within a certain topic such as tuition increase. This study used cognitive dissonance as a self-concept threat. The results showed that there was no statistically significant mean difference on the degree of stereotyping behavior between the cognitive dissonance and non-cognitive dissonance conditions; therefore, the first hypothesis was not supported. This result could be explained by social desirability bias, i.e., a tendency of some respondents to provide answers that are biased in the direction of socially acceptable attitudes or behaviors (Greenstein, 2006). Since the participants were university students, they may have not wanted to be recognized as stereotypical individuals. As a result, the participants who were in stereotype implied condition may have restrained from stereotyping the character (Tom) after reading the story. Furthermore, there was no statistically significant interaction between the effect of cognitive dissonance and degree of stereotyping behavior on total attitude change, which means that the second hypothesis was also not supported. However, a marginally statistically significant main effect of cognitive dissonance on total attitude change was found. The participants in cognitive dissonance condition showed a higher level of total attitude change than the participants in non-cognitive dissonance condition. The higher level of total attitude change of participants in cognitive dissonance condition indicated that writing the counter-attitudinal essay resulted in cognitive dissonance effect, which weakened their initial attitudes towards a tuition increase.

A marginally statistically significant main effect of degree of stereotyping on total attitude change was found as well. The participants who showed a low degree of stereotyping behavior showed a higher level of total attitude change than the



participants who showed a high degree of stereotyping behavior. This result could be explained by the fact that showing a higher degree of stereotyping behavior led to self-affirmation which would have solidified their self-concept. The participants who showed a low degree of stereotyping behavior did not have this solidification of self-concept, which would have led their initial attitudes towards a tuition increase weakened.

### Conclusion

Although a statistically significant result for interaction between the experience of cognitive dissonance and the degree of stereotyping was not found, the trend of the main effects turned out as I predicted (see Figure 1). There were a few limitations of this study. First was the small sample size. The attrition rate, the failure of some research participants to complete an investigation (Pelham & Blanton, 2013), was 65%. Even though this study managed to get 115 participants for part 1 of the study, the total usable data for part 2 were only from 37 participants. Most participants did not complete the part 2 of the study because participants were asked to write an essay regarding tuition increase. Some of the participants who completed the second part of the study ended up arguing according to their point of view even though they were meant to be in the cognitive dissonance condition. The second limitation was the low reliability, the consistency of different measurements of the same thing (Frederick & Wallnau, 2013), of the 8 items stereotype measure ( $\alpha = 0.196$ ). Although I failed to find support for the hypotheses on the effects of stereotyping as generating attitude change, I still believe this is an important topic with real-world implications, and if replicated properly with higher participant numbers and reliable measures, the current study may yield meaningful insights into the underlying mechanisms of stereotyping as a way of self-affirmation.

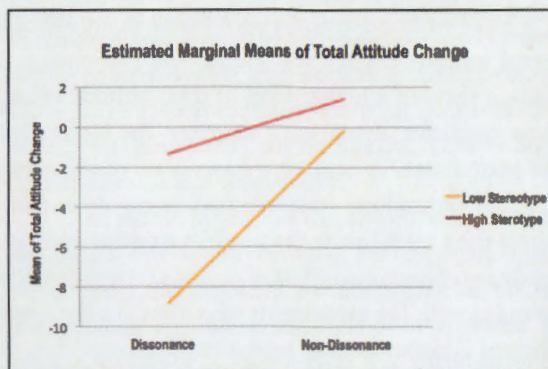
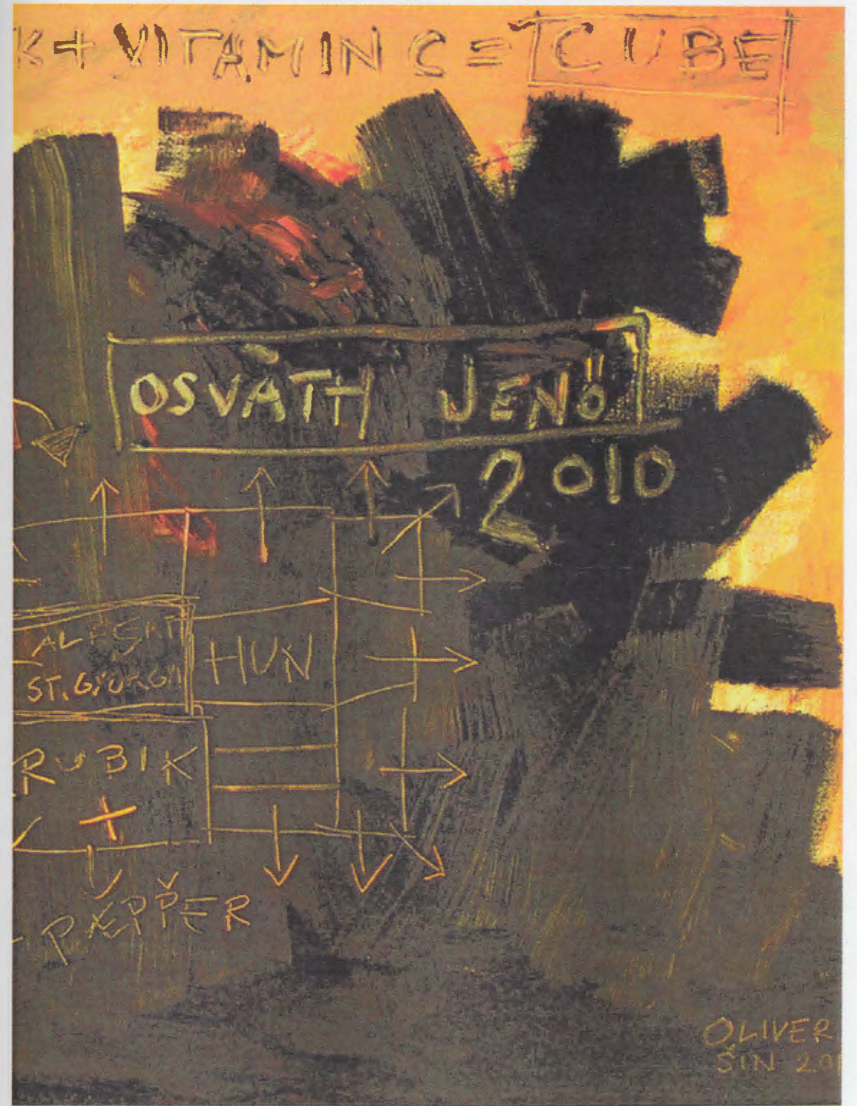


Figure 1

### REFERENCES

- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. *Advances in Experimental Social Psychology*, 4, 1-34.
- Baumeister, R. F., & Bushman, B. J. (2011). Prejudice and intergroup relations. *Social Psychology and Human Nature*, 393-396.
- Fein, S., & Spencer, S. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73(1), 31-44.
- Fein, S., Cross, J. A., & Spencer, S. J. (1995). *Self-esteem maintenance, stereotype consistency, and men's prejudice toward gays*. Paper presented at the 103rd Annual Convention of the American Psychological Association, New York.
- Frederick, G., & Wallnau, L. (2013). *Statistics for the behavioral sciences*. (9th ed., p. 222). California: Wadsworth, Cengage Learning.
- Gilovich, T., Keltner, D., Chen, S., & Nisbett, R. E. (2013). Attitudes, behavior, and rationalization. *Social Psychology*, 245-246.
- Greenstein, T. (2006). *Methods of family research*. (2nd ed., pp. 66-67). California: Sage Publications.
- Pelham, B. W., & Blanton, H. (2013). How do we misinterpret? Common threats to validity. In *Conducting research in psychology: measuring the weight of smoke* (4th ed., pp. 147-151). California: Wadsworth, Cengage Learning.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Steel, C. M., Spenser, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, 64, 885-896.





OLIVER  
SIN 2010



# Dating profile age analysis:

## A supervised learning application

Arielle Friedman, Rushil Singh and Ao Ao Feng

*University of British Columbia, Vancouver, BC, Canada*

We designed an agent that uses supervised machine learning and word frequency analysis to analyze online dating profiles. It classifies the vocabulary that people in different age groups tend to use in their profiles, and can successfully predict the age group of a dating profile's author. We programmed our agent in Java and used term frequency-inverse document frequency (TF-IDF) scores and multinomial logistic regression (MLR). This technique has important potential for market and research applications.

### Introduction

Many mobile phone applications, internet dating websites like *OkCupid*, and social media websites like *Facebook* rely on user-generated data to deliver customized services to their users. This data tends to be gathered through software applications, quizzes and other measures that require active user participation. However, information gleaned through self-surveys has many shortcomings; people aren't always aware of the subtleties of their behaviour, including what words they tend to use and why they use them.

Could user information be gathered using methods that are not subject to these shortcomings? One potential method could be through Natural Language Processing (NLP), which is the study of syntax, words, and grammars for the purpose of gathering data. NLP has been used in a wide variety of research applications for a broad range of purposes. We have used NLP to design a software application which gathers information about how people in different age groups use language to describe themselves on online dating sites. The application gathers demographic information that users may not be aware of including in their profiles. The application can also guess the age range of novel users, which



helps us test the application and provides an entertaining way to engage users.

We have decided to use a particular kind of NLP called word frequency analysis, which gathers information regarding word choices. This approach has some shortcomings; much of the meaning of language emerges from elements which are not included in word frequency analysis, like sentence structure, context, phrase structure, idioms, or compound words. For example, the compound “hot dog” loses all meaning if analyzed in terms of the words “hot” and “dog”. However, when performed on large quantities of data, word frequency analysis preserves semantic meaning for a statistically significant portion of the time, enough to provide useful information. Furthermore, word frequency analysis is easy for an artificially intelligent system to perform accurately on huge bodies of text whereas teaching an agent more complex language structures is more difficult and error-prone. We will look at a few examples where word frequency analysis was used successfully on a large data set.

Vongpumivitch et al. (2009) used word frequency analysis to determine how often applied linguistics researchers used words from the Academic Word List (AWL) in their publications. Academic language is often one of the biggest challenges that teachers and learners of English as a foreign language face, so this study has important pedagogical implications. The researchers found that AWL words accounted for 11.7% of all words used in the entire Applied Linguistics Research Corpus, a data source with over 1.5 million words. This study demonstrates the scale on which word frequency analysis can be used and its potential scope in gathering broad statistical data on huge amounts of textual information.

Baruch Vilensky (1996) used word frequency analysis to determine the relative word choice distance between authors. He found that an agent using word frequency analysis alone could successfully tell whether or not two books were written by the same author. This shows how word frequency analysis can be used to gather information about writers which may not be superficially obvious.

For our “Guess Your Age” dating profile application, we have decided to use supervised machine learning, TF-IDF scores, and MLR to carry out word frequency analysis. We will explain why we made those choices, and how we carried them out.

## Method

Our agent is designed to carry out supervised machine learning. This means that the agent undergoes a training phase where it is given the independent variables (the dating profile) and the dependent variables (the age category) for each data

point. It then undergoes a testing phase where it is given the independent variables (the dating profile) and has to predict the dependent variable (the age category). Supervised machine learning was chosen because it can be done on a relatively small data sets like ours, which included approximately 1000 data points (dating profiles).

We combine two techniques for the machine learning component of our project, TF-IDF, and MLR. TF-IDF scores calculate the relative frequency of each word in a document. The term frequency (TF) score is the frequency of each word in its particular document. This alone would not be enough—many words appear frequently simply because they are common English words, such as “and,” “I” and “in”. We are looking for words which not only appear frequently, but which appear frequently within the dating profile in question, relative to other dating profiles. We use inverse document frequency (IDF) scores for this reason; the IDF value is the proportion of profiles a word appears in. TF and IDF scores are multiplied together to produce a score for each word in each document which accurately reflects the relative frequency of that word in that document—how “significant” it is.

MLR allows us to predict the likelihood of a categorical dependent variable based on independent predictor data (generated from TF-IDF). MLR is a generalized form of logistic regression used when the dependent variable has more than two categories. MLR is what allows the agent to predict the age category of an unknown profile based on the training data it has processed.

One of the most common applications of TF-IDF is in web searches. Rahman et al. (2013) studied the usefulness of TF-IDF scores in dealing with topic drift in web searches. Topic drift is when the contents of a search result drift away from the topic in its heading. They designed a web-page ranking system which uses TF-IDF to improve the relevancy of search results, demonstrating the usefulness of TF-IDF scores in pruning irrelevant data.

The utility of TF-IDF is not limited to written text. Smith et al. (1997) used TF-IDF to design a video search program that reduces each video to a two-minute fragment to speed up search time. These fragments preserve relevant audio keywords while eliminating search-irrelevant information. This is accomplished using TF-IDF scores, demonstrating the power of this tool in a variety of different modalities.

Coyle et al. (2012) designed an agent that, much like our agent, performs word frequency analysis using TF-IDF scores and regression to analyze the effects of various recreational drugs based on user testimonials that people posted on the website *Erowid*. Using word frequency analysis, Coyle’s agent was able to



accurately determine which drug was associated with each testimonial. It could also identify words that were strongly associated with certain kinds of drugs, providing potential insights into those drugs' effects. This study demonstrates the power of these tools in gathering subtle information from what people write about themselves online.

For our data set, we used 900 profiles from *friendfinder.com*. We copied the text of the self-description section of each person's profile (labelled "Introduction" on the site) and recorded the person's age. We divided the profiles into three age groups and designed our program around those categories. We programmed our agent in Java.

During the training phase, the agent applies TF-IDF scores and MLR to a large body of training data to determine the weight of each word relative to each age category. During the testing phase, the agent uses these values to assign the profile being tested a "score" for each category, and selects the category with the highest score. The dating profile's writer most likely falls into this age range category.

We chose the age ranges twenty to twenty-nine, thirty to thirty-nine, and forty plus. We included 300 dating profiles from each category. The techniques we used are scalable, so we could increase the number of age groups for future versions. Initially we wanted the agent to guess the participant's exact age, but if we used our current approach to do so, we would risk over-determination. Over-determination is the risk of information loss when the categories are small and overlapping. If someone uses words that are common to a 28-year-old, they are more likely to be 29 than if they use words that are common to a 58-year-old. But if we use a categorical analysis with exact ages, that information is lost. In future versions, we could program our agent to guess participants' exact age using ordinary least squares regression, which takes a continuous dependent variable.

The use of categories also offsets the issue of how we can be sure that people are reporting their ages honestly. If we were using exact ages, it would be crucial to know that the ages were accurate, but since we are using categories, the ages can be approximate. It is irrelevant if people lie within a category, but if they lie across categories in a systematic way, it could influence the results. This may be taking place, since people may lie to make themselves appear to be in a particular decade (e.g., a 40-year-old might claim to be 39). In future versions, we could offset this by choosing age categories which do not correspond to decades.

### Calculating TF-IDF scores

TF-IDF calculates the relative frequency of each word in each profile. The TF

value equals the number of times the target word appears in that profile divided by the total number of words in that profile. The IDF value is the proportion of profiles a word appears in. The agent calculates this by dividing the total number of profiles by the number of profiles that the target word appears, and then take the tenth log of the result.

### Performing MLR

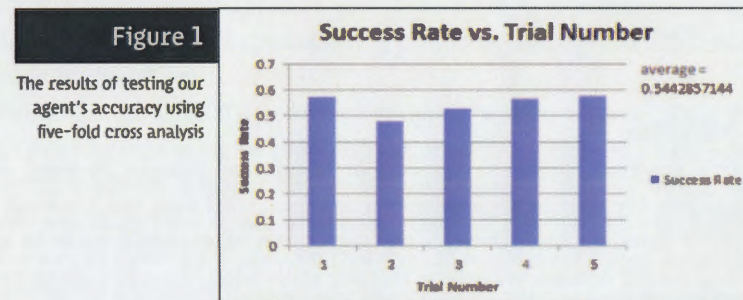
To perform MLR, the agent needs to have three sets of beta-values corresponding to each of the three age range categories. The beta-values for a category are what we have chosen to call the representation of the linear weight of each word relative to that category—a rough estimate of how likely that word is to indicate that category.

### Testing and Assignment

To determine the age of a dating profile's author, the agent calculates the profile's scores for each of the age categories. The score of a profile, for each category, is the dot product of the TF-IDF values for all the words in that profile, with the Beta-value of each word for that category. This yields a number between -1 and +1 for each profile for each category. The highest scoring age category is chosen as the most likely candidate for that profile.

### Evaluation

To test our agent's accuracy, we performed a five-fold cross analysis. We divided our data of 700 dating profiles into five equal parts of 140 profiles each, all of which contained some profiles from each of the three categories. We cycled through using each 1/5th as our testing example, and training on the remaining 4/5th, or 560 profiles (see figure 1). Our agent was accurate ~54.4% of the time. Since each profile could be classified into three possible categories, we would expect





our program to have 33.3% average accuracy by chance, with a standard error of roughly 4% (on 140 test cases). Our program's accuracy was well above the margin of error of what would be expected by chance.

## Results

Our agent can tell us which words corresponded to each age group most closely. Common words in the twenty to twenty-nine year-old age group included "studying," "dude" and "gay." For those in the thirty to thirty-nine year-old age group, "therapy," "fantasies," "grateful," "Hollywood," and "unpredictable." For the forty years or older group, "financially," "secure," "prince," "nonsmoker," "worker," "sexual," and "sincerity" were the most prevalent. Positive beta-values were the indicators that a word was particularly common for that age category, and vice versa for negative (see figure 2).

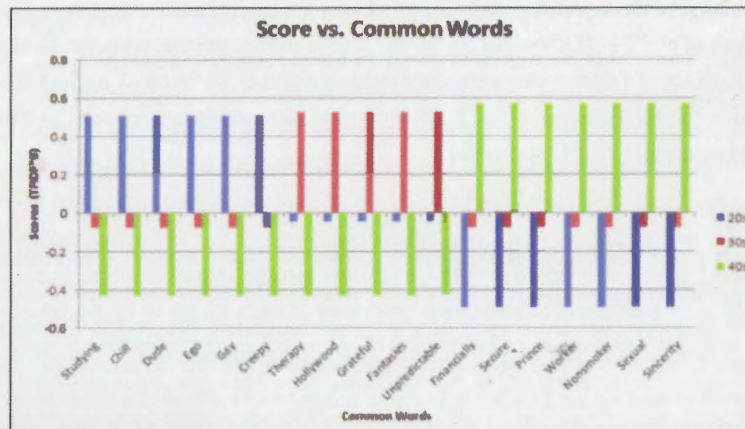


Figure 2

TF-IDF scores for selected words in different age groups

## Discussion

Our program combines two current streams of thought in word frequency analysis: commercial applications on user-generated data, and academic research about demographics. Through the application of academic study to an online commercial domain, we believe that our application, though rudimentary, opens

up new avenues of study and development.

Commercial interest in word frequency analysis is booming. The *Google Ngram Viewer*, a program which allows users to track the frequency of words or phrases in books over time, was an early success, but by no means the last. In the summer of 2013 *OpenAmplify* developed an app called *vs*, the first entertainment app to use word frequency analysis on tweets in real time (Guess, 2013). Users select two topics, *vs*. then breaks down the most recent tweets containing the topic words to determine which is being discussed more frequently on *Twitter*.

There has been significant interest in using word frequency analysis and other forms of NLP to study age demographics. Pennebaker and Stone (2003) published several studies which used word frequency analysis to study demographics. They found that older individuals are more likely than younger individuals to use plural and future-tense verbs, to use positive emotion rather than language emotion words, and to use fewer self-references like "I" and "me." Even though they did not use word frequency analysis exclusively, Pennebaker et al. demonstrated and defended its importance as a tool for gathering information on language use and demographics.

Schwartz et al. (2013) published a study applying word frequency analysis to the *Facebook* messages of 75,000 volunteers who varied along the lines of age, gender and personality. They found out which words and compounds were common for different age groups. Thirteen to eighteen year-olds tended to use more internet slang and emoticons like "xd," ":-)" and "<3" and use words like "school" and "homework." For nineteen to twenty-two year-olds, the words "school," "fuck," and "campus" were most prevalent. The most common words for thirty to sixty-five year-olds were "daughter," "fb friends" and "country." Schwartz et al. used an *open-vocabulary* analysis, which means that they used statistical tools to find out which words were most representative of each category. This is a departure from the previously popular *a priori* analysis, where words are classified into diagnostic categories which are then applied to textual data. We followed Schwartz et al.'s open-ended approach as it allows for more unexpected results to emerge, and is less likely to be constrained by preexisting assumptions about what different word choices mean.

Pennebaker et al. (2003) compared analyzing texts to understanding the layout of a city by driving on the roads in a car, or by flying above in a helicopter. Both methods provide different but equally valid views of the city. While the helicopter will likely miss details of specific streets, it can pick up information about the overall structure of the city that the car would miss. Word frequency analysis is like the helicopter; it passes over specific meanings that a straightforward



reading would catch, but provides linguistic information “from a distance” that has the power to illuminate connections and insights that would otherwise remain hidden. We hope that through our research, we have demonstrated the power of word frequency analysis at revealing and disseminating demographic language trends that might otherwise be undetectable.

## REFERENCES

- Coyle, J. R., Presti, D. E., & Baggott, M. J. (2012). *Quantitative analysis of narrative reports of psychedelic drugs*. Cornell University Library, doi: <http://arxiv.org/abs/1206.0312>
- Guess, A. (2013, July 05). Track sentiment analysis on your mobile device with the new vs. app. *SemanticWeb*, Retrieved from [http://semanticweb.com/track-sentiment-analysis-on-your-mobile-device-with-the-new-vs-app\\_b38317](http://semanticweb.com/track-sentiment-analysis-on-your-mobile-device-with-the-new-vs-app_b38317)
- Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547-577. doi:10.1146/annurev.psych.54.101601.145041
- Pennebaker, J., & Stone, L. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, *85*(2), 291-301. doi: 10.1037/0023514.85.2.291
- Rahman, M., Ahmed, S., Islam, S., & Rahman, M. (2013). An effective ranking method of webpage through tf-idf and hyperlink classified pagerank. *International Journal of Data Mining & Knowledge Management Process*, *3*(4), 149-156. doi:10.5121/ijdkp.2013.3411
- Smith, M., & Kanade, T. (1997). Video skimming and characterization through the combination of image and language understanding technique. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 775 - 781. doi: 10.1109/CVPR.1997.609414
- Schwartz, H., Eichstaedt, J., Kern, M., Dziurzynski, L., Ramones, et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, *8*(9), e73791. doi:10.1371/journal.pone.0073791
- Vilensky, B. (1996). Can analysis of word frequency distinguish between writings of different authors?. *Physica A*, *231*, 705-711.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the academic word list (awl) and non-awl content words in applied linguistics research papers. *English for Specific Purposes*, *28*, 33-41. doi: 10.1016/j.esp.2008.08.003



the oubliettes of dreams

Eleanor Qu



# Sound changes of the /h/-phoneme from Old English to Modern English

Derrick Buchanan

*Carleton University, Ottawa, ON, Canada*

The goal of this paper is to investigate the sound changes that occurred to the /h/-phoneme from Old English to Modern English. What I found is that there is a myriad of natural processes of language change as well as many sociolinguistic implications that have influenced the /h/-phoneme. Instances of natural processes discussed in this paper are loan words, /h/-dropping, glide cluster reduction, loss of medial /h/, weak aspiration, and zero realization. Instances of sociolinguistic implications discussed in this paper are the divide of upper and lower class, proper access to education, prestige, and negative stigmas. Importantly, distinctions between different dialects and writing systems will be made which helps reveal the progression of the changes that occurred to the /h/-phoneme.

## Introduction: Brief Historical Overview

Currently as well as historically there exist countless phonetic variations between the dialects of English: these phonetic variations exist across different regions of the world, as well as throughout different points in time. The multiform spelling and pronunciations of words across the dialects of English are owed to various historical, cultural, and other language influences (Stockwell & Minkova, 2001). The focus of this paper regards investigating the changes that have occurred to the voiceless glottal fricative—the /h/-phoneme in English—that made it the way it is now. Particular emphasis will be placed on the phonological changes of the phoneme across English dialects, as well as the sociolinguistic and historical



implications of those changes. I will explore the changes of the /h/-phoneme under these perspectives by looking primarily at its use in Old English (approx. c. 450-1066), with some examples from Middle English (approx. 1066-1776) and finally looking at its use in Modern English (approx. 1776- present) (Stockwell & Minkova, 2001). These dates are intended only to establish a relative frame of reference as I allude to a given era; I do not intend to be overly scrutable concerning specific dates throughout this paper. The majority of the changes to the /h/-phoneme discussed in this paper occurred due to natural processes, and the remaining changes occurred due to sociolinguistic implications. Although it is certainly arguable that sociolinguistic language changes occur due to natural processes, for the purpose of this paper, sociolinguistic influences will be regarded simply as language changes that occurred due to the effects of society. Instances of natural processes discussed in this paper are loan words, /h/-dropping, glide cluster reduction, loss of medial /h/, weak aspiration, and zero realization. Instances of sociolinguistic implications discussed in this paper are the divide of upper and lower class, proper access to education, prestige, and negative stigmas. In the remainder of this paper I will explore all of the aforementioned language changes and influences.

### Features of Old English: Relationship Between Spelling and Pronunciation

Old English has distinctive language features that dictate the spelling and pronunciation of a word be on a one to one relationship. This means that a feature of Old English is the pronunciation of certain vowels or consonants in words which are now regarded as silent in Modern English words. The Modern English “word knight, for example, has only three sounds that are audible, but the Old English word from which it [knight] is derived—*cnicht*—every letter is given its full phonetic value” (Alston, 1962, p. 8). Since silent letters are not a feature of Old English, it is very useful for exploring and understanding the changes and loss of specific phonemes that have occurred over time. Old English is useful for understanding changes to phonemes because when compared to Modern English the current pronunciation can be compared to the old pronunciation and the differences that are revealed can be understood as changes that occurred over time. In Old English the /h/- phoneme occurred commonly in speech, as well as consequently or subsequently, whichever may be the case, the letter <h> also shows prevalence in the writing systems of Old to Modern English. We will come to see how both the writing systems and speech, letter and phoneme- /h/ thusly, influenced each other to change whilst still preserving the mere aspiration that is “h”.

### Old English Allophones of /h/ and Introduction to Phonetic Changes

There were a few other overall sound changes that took place environmentally which affected [h] and its allophones from Old English to Modern English. Examples of these changes are the loss of intervocalic [h], glide cluster reduction, and the loss of medial [h] (Ringler, 1971). These losses can be roughly categorized into an umbrella term that expresses the overall change that occurred with the loss of [h], known as /h/-dropping. Originally in Old English, [h] frequently occurred used compared to its less common use in the majority of Modern English dialects. It makes sense for there to be a decline in the use of [h] throughout history as the /h/-phoneme is a very weak sound, naturally increasing its potential to be manipulated. [h] already shows great sound alterations based on its phonetic environment; the Old English allophones of [h], [X] and [ç] can attest this. In

Old English <h> is pronounced [h], as in Modern English, at the beginnings of syllables, but elsewhere it is pronounced approximately like German ch in *Nacht* or *ich*—that is, as a velar [X] or palatal [ç] unvoiced spirant (pronounced with the tongue against the velum [soft palate] or, after front vowels, against the hard palate)” (Baker, 2007, p. 15).

The consequence of these varying sound alternations is the weakening of the pronunciation of [h] which is a contributing factor to the overall decline in the use of the /h/-phoneme.

### Standard Initial /h/-dropping and Circumstantial /h/-dropping

The declining use of [h] in English is owed mostly to the phenomenon known as /h/-dropping: “a dialectal feature associated to the phonetic realisation of the letter <h>....So-called /h/-dropping is basically the absence of the /h/-phoneme— aspiration—in certain phonetic environments, especially at the beginning of a word [or syllable]” (Lopez, 2007, p. 158). The phenomenon of /h/-dropping still persists today in different English dialects, however historically it originated from various manipulations to [h]. Although it is notable that there is a slight distinction between what I will call standard initial /h/-dropping, and circumstantial /h/-dropping. Standard initial /h/-dropping can be characterized by sheer removal of the phoneme due to cultural or sociolinguistic influence. According to this, words like *happy*, would sound like [æpi] and this dropping of initial [h] appears to be a representation of cultural influence and sociolinguistic status. As opposed to circumstantial /h/-dropping which persists presently but originated mostly historically, and was due to phonetic environments. For instance, one case from *Bright's Old English Grammar and Reader* is the aforementioned “intervocalic h,



which disappears, and the vowel which followed it is absorbed into the vowel or diphthong which preceded it, by compensation lengthening it" (Ringler, 1971, p.57). This example importantly informs the understanding of circumstantial /h/-dropping by showing how it differs from standard initial /h/-dropping. Standard initial /h/-dropping occurs much more deliberately and most often at the beginning of words, where as circumstantial /h/-dropping occurs in more complex environments where [h] is merely absorbed into a vowel or diphthong as opposed to being dropped completely.

### Phoneme /h/ in Glide Cluster Reduction

Furthermore, throughout the transition from Old to Middle English, a further loss of the /h/-phoneme became prominent in glide cluster reduction. Glide Cluster Reduction involves the zero-realisation of /h/ when it occurs in glide clusters such as /hw/ and /hj/. Historically glide clusters like <hw> and <hj> were "pronounced /hw/ and /hj/, but in some dialects they eventually became [w] and sometimes [j]. Therefore, words like 'where' began to be pronounced [wecr] or [wec] rather than [hwecr] or [hwec]" (Lopez, 2007, p. 158). These glide clusters are prominently part of the Old English phonemic inventory, however, as pronunciation is simplified, it is accepted as standard that most English dialects abide by a reduction in glide clusters. Wells (1982) provides evidence that most English dialects abide by a reduction in glide clusters by noting that it "characterizes most accents of England and Wales, the Southern hemisphere, and the West Indies, and also some American speech, but not the accents of Scotland and Ireland" (p. 228). It is important to remember that, even though there are natural transitions (like simplified pronunciation patterns), such phonetic transitions are not always reflected in the syntax of the writing system. For instance, recall silent letters and the example of "where". Historically the /h/-phoneme would have been pronounced when this word is uttered, however in Modern English the /h/-phoneme in "where" is silent. Hence, for words like "where", the /h/-phoneme is being regarded as silent even though [h] is still reflected in the syntax of the writing system. In other words, etymological changes and phonetic changes do not always coincide. I will explore the implications of these important notions later as they become more relevant in the discussion of cultural and sociolinguistic influence.

### Medial Loss of /h/ and its Old English Allophones

The last main environmental loss of [h] to discuss is the loss of medial [h]. The rule for this, according to Bright's Old English Grammar and Reader, is that "Medial h (but not hh) preceded by r or l and followed by an inflectional vowel disappears,

and, in compensation, the stem-vowel is lengthened: Eg. Mearh, → mēares, horse; seolh, → seōles, seal" (Ringler, 1971, p. 56). The loss of medial [h] encompasses the disappearance of [h] in its medial position, and thus would also be responsible for the loss of the Old English allophones [X] and [ç]. For instance, let us use the Old English version of the word daughter as an example. In Old English daughter was actually dohtor, "which was pronounced ['dɔxtər]. The development of this word gave ['dɔxtər] and then ['dɔxtər], ['dɔtər], ['dɔ:tər], and ['dɔ:tər(-)], where the original allophone of /h/, that is to say [X] —which became [x] later on— has certainly been lost" (Lopez, 2007, p. 159). Given this succession of changes it seems apparent that the [X] allophone was lost due to a number of adaptations which lead to a more simplified pronunciation pattern. Similarly in these circumstances with the allophones, even though they are dropped in the medial position they still influence the pronunciation of the vowel by lengthening it. Now that I have accounted for the loss of [X] allophone, I will now account for the medial loss of [ç] allophone. The disappearance of the [ç]-allophone occurred quite similarly to the disappearance of the [X] allophone. For instance, consider the use of medial /h/ in the Old and Middle English word *riht* "right". This word was originally pronounced [riçt] which developed into [raiçt], [raixt], and [rait] (Lopez, 2007). After following the sound transition it remains evident that the [ç] allophone was also lost in Modern English. This covers the overall loss of medial [h] and its allophones; and all together the last three paragraphs encompass the varieties of circumstantial /h/-dropping that occur in phonetic environments.

### Cultural and Sociolinguistic Influence on /h/- dropping

Interestingly, initial /h/-dropping did not solely occur due to the aforementioned stipulations and constraints of pronunciation and phonetic environment; but also occurred due to confusion after the influence of certain French loanwords like *herb*, *host*, *honour*, or *honest*, were introduced to English. This brings us back to the cultural and sociolinguistic influence on [h] that I made reference to earlier, where these influences caused a confusion amongst English speakers that lead to a zero-realisation of the /h/ phoneme in certain environments. According to Lynda Mugglestone (1995), "French loans regularly 'dropped their [h]s', and at some date, such habits seem to have been extended into native [English] words as well" (p. 110). Support for this claim can be attained by analyzing the development of the aforementioned French loanwords in Latin, French, and English. For instance, the word *herb* initially comes from the Latin word *herba*. Given the phonological rules of Classical Latin, *herba* was actually pronounced /'herba/, with a clear initial aspiration (Lopez, 2007). However, due to the weak quality of the initial



aspiration the sound was eventually lost in Latin which transitioned to become “[erba]. Then this word developed into Old French *erbe*, probably pronounced [erb]” (Lopez, 2007, p. 161) which would later be loaned to English. This example of the transitions that occurred to different loanwords alludes to one of the basic claims in the current paper; despite the phonological rules of classical Latin (and other languages), the naturally weak quality aspiration of /h/ prompted a gradual decline in its pronunciation in varying phonetic environments. However, aside from the importance of the last claim, the ignorance of English speakers seems to be playing the most viable role in the cultural influence on /h/. Consider the fact that the word *erbe* was loaned to the lexis of Middle English from Old French. Since *erbe* was a loanword, there was confusion as to its correct spelling—whether it should be *erbe* or *herbe*. This caused further confusion for how to correctly pronounce the word in Middle English. According to Lopez (2007), despite the differences in spelling, either word was pronounced “[erb], as the presence of an initial <h> in the second spelling of this word was restored only for etymological but not phonetic reasons” (p. 161). The English speakers’ lack of knowledge concerning the spelling of the newly introduced loanwords rendered them unable to effectively determine the appropriate circumstances where [h] should remain against those where it should be dropped. This confusion between cultures further influenced the habit of /h/-dropping which inevitably caused the habits to be extended into native English words. Interestingly, these habits and language misunderstandings actually still persist today, and are rather dependent on one’s susceptibility to language trends and access to proper education. In the next section I will examine the implications of some of the latest and current trends in /h/-dropping.

### Cultural and Sociolinguistic Implications of /h/- dropping

Once /h/-dropping was recognized and extended habitually into native English dialects, the sociolinguistic implication of prestige contributed to the perpetuation of confusion because of a divided social class. The divide in the social class can essentially be narrowed down to those with proper access to education and those without proper access to education. Individuals with proper education would be aware of the correct spelling of words. This means that those who have a proper education would eventually become aware that words which they had always pronounced without initial aspiration are actually correctly spelled with an initial <h>. This caused a confusion among divided social classes because the upper class who became aware of the correct spelling of words with an initial <h> were now trying to aspirate these words (Lopez, 2007). Conversely, the lower class individuals

that did not have proper access to education never learned that these words were correctly spelled with an initial <h>, and hence the lower class continued to not aspirate it (Lopez, 2007). This means that the habits of the loanwords managed to affect the native words, sparking confusion regarding /h/-dropping. Then, the habits from the loanwords ended up becoming associated with hypercorrection in the upper class and careless speech amongst the lower class, which in turn perpetuated the confusion and led to different accepted uses of /h/ in different dialects of English. Currently these preferences to /h/-dropping cannot really take precedence over one another since they have all now developed into accepted English dialects within their own regions. However, it is worth noting that there is still sometimes a negative stigma attached to /h/-dropping in British English, like the example from earlier of *happy* becoming [æpi].

### Conclusion

It is certainly evident that the /h/-phoneme underwent serious natural and sociolinguistic changes from Old English to its current use now in Modern English. The major change identified in this paper that occurred to the /h/-phoneme is its heavy disappearance due to an overall tendency for it to be weakened to the point of zero-realisation. I can conclude that both Latin and Old French had a large impact on the historical development of the /h/-phoneme in Middle English, and subsequently Modern English. Also, it should be understood that the etymological restoration of <h> is what caused the major sociolinguistic confusion among social classes, as it was only after these restorations of <h> in the spelling that words with an initial <h> began to again be pronounced with initial aspiration. This is however only in “Standard British English, remaining unaspirated in the United States and, of course, in those dialects of England where /h/-dropping is a current feature, regardless of the French or native origin of the word” (Lopez, 2007, p. 161). Overall, it is important to keep in mind, that like everything else from (c. 450-1066), the /h/-phoneme has gradually changed, and that is why it is currently used the way it is in Modern English.

### REFERENCES

- Alston, R. (1962). *An introduction to old english*. Toronto, Canada: Copp Clark Publishing Company.
- Baker, P. S. (2007). *Introduction to Old English second edition*. USA: Blackwell Publishing.
- Haker, M. (2004). *New perspectives on english historical linguistics*. John Benjamins Publishing Company.
- Lopez, I. (2007). *The social status of /h/ in English*. University of Salamanca.
- Mugglestone, L. (1995). *Talking proper: The rise of accent as social symbol*. Oxford: Clarendon.



- Ringler, R. N. (1971). *Brigh's Old English grammar and reader*. USA: Holt, Rinehar and Winston Inc.
- Stockwell, R., Minkova, D. (2001). *English words: History and structure*. Cambridge: Cambridge University Press.
- Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.
- Williams and Anderson. (1935). *Old English handbook*. Massachusetts: The Riverside Press Cambridge.



*straight to the future*

**Stefano Menicagli**



# Speaking, remembering and feeling the self: How autobiographical memories bring forth the experience of a self through language

Mauro Silva

University of Texas, Austin, TX, USA

The self has been explored by many traditions and in many contexts throughout human history. In this thesis, I propose that the self, as a phenomenal experience, arises from autobiographical memories mediated by self-reference in language. I resort to philosophical texts, neuroimaging studies and works from academic literature to support this notion. Relevant findings are analyzed and placed in the context of a modern discussion on selfhood, consciousness and personal identity. In addition, plausible implications for psychotic disorders and brain disease are discussed.

## Introduction

I am. *Yo soy*. At least that is the proposition that precedes what I say and write about what I like to call “me” and what is presupposed when I think about my own past, inclinations and thoughts. We describe this representation as the self; a rationalization by cognizant beings that make it possible to talk about an external world. The interpretation of the self as an actual entity has been challenged by different streams of philosophical thought. Others assert the validity of the self as a finite characteristic of the human mind—some even seek to find its anatomical location (Berrios & Markova, 2003). There is, however, something to be said about *oneself* in the language one uses to convey meaning, either to others or within our consciousness. One would also think that an idea like the self is rooted on one’s own memories, including life narratives, which are revised experiences and events from one’s past that have shaped perceived self traits. How does the self, then, specifically relate to language and autobiographical memory?



One of the most remarkable features about the experience of selfhood is the ability to unify thoughts and actions under a first-person agent across time.<sup>1</sup> Each of us experiences this temporal continuity seamlessly and it is so imbued into our normal operation that we fail to notice it, outside of deep introspection. Naturally, autobiographical memory, memory of one's own actions, is construed as the content material from which the self establishes its procession in time. However, it is the sense of familiarity and ownership of such recollections that characterizes the *feeling* of selfness. As other forms of memory lack this phenomenal quality, the investigation of what makes autobiographical memory unique can provide insight on the nature of experiencing "self" (Conway, Singer, & Tagini, 2004).

Previous studies have shown a correlation between the acquisition of language and the emergence of autobiographical memory (Fivush, 1998). It has also been shown that retrieval of memories of one's own experiences greatly depends on language-specific cues (Matsumoto & Stanny, 2006). These observations demonstrate that some aspects of language modulate the encoding and decoding of autobiographical memory. From this, an important link between language and the self can be elucidated.

Language has been previously related to layer-based concepts of self. For example, an individual's personal narrative, considered an aspect of the reflective self, has consistently been shown to change at different stages of language mastery (McAdams, 1999). Previous research suggests a social-environmental relationship between some language aspects and certain selfhood notions such as the idea of a social self (Fivush & Nelson, 2004; Kivinen & Piironen, 2012). However, there is more to be discussed about language and its link to more fundamental aspects of the self. Although language may affect the experience of self, a clear relationship has not been detailed. This paper seeks to reduce this deficit extant in current conceptions of the self.

I propose that the experience of a self is constructed by autobiographical memories that are shaped with language acquisition. Self-referential factors in language mediate the nature of memories, which in turn contribute to their ownership, fundamental characteristic of self-like experiences. Initially, this paper will examine the history of ideas about the self, from early philosophy meditations to recent scientific approaches, in the context of introducing an interdisciplinary definition of the self. It is to be argued in this paper that certain aspects of language

<sup>1</sup> Various authors, including Proulx and Chandler (2007), challenge the assertion that a self, unified in time and number, is a cross-culturally accepted notion. In fact, they suggest that multiplicity of selves within an individual, at least to some extent, is a normative view in some cultures. Nevertheless, we will assume there is a psychological basis for *some degree* of self unification at the very least and that this occurrence may rely on autobiographical memories.

play a major role in the phenomenology of autobiographical memory. From the analysis, a relationship between language and self will be inferred. Lastly, the discussion will address the potential implications that the language and memory enquiry may have on our current understanding of what it really means to say "I am" or "Yo Soy".

### What is the Self?

The definition of the self to be employed in this paper is the following: *The self is a distinct phenomenal experience, time-oriented but temporally independent, in which certain ideas and dispositions entering the stream of consciousness are perceived with a sense of ownership and agency. The experience of self can be subdivided into two dimensions; minimal self (pre-reflective) and extended self (reflective).*<sup>2</sup> *The self can only be accessed from a first person perspective and consideration of others' selves is only possible through an appropriate theory of mind.* The first sentence of the definition establishes that selfhood is looked upon as an internal phenomenon, which is distinct from personhood, a concept that presupposes a connection with the external world and executive action. Moreover, this part describes that although the self experience can reference other experiences in time, the self alone cannot be referenced to a particular time (Gallagher, 2000). The second sentence alludes to the observation that the self experience has a non-conscious passive and primal component and an introspective, more complex aspect. The third aspect of our definition provides a valid framework for the study of the self. Strictly, we can only know about the self from the first person perspective. However, a theory of mind, namely inferring that others experience selfhood and possess discrete agency, allows us to discuss the self from second and third person perspectives (Brüne, 2005). We address the historical background and motivation behind this definition and how it may relate to memory and language.

For a long time the discussion about the self centered on its metaphysical quality, whether it represents a tangible entity or not. Evidence for the search of a concrete self can be traced back to 800 BCE in the *Katha Upanishad*, a sacred text of ancient India. In the document, the author describes the self as what sits in a chariot, which represents the body, controlled by the mind, which in this analogy is represented by the reins (Muller, 2004). Opposing this substantive view, Xuanzang, a prominent figure in 7th century Chinese Buddhism, reflects on the self as a construction based on false ideas, like the differentiation of selves based

<sup>2</sup> A more basal or proto self is referenced in Damasio (2000), Gallagher (2000) and Nelson et al. (2009), which concerns about self-discrimination in sensory processing. The primitive proto self may underlie mirror self-recognition in infants younger than two years of age and in great ape species (Keenan, Gallup, & Falk, 2004). However, I will omit this form of self-discrimination from the definitional framework.



on bodies, for instance (Bonevac & Phillips, 2009).

In early Western Philosophy, the self as an entity became a key framework for questions about human nature. Converging upon the chariot imagery described in the *Katha Upanishad*, Plato alludes to this charioteer (the self) who controls the three horses that represent the three parts of the soul (Plato, 1871). More explicitly, Aristotle describes the self as a primary substance, something that cannot be reduced to simpler parts, and ascribes to it the unity of mental and physical properties (Bonevac & Phillips, 2009). Modern Western philosophers built upon this foundation to elaborate a conception of the self that remained widely uncontested in Western culture for centuries. To avoid entering the millenary debate about the self's reality, the definition outlined above regards the self in terms of its phenomenal qualities—what it is meant by saying that someone is engaged in an experience of self from the first person perspective. As such, a metaphysical claim about phenomenal experiences, mind and consciousness is outside the scope of this paper.

Modern western philosophers advanced upon an epistemological approach to the self. Most famously, Rene Descartes asserted that the only thing we can be sure of beyond any doubt is the self. This view is condensed in his well-known statement, *Cogito ergo sum* (I think, therefore I am) (Descartes, 1993). This hallmark realization in Western thought would come under fire with David Hume who cast doubt on the privileged standpoint of knowledge about the self. Particularly, Hume claimed that the idea of self is derived from our imposition of causality into the world, namely positing that mental events related to the self are causally conjoined to form unity. Hume points out, "I never can catch myself at any time without a perception and never can observe anything but the perception" (Hume, 1739). By this, Hume asserts the self is nothing but a construction, a set of experiences connected by memories, but not an entity of experience itself from which knowledge can be derived. Alternatively, this paper proposes a self that can be construed as an experience with specific phenomenal qualities modulated by language and memory.

Furthermore, my definition of the self may be of practical use in discussing "self-disorders", such as schizophrenia. Since the end of the 19th century, approaches in psychiatry and psychoanalysis focused on a practical view of the self and described disorders in relation to the state of self in affected patients (Berrios & Markova, 2003). Additionally, I seek that the definition of self-derived here can be employed in the context of a discussion about loss of self in patients with diseases affecting autobiographical memory, like Alzheimer's disease. This inclusiveness allows us to derive pertinent implications about alteration of the

self-experience in various diseases from our analysis of the relationship among language, memory and the self.

### Autobiographical Memory, Language, and the Self

Psychologists have categorized memory into distinct systems. Autobiographical memory is a form of declarative memory, as opposed to non-declarative memory (e.g. learning to ride a bicycle, fear conditioning), which many other species seem to possess (Conway, 2009). Within declarative memory, autobiographical memory falls under the sub-category of episodic memory, which is situated in a temporal context, as opposed to semantic memory (e.g., the capital of France, a sandwich is a food object), which mainly references factual information (Conway, 2009; Tulving, 1985). Fivush and Nelson (2004) define episodic memory as "an explicit memory of an event that occurred in a specific time and place in one's personal past" (574). In the brain, tasks involving episodic memories, including autobiographical memory, have been associated with activation of medial areas in the cortex (Northoff et al., 2006). In particular, functional imaging has demonstrated that activation of the ventromedial prefrontal and posterior cingulate cortex is highly correlated with autobiographical recall tasks (Gusnard et al., 2001; Spreng et al., 2009). However, the only definitional distinction between autobiographical memory and other forms of episodic memory lies on its subjective quality; that the individual experiences a sense of self-ownership when invoking them. Thus, even though it may seem obvious that one's autobiographical memories feel like one's own, this observation represents the basis for linking autobiographical memory to the feeling of self across time, as it will be discussed below.

As described in our definition, a part of experiencing a self is the ability to ascribe ownership to experiences at a specific point in time. This is a crucial aspect of the reflective self. For the most part, we come to make sense of ourselves through the many different personal narratives contributed by our own memories and the recounts of others (McAdams, 1999). However, it is the trans-temporal nature of the self-experience that represents the most distinctive feature of the pre-reflective self. We integrate our autobiographical memories to form a sense of self-continuity in time; a self that is not associated to a particular timeframe (Gallagher, 2000). This would indicate that autobiographical memory relates to the self in both its reflective and pre-reflective dimensions.

Autobiographical memories not only provide the content for the reflective (extended) self but also help provide phenomenal context for the pre-reflective (minimal) self. Meta-analyses of imaging studies showing cortical activation in self-referential tasks, demonstrate a statistically significant, positive correlation



between the brain areas activated in tasks involving the recall of autobiographical memories and those activated in performing other self-referential processes (e.g. own personality judgment, self-facial recognition, etc.), across multiple sensory input modalities (Northoff et al., 2006; Spreng et al., 2009). Not surprisingly, this analysis empirically supports the notion that memories of our own past are functionally related to aspects of what we consider the extended self. Interestingly, similar studies show activation of the same regions in tasks involving self-projection into the future (Buckner & Carroll, 2007). More groundbreaking, however, is what Northoff et al. (2006) propose about the phenomenal quality of the self in light of their findings. The authors suggest that “[t]he more the respective stimulus is associated with the person’s sense of belongingness, the more strongly it can be related to the self” and that, ultimately, the strength of the self-stimulus relation weighs in the “mineness” or ownership that accompanies the response to the stimulus (441). This implies that the extent to which the memory is *deemed* as belonging to oneself contributes to the *feeling* of selfness, characteristic of the pre-reflective self. It follows that an underlying cognitive process of identifying selfness is required for the phenomenal experience of self. Considering previous research, language emerges as a natural candidate mediator for this self-discriminating cognitive capacity given that autobiographical memories are likely dependent on certain language aspects explored next.

For instance, a connection between language and autobiographical memory can be inferred from the well-known phenomenon of childhood amnesia. Although children as young as two years old show signs of recalling past events, most adults cannot recall personal events occurring prior to three and a half to four years of age (Simcock & Hayne, 2013). The age disparity between forming memories of the past and the offset of childhood amnesia can be explained as a phenomenon resulting from the gradual acquisition of increasingly sophisticated and descriptive language, which ultimately permits the consolidation of truly autobiographical memories and not just events in the past. It seems that language not only serves as a vehicle to convey the content of memories, but it also provides the scaffold for organizing the encoded experience (Fivush & Nelson, 2004).

Studies of memory recall in children provide further evidence for a language mediation of autobiographical memory. For instance, Simcock and Hayne (2002) assessed verbal and behavioral recall of engaging play activities in two- and three-year old children, six to twelve months after the activity took place. Their results showed that “in no instance during test did the child use a word or words to describe the event that had not been part of his or her productive vocabulary” at the time of memory formation (Simcock & Hayne, 2013, p. 230). This observation

suggests that linguistic proficiency circumscribes a given memory’s richness. Other studies showing a correlation between complex autobiographical memories and linguistic proficiency also support this notion (Fivush & Schwarzmueller, 1998). Additionally, subjects with aphasia display episodic memory deficits, whereas individuals with episodic memory impairments do not show signs of aphasia (Barba, Frasson, Mantovan, Gallo, & Denes, 1996). The preceding empirical observations strongly suggest that the capacity for autobiographical memory depends on language capacity.

The specific aspects of language that may constitute the scaffolding for constructing autobiographical memories, and ultimately the self experience, are to be determined. I propose that language provides the conceptual self-reference context that allows the formation of autobiographical memories and the feeling of self. Concretely, linguistic elements that instantiate self-reference (e.g., I, mine, etc.) provide the paradigms for engraving in memories the “metadata” about the selfness of the experience, as it has been set forth in theories of self-awareness (Stamenov, 2003). In addition, other constructions in language serve to place the ego in opposition to other interlocutors and environmental actors. The use of “I” as subject, object and owner in increasingly complex language exchanges would help create the acknowledgement that one’s own memories are distinct from those of others. This process describes a plausible mechanism for the formation of a theory of mind, that is, the realization that others’ may have desires, thoughts, etc. that may be different from one’s own, which would coincide developmentally with the emergence of theory of mind, which is typically around four years of age (Fivush & Nelson, 2004).<sup>3</sup> The particular elements in a language that dictate the rules of reference to self and others would then be implicated in the formation of autobiographical memories. In line with this notion, studies reveal that individuals from Western cultures, who employ languages with distinct self-referential structure, consistently recall earlier autobiographical memories than individuals belonging to Eastern cultures, whose language possesses constructions that are less self-centered (Fivush & Nelson, 2004). Further, neuroimaging scans have shown that Chinese individuals show similar activation when presented with self and mother-related trait judgment tasks, and overall greater activation than for the public person trait judgment. In contrast, American individuals did not show increased activation for the mother-related over the public person trait judgment task (Han & Northoff, 2009). As such, the degree of acquisition of language aspects that support self-reference would be a parameter to determine how much a memory is owned.

In the context of discussing the self experience, the richness of the



autobiographical experience can be placed in terms of its self-referential quality; an experience with sufficient self-ascribing elements will be conceived as one's own. Only after a stage in language mastery that coherently sustains self-reference is attained the self-stimulus relation becomes strong enough for the memories to be conceptualized as autobiographical. Once this occurs, retrieved autobiographical memories will elicit their distinct phenomenal quality of ownership, as proposed in Northoff et al. (2006). Further evidence that autobiographical memory may set the stage for the phenomenal self experience is given by noting that the same region activated during self-reference shows the highest metabolic activity during resting, non-task related conditions (Gusnard et al., 2001). This could plausibly represent the basis for the continuous pre-reflective self experience. The conceptualization of the phenomenal experience of self in terms of language and memory can have profound implications for diverse philosophical, social and clinical considerations.

### Implications

The outlined discussion provides evidence for a functionalist expression of what it is meant by "the self". We can come to understand the self as we rationalize fear. There is no "fear organ" or brain region exclusively responsible for fear, but we understand the psychological and physiological mechanisms that are taking place when we experience fear. Although fear is not a tangible thing, society and academia treat fear as a real phenomenon with emotional and bodily properties. In a similar way, I sought to characterize the self not in terms of physical substance or epistemic quality but rather in terms of the causal relationship between language and memory that may give rise to a feeling of self. By this approach, we circumvent a primordial standoff in philosophical discussions about the self. If it can be admitted that the feeling of self does have a neuropsychological basis, the discussion may then turn toward understanding how does the diverse quality of the self experience across different cultures may relate to personal identity notions. That is, rather than worrying about whether there is an irreducible *me*, I may instead inquire if and how the distinct way I *feel myself* can influence how I perceive and rationalize my own individuality and that of others. However, this view does invite a deeper exploration on the intrinsicity of the self experience in human consciousness.

The current thesis has brought about an instance where language is said to mediate an experience of self. The assertion that language influences a phenomenal experience has been proposed before in discussions about human consciousness (Kivinen and Piironien, 2008). For instance, Rosenthal (2008)

argues that language likely mediates the inferential association of behavior to inner mental states (838). Nevertheless, the inquiry above suggests a specific conduit by which language may mediate the phenomenal experience, namely by bestowing autobiographical memories with their "mineness" quality. Of course, given that the self-stimulus relation's role in provoking the feeling of ownership applies to other self-referential processes (e.g., prospection, navigation) as well, these would be expected to elicit similar phenomenal experiences. Whether language mediates the phenomenology of other self-referential processes is to be investigated. Nevertheless, it is the emergence of autobiographical memory that better correlates with the appearance of complex self-consciousness in children, which could foster the feeling of self (Keenan et al., 2004, p. 304). Also, it is unclear whether consciousness necessarily gives rise to a self experience, and if self-awareness presupposes primitive consciousness at the very least. As it turns out, this inquiry is important when dealing with artificial intelligence (Gallagher, 2000). This discussion may be enriched if language mediated encoding of memories is understood as a pathway to experiencing a self.

Furthermore, an interdisciplinary understanding of the self experience can lead to earlier diagnosis of psychotic disorders and better treatment outcomes for patients. Schizophrenia spectrum disorders are widely characterized by altered self-demarcation (distinguishing self from the external world) and disturbance of presence (alienation from self) (Nelson et al., 2009). Given that these symptoms are phenomenal and psychological in nature, clinical assessments that take into account memory and language's role in self experience could be devised. Such tests could assess self-reference capacity in language and autobiographical recall proficiency, and yield objective clinical markers on the state of self. This type of evaluation would result in a more reliable diagnosis given that, in current practice, phenomenological affections are determined from clinical conversations, which are dependent on subjective interpretation from the patient (Møller & Husby, 2000). Additionally, a linguistic approach to develop therapies for coping with schizophrenia and other personality altering disorders could be explored. Though speculative, such method could take the form of a specialized behavioral and speech therapy stressing self-reference that could help mitigate the symptoms of self-related disorders. Indeed, the differential use of first person pronouns and emotionally-laden words has been described as a successful predictor of treatment outcome in personality disorder patients (Arntz, Hawke, Bamelis, Spinhoven, & Molendijk, 2012). In addition, identifying the neural correlates of self experience can aid clinicians and families better understand the condition of "loss of self" undergone by patients suffering from neurodegenerative diseases. A multi-faceted



account on the self experience can translate into a diversified set of tools for characterizing and treating symptoms in various ailments.

## Conclusion

The self has taken many interpretations throughout human history. Here, I have proposed that language may play an underlying role in the phenomenal experience of self by providing autobiographical memories with a context of self-reference. The acquisition of self-reference through language elements signals the consolidation of autobiographical memories which in turn elicit the self experience. Adopting this view will enrich a model philosophical discussion on self and personal identity. Though a comprehensive view on consciousness and the nature of phenomenal experiences is still missing, outlining the functional dependence of some of the elements that come into play in evoking self experience can get us closer to this end. In addition, looking upon the self in terms of its relation to language can open the door to new ways of exploring diseases that affect self experience. Future research should aim to characterize which specific language constructions foster autobiographical memory and how does language impairment interfere with memory recall and self experience. It may be that referring to oneself may feel different, not only depending on one's memories but on one's language as well, such that even though I am and *Yo soy* mean the same, they may not evoke quite the same feeling within.

## REFERENCES

- Arntz, A., Hawke, L. D., Bamelis, L., Spinhoven, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behaviour Research and Therapy*, 50(3), 191–202. doi:10.1016/j.brat.2011.12.007
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110–8. doi:10.1016/j.tics.2009.12.006
- Barba, G. D., Frasson, E., Mantovan, M. C., Gallo, A., & Denes, G. (1996). Semantic and episodic memory in aphasia. *Neuropsychologia*, 34(5), 361–7.
- Berrios, G. E., & Marková, I. S. (2003). The self and psychiatry: a conceptual history. In Kircher, T., David, A. (Eds.), *The Self in Neuroscience and Psychiatry* (pp. 9–39). Cambridge University Press.
- Bonevac, D. A., & Phillips, S. H. (2009). Introduction to World Philosophy: A Multicultural Reader. Oxford University Press.
- Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin*, 31(1), 21–42.
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, 11(2), 49–57. doi:10.1016/j.tics.2006.11.004
- Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, 47(11), 2305–13. doi:10.1016/j.neuropsychologia.2009.02.003
- Conway, M. A., Singer, J. A., & Tagini, A. (2004). The self and autobiographical memory:

- Correspondence and coherence. *Social Cognition*, 22(5), 491–529.
- Damasio, A. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Orlando, FL: Mariner Books.
- Descartes, R. (1993). *Meditations on first philosophy* (D. A. Cress, Trans.). Indianapolis, IN: Hackett Publishing.
- Fivush, R. (1998). The stories we tell: How language shapes autobiography. *Applied Cognitive Psychology*, 12(5), 483–487.
- Fivush, R., & Nelson, K. (2004). Culture and language in the emergence of autobiographical memory. *Psychological Science*, 15(9), 573–7. doi:10.1111/j.0956-7976.2004.00722.x
- Fivush, R., & Schwarzmueller, A. (1998). Children remember childhood: implications for childhood amnesia. *Applied Cognitive Psychology*, 12(5), 455–473. doi:10.1002/(SICI)1099-0720(199810)12:5<455::AID-ACP534>3.0.CO;2-H
- Gallagher, I. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14–21.
- Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(7), 4259–64. doi:10.1073/pnas.071043098
- Han, S., & Northoff, G. (2009). Understanding the self: A cultural neuroscience approach. *Progress in Brain Research*, 178, 203–12. doi:10.1016/S0079-6123(09)17814-7
- Keenan, J., Gallup, G. G., & Falk, D. (2004). *The face in the mirror: How we know who we are*. New York, NY: HarperCollins.
- Kivinen, O., & Piirainen, T. (2011). *On the distinctively human: Two perspectives on the evolution of language and conscious mind*.
- Matsumoto, A., & Stanny, C. J. (2006). Language-dependent access to autobiographical memory in Japanese-English bilinguals and US monolinguals. *Memory*, 14(3), 378–90. doi:10.1080/09658210500365763
- McAdams, D. P. (2008). Personal narratives and the life story. In J. P. Oliver, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 242–262). New York, NY: Guilford Press.
- Møller, P., & Husby, R. (2000). The initial prodrome in schizophrenia: Searching for naturalistic core dimensions of experience and behavior. *Schizophrenia Bulletin*, 26(1), 217–32.
- Muller, F. M. (2004). *The Upanishads Part II: The sacred books of the East Part Fifteen* (p.404). Kessinger Publishing: Whitefish, MT
- Nelson, B., Fornito, A., Harrison, B. J., Yücel, M., Sass, L. A., Yung, A. R., Wood, S. J., Pantelis, C., & McGorry, P. D. (2009). A disturbed sense of self in the psychosis prodrome: linking phenomenology and neurobiology. *Neuroscience and Biobehavioral Reviews*, 33(6), 807–17. doi:10.1016/j.neubiorev.2009.01.002
- Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–57. doi:10.1016/j.neuroimage.2005.12.002
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi:10.1126/science.1107621
- Plato. (1871). *Phaedrus* (B. Jowett, Trans.). New York, NY: Charles Scribner's Sons.
- Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, 46(3), 829–40. doi:10.1016/j.neuropsychologia.2007.11.012
- Simcock, G., & Hayne, H. (2013). Breaking the barrier? Children fail to translate their preverbal memories into language. *Psychological Science*, 13(3), 225–231.
- Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default



mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, 21(3), 489–510.  
doi:10.1162/jocn.2008.21029

Tulving, E. (1985). How many memory systems Are There? *American Psychologist*, 40(4), 385–398.



B E N G R E E N  
 T H E R E N C E T A O  
 T H E N T A I N P R I M E S  
 C A R B O N T A I T R A R I L Y  
 L A R G E A R I T H M E T I C  
 P R O G R E S S I O N S



$$(G, m+1) = G(0, \dots, 0)(\log R)^{m+1} + \sum_{j=1}^m O(\|G\|_{c_j})$$

