

What can Milgram and Zimbardo teach ethics committees and qualitative researchers about minimizing harm?

Research Ethics
2014, Vol. 10(2) 86–96
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1747016114523771
rea.sagepub.com


Martin Tolich

Sociology, University of Otago, New Zealand

Abstract

The first objective of this article is to demonstrate that ethics committee members can learn a great deal from a forensic analysis of two classic psychology studies: Zimbardo's Stanford Prison Study and Milgram's Obedience Study. Rather than using hindsight to retrospectively eradicate the harm in these studies, the article uses a prospective minimization of harm technique. Milgram attempted to be ethical by trying to protect his subjects through debriefing and a follow-up survey. He could have done more, however, by carrying out what ethics committees routinely insist on today for those researching sensitive topics. The establishment of counselling supports to identify harm to participants would have minimized additional harm. Were these in place, or in Zimbardo's case had the Stanford Ethics Committee properly identified Zimbardo's conflict of interest – he was both a principal investigator and the prison warden – how much harm could have been minimized? The second aim is to examine how some qualitative authors routinely demonize these classic studies. It might appear that there are too few cases of unethical qualitative research to justify such an examination; however, this article identifies a number of recent examples of ethically dubious qualitative research. This would suggest that qualitative research should examine its own ethics before poaching from psychology.

Keywords

ethics, Milgram, qualitative research, sociology, Zimbardo

Corresponding author:

Martin Tolich, Sociology, University of Otago, PO Box 54, Dunedin, 9054, New Zealand.

Email: martin.tolich@otago.ac.nz

Unhelpful criticism of unethical social science research

Philip Zimbardo's (1973) Stanford Prison Study and Stanley Milgram's (1974) Obedience study are convenient shorthand fall guys for justifying the necessity of ethics review. As with Adam and Eve's original sin producing the fall of man in the Christian faith, Zimbardo and Milgram are cast in this role, not only for use in psychology, but emblematic of the need to evaluate behavioral research designs prospectively for all social sciences (Librett and Perrone, 2010). These studies are described as having iconic status (Haggerty, 2004), notorious (Homan, 2006; Israel and Hay, 2006; King et al., 1999), a disgrace (Lahman et al., 2011), infamous (Nicholls et al., 2012; Williams-Jones and Holm, 2005), highly publicized (Lopus et al., 2007), well known violations of participants rights (Fitch, 2005), celebrated horror stories (Wiles, 2012: 69), and enduring examples of abuse and deception of research participants (Librett and Perrone, 2010). Haggerty (2004) characterizes Zimbardo and Milgram, together with Laud Humphrey's Tea Room trade, as the inescapable referents in any discussion of research ethics in the social sciences. Each study raises important questions about informed consent, deception, and manipulation of subjects, all of which are issues that ethics committees continue to grapple with (Haggerty, 2004: 399).

The references in the articles listed above are usually followed by detailed accounts of both studies without any reflection on how these studies could have been performed ethically. For example, the starting point for this article, *What are Qualitative Research Ethics* (Wiles 2012), details the Milgram and Zimbardo studies without critique, even though, as the author admits, Zimbardo and Milgram 'were behavioural experiments rather than qualitative studies' (Wiles, 2012). Without a systematic examination of these studies, Wiles' (2012) textbook warns readers not to do this type of study. A consolation Wiles (2012) provides for her readers is that 'these ethical horror stories' are relatively few and far between in qualitative research (Wiles, 2012: 70). Missing in Wiles' book are qualitative studies with ethical dilemmas open for discussion, such as Ellis's *Fisher Folk* (1986), Venkatesh's *Gang Leader for a Day* (2008), Whyte's classic, *Street Corner Society* (1943), or the new qualitative sub-discipline, autoethnography (see Tolich, 2010).

Wiles (2012) is not alone in giving undisputed verdicts on these two cases. It is common to use these fall guys to legitimize ethics review, as if ethics review is essential to halt the abuse of research participants from happening again. Yet if the cases are not reconsidered prospectively, as opposed to being damned retrospectively, then any learning that can be gained from these two studies will be lost.

The retelling of these horror stories builds an ethics lore, and socializes any novice ethics committee member into these worst-case scenarios, simultaneously legitimatizing ethics review. This article does not question the need for ethics

review – the author is an ethics chair and a qualitative researcher, so understands the importance of ethics review – but he does argue that ethics training for ethics committee members requires constructive criticism rather than blanket condemnation.

The Canadian Tri-Council ethics statement has a useful CORE (Course on Research Ethics) online tutorial (<http://tcps2core.ca/welcome>) that begins instruction for novice researchers and ethics committee members by using a series of PowerPoint slides of the Stanford Prison study, which highlights how the project escalated out of control. The slide show provides an excellent overview of the study and the ethical abuse, yet as a training program designed for prospective ethics committee members, it does not inform these learners how they could have been conducted differently. The only lesson is ‘don’t do research this way’.

Even those that are critical of ethics review ‘take the oath’ (denouncing these cases), accepting the lore that condemns the horror stories. Social scientists that challenge ethics review vociferously, such as Haggerty (2004), who positions ethics review as ‘ethics creep’, recognize Zimbardo et al.’s (1973) and Milgram’s (1974) research as ethically problematic. Like Wiles (2012), Haggerty (2004) provides details of both cases without reflecting on how these studies could have been conducted differently. Yet also like Wiles (2012), once establishing the rogue status of these cases without critique, Haggerty (2004: 400) concurs, ‘notwithstanding the iconic status of the early examples, the harms that social science ethics committees routinely try to mitigate are generally of a considerably lower magnitude’.

Israel and Hay (2006: 1) claim researchers are angry and frustrated by ethics review, but also ‘take the oath’ by providing an overview of Milgram research with its mitigating features. They argue that the study was not altogether harmful: fewer than 1 percent regretted that they had participated in the research (Israel and Hay, 2006: 106). How should novice researchers read Israel and Hay’s (2006) less than 1 per cent? Is 1 per cent a benchmark in terms of acceptable confidence levels? Neither these authors, nor any of the authors mentioned above, provide steps Milgram could have taken to alleviate the harm to these study participants. As for the 1 per cent, Faden and Beauchamp (1986: 174–177) say that, although Milgram produced important results, the price was too high, and researchers must anticipate possible outcomes and describe them to potential subjects.

Recounting these two studies without elaborating how they could have done it differently embeds a sense that there is nothing to learn from these worst-case studies other than that injury befalls projects that do not eradicate harm. The other outcome is that surveillance of social science by ethics committees is warranted to avoid this recurring. But are these the only lessons available to us?

These questions are highly relevant. First, would a modern ethics committee, if given the same studies, be able to predict with any certainty the eventual harm that

resulted in either study? Second, how would these contemporary ethics committees compare with imminent psychologists of the day, who predicted that only 1 percent of the subjects would go all the way on the shock meter and thus harm the learner? Milgram (1974) found that 66 percent went all the way. Third, why do ethics committee members believe that they are capable of predicting uncertainty, when equipoise is a fundamental characteristic of research?

We literally do not know what the outcome of research will be; that is why the research is being carried out. This uncertainty makes the research unpredictable in two important ways. First, it makes the potential benefits of the research difficult to weigh; second, it also makes the potential harms to the research participant difficult to weigh. This is important because the risks of the research need to be weighed against the benefits, and given that both the risks and the benefits are often uncertain, this is very difficult (Wilson and Hunter, 2010: 51).

Prediction of harm with any certainty is not necessarily possible, and should not be the aim of ethics review. A more measured goal is the minimization of harm, not its eradication. How could either Zimbardo et al.'s (1973) or Milgram's (1974) studies be modified to minimize harm in a contemporary study? Moreover, how were studies replicating Milgram's studies being conducted under the ethics radar? The BBC News (n.d.) replication of the study, 'people still willing to torture', sought no ethics approval and suffered no backlash. Burger's (2009) 'would people still obey today?' research design and ethical approval of the research was predicated on Milgram's results. Previously, how did the Asch experiment that preceded it not warrant a similar ethical rebuke then and now?

The Asch experiments

The Asch (1956) experiments into human weakness which were what Milgram was attempting to replicate are worthy of comparison. In the Asch (1956) experiments 75 percent of subjects acted against their own interest, and yet Asch (1956) received little rebuke despite not gaining informed consent, deceiving his subjects, and producing within them acute embarrassment or shame. Fitzgerald (2005: 325) claims that it was not the harm that offends, but rather the darker side of humanity Milgram revealed.

Much of the moral outrage in relation to the Milgram research was not because the participants might have been harmed, but that the research revealed that every day, basically good people (our peers) could under certain conditions behave in ways unfathomable to most people living in comfortable circumstances, and this was just not acceptable. The idea that the other could be us was too distressful, and people did not want to hear about such things.

Those creating the ethics lore about Milgram (1974) and Zimbardo et al. (1973) do not ask: did Milgram or Zimbardo do anything well? If a similar study was

presented to an ethics committee today, what could they draw upon from the articles and books listed in the first paragraph above, to inform the present? Hindsight provides twenty-twenty vision, but any ethics committee member reviewing a similar research project using deception can only benchmark Milgram as a worst-case scenario. There is a great deal to learn about ethics review from a retrospective analysis of these two cases.

Milgram's obedience studies

How should an ethics committee review a contemporary version of the Milgram study if they want to minimize harm? As stated above, even the imminent psychologists could not predict harm. Seeking to minimize harm is best practice, and Milgram was proactive and innovative (see Sieber and Tolich, 2012: Chap. 4). He was the first to use debriefing to ensure that his subjects re-entered their world in a good frame of mind. He also used a 'quasi' reference group of fellow social scientists who followed the results of the trial with great interest, although no colleagues at Yale raised an alarm about his methodology. He used a follow-up survey to evaluate the effects of the study on his subjects. These were sound interventions, but he could have done more, and this next suggestion is standard practice on most ethics committees today.

After his post-research survey found that 1 percent regretted taking part in the study, Milgram did not follow up to attempt to alleviate the harm that those subjects suffered. In hindsight, a qualitative follow-up study with face-to-face meetings would have better captured the nuance of how subjects experienced the obedience trials, better than a pen and pencil survey. A qualitative study would most probably have led to some support for the distressed subjects. Equally, had Milgram encouraged any persons suffering an adverse event to be directed toward a New Haven counselling service, would the counselling service have been compelled to contact the researcher warning him of the mounting adverse events? Yes, the counselling service would be morally obliged to act as a reference group, independent of the research team. This ethical consideration is embedded in most Institutional Review Board (IRB)s' review processes today. How easily could this worst-case scenario be tamed?

Zimbardo's prison study

How would the 14-day Zimbardo prison study, which was stopped prematurely after 6 days, be reviewed today? How many would find fault with the way the study was reviewed then by the IRB? Few find fault with the Stanford IRB, who failed ethics 101 by not recognizing Zimbardo's conflict of interest as both a principal researcher and the prison warden. It is impossible now to predict what effect

this recognition would have had on the overall study, but had the role of prison warden been delegated to another, would the researcher have noticed the increasing level of conflict between the guards and the prisoners and its ensuing harm? This recognition would have minimized harm. Ethics committees can make errors, but these are not part of ethics committee lore.

Macquarie University's online ethics training (http://mq.edu.au/ethics_training/) session goes some way towards changing this lore. They provide details of the Zimbardo case but do so analytically. They state that Zimbardo acknowledges that the research was unethical by violating the basic Nuremberg tenet – subjects believed that, once in the prison study, they could not leave. Macquarie University reports that the study was curtailed only when an outsider, Christine Malach (Zimbardo's Graduate Student), questioned the ethics of the study. They also point out Zimbardo's conflict of interest as he was both the principal investigator and the Prison Warden. However, the online course does not take the learning that one step further. Having established that a conflict of interest took place, they do not make the link that the Stanford IRB committee approved it. The Stanford IRB is culpable in this worst-case scenario (Sieber and Tolich, 2012). Whereas Milgram's study took place in an ethical vacuum without IRB oversight (Blass, 2004: 70), Zimbardo et al.'s study was reviewed by an IRB; he gained the subjects' informed consent. The study was deemed innocuous role-playing – a bit like cops and robbers – and approved by the Stanford University IRB.

Neither the members of the human subjects research committee nor I imagined in advance that any such external authority was necessary in an experiment where college students had the freedom to stay or go any time the going became rougher than they could handle. Before the experiment, it was just 'kids going to play cops and robbers' and it was hard to imagine what could happen within a few days. It would have been good to have had advance hindsight operating (Zimbardo, 2007: 235).

The Stanford IRB failed to acknowledge a basic feature of his research design: the conflict of interest producing a domino effect on the entire study. Zimbardo's conflict of interest resulted in his failure to perceive the possible harm to his subjects (both prisoners and guards) in the rapid escalation of violence.

There is much learning to be gained from considering these studies afresh, and training members of ethics committee members to dissect them prospectively rather than blindly condemning them retrospectively would be beneficial to members and researchers. Training novice ethics committee members should involve encouraging them to treat any project submitted for review as an opportunity to assist the researcher to find more ethical ways to conduct the research. The next section reviews examples that qualitative research should acknowledge as their own and worthy of forensic review.

Examples of dubious qualitative research

This article began after the author read *What are Qualitative Research Ethics?* (Wiles, 2012), in which Zimbardo and Milgram are dismissed as celebrated horror stories without providing detailed analysis on how the studies could have been made ethical. This was an opportunity lost. Wiles (2012) then goes on to state that these two social behavioral studies were not qualitative research. Examples of qualitative research, she claims, are few and far between. Humphreys' (1973) *Tearoom Trade* is recognized as another celebrated horror story, again with no attempt to rehabilitate the rogue status. Moreover, the net of possible exemplars is not cast sufficiently to bring contemporary qualitative research like Ellis (1986, 1995), Venkatesh (2008), Whyte (1981), Vidich and Bensman (1968) or the newly emerging sub-discipline of autoethnography into the frame. Thus, qualitative research has sufficient examples of contemporary dubious ethical research that it does not need to look to psychology for its learning. Each of the following warrants forensic analysis of its ethics considerations.

The fisher folk

Carolyn Ellis's 'Emotional and ethical quagmires in returning to the field' (1995) presents an account of dealing with her own distress when she realized the pain that her study of fisher folk (1986) in a Chesapeake fishing community had caused her informants. Upon Ellis's return to the fishing village, she discovered that the research participants, many of whom she considered friends, were outraged by her book. Ellis reports that they felt the book had made them look stupid. The key informants felt that, because they could identify themselves, others would also identify them and their personal thoughts (Ellis, 1995). The learning here is that pseudonyms, usually thought to be a means to minimize harm and used to secure confidentiality, had failed to obscure identities with relational informants (Tolich, 2004).

Gang leader for a day

Venkatesh's (2008) *Gang Leader for a Day*, an ethnographic memoir of his research conducted without prior ethics approval and with little regard for ethical reflexivity, should be read by all researchers and ethics committee members. It is a fine example of how research is conducted when no ethical review is undertaken, and without implementing ethical responsibility, i.e. thinking in advance on how to protect those who are brought into the study. Venkatesh's memoir is a candid description of a study with little ethical consideration.

Over and above his enrolment at the University of Chicago, Venkatesh becomes a rogue sociologist and learns the secrets of an inner city slum high-rise tower. On one occasion he seeks to triangulate the secrets he has learned from people who

trusted him, with two gatekeepers whose official and unofficial positions dominated the residence. Venkatesh (2008: 200–201) explains the logic of his triangulation without reference to ethics.

‘Hey, you know what, I could actually use the chance to tell you [JT and Ms Bailey] what I’ve been finding,’ I said, taking out my notebooks. ‘I’ve been meeting so many people, and I can’t be sure whether they’re telling me the truth about how much they earn. I suppose I want to know whether I’m really understanding what it’s like to hustle around here ...’. For the next three hours, I went through my notebooks and told them what I’d learned about dozens of hustlers, male and female. There was Bird, the guy who sold license plates, Social Security cards, and small appliances out of his van. Doritha the tax preparer. Candy, one of the only female carpenters in the neighborhood. Prince, the man who could pirate gas and electricity for your apartment. JT and Ms Bailey rarely seemed surprised, although every now and then one of them perked up when I mentioned a particularly enterprising hustler or a woman who had recently started taking in boarders.

I finally left, riding the bus home to my apartment. I was grateful for having had the opportunity to discuss my findings with two of the neighborhood’s most formidable power brokers.

This rogue sociologist provides students of qualitative research with an opportunity to read this book-length narrative and apply ethical principles to it.

Venkatesh’s (2008) *Gang Leader for a Day* should be read with a companion text – Mitchell Duneier’s *Sidewalk* (1999), an ethnographic narrative account of people who could just as easily be documented in Venkatesh’s book. Unlike Venkatesh, Duneier practices impeccable ethical considerations and demonstrates a seamless practice of ethical research.

Autoethnography

Autoethnography as a body is a qualitative research technique that invites ethical reflection. The sub-discipline questions if it requires prior ethics approval, given that by its very name the focus of the study is the self, thus gaining ethics approval from one self is redundant (Rambo, 2007). Leaders of this sub-discipline (Ellis, 2007; Richardson, 2007) do not afford autonomy, voluntary participation, or informed consent, even for close friends (Tolich, 2010). Yet rarely is an autoethnography solely about the author. Invariably these studies focus about self’s relationships with others, and given that others are brought into the research involuntarily, autoethnographers must demonstrate their respect for persons by anticipating the needs of both the other and the self before the research writing begins (Tolich, 2010).

Students of research ethics can review autoethnography to gauge if it is the person’s story or if others should have been given the right to appear in the research.

As Tolich (2010) claims, the word ‘auto’ is a misnomer. The self may be the focus of research but the self is porous, leaking to the other without due ethical consideration. Topic choice can inadvertently harm the researcher.

Discussion

This article had two objectives. First it wanted researchers and ethics committee members to reconsider two classic psychological studies that have been damned for their ethical breaches. The article found Zimbardo’s IRB poorly advised him. The IRB failed to identify his obvious conflict of interest inherent in his roles as principal investigator and prison warden. The lesson to be learned here is that ethics committees can on occasion make ethical blunders. The lesson in Milgram’s case is in acknowledging that contemporary ethics committees would likely mitigate harm by insisting that research on sensitive topics require the participant information sheet to inform the research participant of access to counselling should they feel that the research has adversely affected them by participating in the research. Had Milgram’s participants had access to counselling, then would the counsellor have taken on a similar role that Christina Malach performed for Zimbardo, and with fresh and objective eyes curbed both researchers’ excesses?

The article’s second objective was suggesting that qualitative research look within itself for examples of dubious ethical behavior rather than focusing on the two psychological classics. Venkatesh (2008), Ellis (1995), Whyte (1981), Vidich and Bensman’s (1968) and the newest iteration of qualitative research, autoethnography (Tolich, 2010), are all worthy of forensic review. Novice ethics committee members and postgraduate students each choosing an ethical principle or concept and trawling through Venkatesh’s (2008) book forensically will be well rewarded. What makes this task easy and highly educational is that there are few ethical considerations given in the book. There is no informed consent, no voluntary participation, no minimization of harm, no proper storage of data, no sense of what to do when others are put in danger, nor any recognition that the people he studied were vulnerable persons. Identifying the concepts is only part of the forensic analysis. The next step would be to debate which of the harms was the greatest.

Declaration of conflicting interest

The author declares that there is no conflict of interest.

Funding

This research article was written with the support of a Royal Society of New Zealand Marsden Grant #UOO08185.

References

- Asch S (1956) Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs* 76(9): 1–70.
- BBC (n.d.) News people ‘still willing to torture’. Retrieved from <http://www.bbcprisonstudy.org/>.
- Blass T (2004) *The Man who Shocked the World: The Life and Legacy of Stanley Milgram*. New York: Basic Books.
- Burger JM (2009) Replicating Milgram: Would people still obey today? *American Psychologist* 64: 1–11.
- Duneier M (1999) *Sidewalk*. New York: Farrar, Straus and Giroux.
- Ellis C (1986) *Fisher Folk: Two Communities on Chesapeake Bay*. Lexington, KY: University of Kentucky Press.
- Ellis C (1995) Emotional and ethical quagmires in returning to the field. *Journal of Contemporary Ethnography* 24: 711–713. doi:10.1177/089124195024001003.
- Ellis C (2007) Telling secrets, revealing lives: Relational ethics in research with intimate others. *Qualitative Inquiry* 13: 3–29. doi:10.1177/1077800406294947.
- Faden RR, Beauchamp TL and King NM (1986) *A History and Theory of Informed Consent*. New York: Oxford University Press.
- Fitch KL (2005) Difficult interactions between IRBs and investigators: Applications and solutions. *Journal of Applied Communication Research* 33(3): 269–276.
- Fitzgerald MH (2005) Punctuated equilibrium, moral panics and the ethics review process. *Journal of Academic Ethics* 2(4): 315–338.
- Haggerty K (2004) Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology* 27(4): 391–414.
- Homan R (2006) Problems with codes. *Research Ethics Review* 2(3): 98–103.
- Humphreys L (1975) *Tearoom Trade: Impersonal Sex in Public Places*. New York: Aldine.
- Israel M and Hay I (2006) *Research Ethics for Social Scientists*. London: SAGE.
- King N, Henderson G and Stein J (1999) *Introduction – Relationships in Research: A New Paradigm*. Chapel Hill, NC: University of North Carolina Press.
- Lahman MK, Geist MR, Rodriguez KL, Graglia P and DeRoche KK (2011) Culturally responsive relational reflexive ethics in research: The three rs. *Quality & Quantity* 45(6): 1397–1414.
- Librett M and Perrone D (2010) Apples and oranges: Ethnography and the IRB. *Qualitative Research* 10(6): 729–747.
- Lopus JS, Grimes PW, Becker WE and Pearson RA (2007) Effects of human subjects requirements on classroom research: Multidisciplinary evidence. *Journal of Empirical Research on Human Research Ethics* 2(3): 69–78.
- Milgram S (1974) *Obedience to Authority*. New York: Harper.
- Nicholls SG, Brehaut J and Saginur R (2012) Social science and ethics review: A question of practice not principle. *Research Ethics* 8(2): 71–78.
- Rambo C (2007) Handing IRB an unloaded gun. *Qualitative Inquiry* 13: 353–367.
- Richardson L (2007) *Last Writes: A Daybook for a Dying Friend*. Walnut Creek, CA: Left Coast Press.
- Siebert JE and Tolich MB (2012) *Planning Ethically Responsible Research* (Applied Social Research Methods, Vol. 31). Thousand Oaks, CA: SAGE.
- Tolich M (2004) Internal confidentiality: When confidentiality assurances fail relational informants. *Qualitative Sociology* 27(1): 101–106.

- Tolich M (2010) A critique of current practice: Ten foundational guidelines for autoethnographers. *Qualitative Health Research* 20(12): 1599–1610.
- Williams-Jones B and Holm S (2005) A university wide model for the ethical review of human subjects research. *Research Ethics Review* 1(2): 39–44.
- Venkatesh S (2008) *Gang Leader for a Day: A Rogue Sociologist Takes to the Streets*. New York, Penguin.
- Vidich A and Bensman J (1968) *Small Town in Mass Society: Class, Power and Religion in a Rural Community*. Princeton: Princeton University Press.
- Whyte WF (1981/1943) *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: University of Chicago Press.
- Wiles R (2012) *What are Qualitative Research Ethics?* London: Bloomsbury Academic.
- Wilson J and Hunter D (2010) Research exceptionalism. *The American Journal of Bioethics* 10(8): 45–54.
- Zimbardo P (2007) *The Lucifer Effect*. London: Random House.
- Zimbardo P, Haney C, Banks W and Jaffe D (1973) The mind is a formidable jailer: A Pirandellian prison. *New York Times Magazine*, 8 April: 38–60.