
Controversial Psychological Research Methods and Their Influence on the Development of Formal Ethical Guidelines

Katie Youngpeter

Metropolitan State College of Denver

Abstract: Presently, researchers must give special ethical concerns to monitor participant treatment when they are utilizing controversial research methods. Such methods include, but are not limited to deception, coercion, and potentially inducing psychological harm. Throughout the course of the history of psychology as a science, a number of different ethical guidelines have been developed and employed in a variety of ways. The interest in developing guidelines for psychological research was stimulated by several studies, particularly Watson and Raynor's research on Little Albert (1920), Milgram's interest in obedience (1963; 1964), Humphrey's naturalistic observations of tearrooms (1973), and Zimbardo and colleague's prison experiment (1973). Each of these studies has in part contributed to the current ethical standards for psychology research on humans in the United States. Current APA guidelines are briefly discussed.

Whenever research involves human participants, researchers must give special ethical concerns to monitor participant treatment. In general, most people would not support the idea of suppressing an individual's civil rights for the sake of expanding our knowledge of human nature; on the other hand, researchers often are unable to predict the results of their manipulations, thereby making it unreasonable to reprimand them on the basis of their findings alone. Moreover, researchers must be particularly cautious when using deception, coercion, and potential psychological harm, as well as consider voluntary participation and overall confidentiality. Indeed, the aforementioned factors were the main focal points for the

development of formal ethical codes for psychological research (Kimmel, 1996), and there were several important studies that brought these issues to light (e.g., Haney, Banks, & Zimbardo, 1973; Humphreys, 1975; Milgram, 1963; Watson & Rayner, 1920). These studies are important to review and appreciate because each has ultimately contributed, in part, to the current guidelines of the American Psychological Association (see APA, 2002)

Ethics is the study of proper action (Gravetter & Forzano, 2006). Changes in research interests over the past several decades have substantially influenced the evolution of ethical standards, guidelines and codes for research within the field of psychology.

Compared to other sciences, psychology is a relatively new field, so, understandably, the concern for creating guidelines for researchers is also a novel enterprise (Gravetter & Forzano, 2006; Kimmel, 1996). However, the general public, as well as researchers, tended to be unconcerned with the creation of formal ethical codes for this field until the growing popularity of behaviorist studies on people, and later, socio-psychological research. In particular, researchers had little regard for the ethics revolving around how behavioral traits emerge in uncomfortable situations (Haney et al., 1973; Humphreys, 1975; Milgram, 1963; Watson & Rayner, 1920).

During psychology's infancy, researchers were indifferent toward ethics because they often based their studies on introspection rather than social manipulations (Hergenhahn, 2001). Initially, psychophysics gained popularity in the early 1900s and was characterized by the simple recording of participant responses after exposing them to various types of stimuli, such as colored lights or slight pricks on the skin (Hergenhahn, 2001). While participants may not have known what exactly was being studied, they were aware that they were participating in research, communicated directly with the researcher about their experiences, and usually did not have to be concerned about the possibility of feeling humiliated, embarrassed, or otherwise harmed due to the tests that were administered to them (Hergenhahn, 2001).

With the rise of behaviorism in the 1920s came a new interest in environmental factors and their influence on behaviors. For example, a researcher may have been curious about how reinforcement and punishment could respectively lead to likes and dislikes (Hergenhahn, 2001). Watson was one such researcher, and

the major focus in his famous study on "Little Albert" was how human emotional responses are created (Watson & Rayner, 1920). Watson and Rayner hypothesized that fear could be conditioned in a baby, which suggested that fear of specific objects or events was not innate and that fears were not the results of adverse sexual experiences (the popular Freudian theory at the time [Watson & Rayner, 1920]). Watson and Rayner created an experimental method to test this, and therein laid the ethical concerns with their work.

Using Albert, an 11-month-old boy with a very relaxed demeanor, Watson and Rayner (1920) attempted to create a novel fear of a white rat. Albert initially was not afraid of the rat, and indeed was very curious about it; but, as soon as Albert touched the animal, Watson hit a large metal bar which yielded a loud, unpleasant noise that was found beforehand to upset the baby. It took only seven rat-sound pairings for the rat alone to elicit an extreme response in the child, characterized by withdrawal and crying. The association persisted strongly for one week, and Albert was again tested after about one month – the fear was still present, albeit somewhat weaker. The response had also generalized toward other white, furry objects, including a rabbit, dog, fur coat and Santa mask. However, Albert was still content to interact with dissimilar objects such as toy blocks while in the laboratory environment (Watson & Rayner, 1920).

Watson and Rayner's (1920) methodologies helped facilitate the growing concern for participant wellbeing, particularly for those unable to give consent or even assent, such as infants (Baumrind, 1964; Fischer, 2005). Despite the child's obvious discomfort, Watson and Rayner continued with the study. To make

matters even more controversial, Watson and Rayner failed to conduct the final part of the experiment, which involved extinguishing the fear, because Albert's mother had to move. Watson and Rayner responded to this unfortunate outcome by assuring the reader that Albert, like all babies, would have likely developed a similar association anyway out in the "rough and tumble" world (Watson & Rayner, 1920, p. 2). While fearing fuzzy, white objects may not have adversely impact Albert's life, critics began to feel a rising concern about experimenters possibly instilling serious, permanent psychological damage in the very people they are relying on for their data (Gravetter & Forzano, 2006). Despite these criticisms of his research, however, the APA offered Watson high praise in 1957 for revolutionizing modern psychology (Skinner, 1960).

Although most psychologists assumed that all researchers would be reasonably responsible for their participants' wellbeing (Gravetter & Forzano, 2006), there were important exceptions in the 1940s during World War II. These war crimes, isolated to Germany and Japan, involved disturbing manipulations of prisoners that left them severely psychologically or physically damaged or even resulting in death (Fischer, 2005). Under normal circumstances, one would hope that a researcher would do all possible to keep his participants comfortable – if anything to not instill any discouragement toward research. War crimes are unique, however, because the participants are viewed as *subhuman* and thus treated as such to minimize feelings of guilt (Fischer, 2005). While these crimes tended to involve physiological interests rather than psychological ones, they are still important to psychologists because they ultimately led to the creation of the Nuremberg Code in 1947 after the

researchers were tried for their crimes in Nuremberg, Germany (Fischer, 2005).

The Nuremberg Code (Kimmel, 1996) was created as a set of guidelines specifically for researchers in the biomedical field, but psychologists adopted it as their own because the Nuremberg Code focuses on the treatment of live participants (Kimmel, 1996). Moreover, during the 1950s, there was a rising interest in how people interact with others, as well as how they respond in various social situations. This interest was stemmed by the widespread adoption of Functionalism, a school of thought which emphasized studying cognitive processes (Hergenhahn, 2001). Because these processes are almost always influenced by our environment, psychologists began asking questions about what factors are important in guiding behaviors. This resulted in an interest in ethics for this field because, as previously mentioned, methods such as deception can be unclear in terms of whether or not they harm participants (Gravetter & Forzano, 2006; Kimmel, 1996). If anything, social psychologists had to be diligent in their use of informed consent and debriefing procedures.

After the development of the Nuremberg Code, almost all psychologists accepted and followed its guidelines, but it was nonetheless informal and did not necessarily dictate how a study should be conducted (Gravetter & Forzano, 2006; Kimmel, 1996). It was not until the late 1960s that all institutions that conducted funded research on humans were required to have an Institutional Review Board (IRB; Pope, 1992) and not until 1978 that a set of guidelines, called the Belmont Report, was adopted (Fischer, 2005; Gravetter & Forzano, 2006; Pope, 1992). An IRB is a diverse group of individuals who review potential

research topics and determine if they meet the guidelines as established by the overseeing ethical board (Gravetter & Forzano, 2006). If a problem arises, the researcher is contacted and asked to revise his or her methodology. The rules arising from the Belmont Report have not changed drastically over time, though, mainly because it is nearly impossible to impose strict rules on psychological research; there will always be important exceptions where stringent guidelines would limit the usefulness of a research study (APA, 2002; Kimmel, 1996).

Despite growing efforts to protect participants, a number of controversial studies emerged. Each controversial study was considered ethically sound according to APA, the established research overseer at the time (Kimmel, 1996) and yielded fascinating results. However, upon being revealed to the public, the results and/or methods of these studies called for further revisions of ethical rules. These studies include, but certainly are not limited to, the following: Milgram's numerous studies on obedience (1963, 1965); Humphreys' field studies on "teahouses" (1975); and the Stanford study on prison life (Haney et al., 1973).

MILGRAM'S OBEDIENCE STUDIES

Stanley Milgram was a social psychologist who was intrigued by the idea of negative obedience, where people are guided by an authority figure to commit immoral acts (Milgram, 1963). He was inspired mainly by war crimes, such as those that led to the Nuremberg Code, and wanted to gather clues about what makes a person obey so blindly – is it an innate character flaw? Or, can anyone be induced to hurt or even kill someone else simply due to authoritative demands? Milgram devised an ingenious setup to test obedience in humans (see Milgram,

1963; 1965). After gathering a sample of 40 males with various career backgrounds, he told the participants that they were participating in research on the effects of punishment on memory. In each session, the participant was always assigned the role of the teacher while a confederate played the learner and was ultimately strapped to an electrical chair that could be controlled by the teacher in another room. The two communicated via an intercom system (Milgram, 1963).

Although all participants believed the setup to be genuine, the learner would never receive any actual shocks. The participant's job was to read to the learner a list of words and wait for him to repeat them in order. If he was incorrect or did not respond he was given a shock. Each time this occurred, the participant had to administer a fake shock that was 15 volts more intense than the last (the maximum was 450 volts). The shock machine was labeled with phrases such as "Danger: Severe Shock" (Milgram, 1963, p. 373). If a participant expressed any concern during the session, the experimenter urged him to continue by saying, for example, "It is absolutely essential that you continue" (Milgram, 1963, p. 374). But, participants were told at the beginning of the experiment that they were free to leave whenever they wished to do so.

Much to the experimenters'—and later, the public's surprise—30 of the participants continued to follow the procedure and administer shocks until 450 volts was reached. Although Milgram expressed concern for the wellbeing of his participants, as some were showing extreme agitation, he decided not to terminate the study on the basis that each prior participant seemed to recover relatively well after his session ended. Additionally, they were each told afterward about the

nature of the deception and that they did not actually harm anyone, plus they were invited to discuss the study with the confederate (Milgram, 1963; 1964; 1965). Most participants stated in later questionnaires and interviews they were *glad* to have participated because they too were given insight on the dangers of obedience and that they would evaluate their actions more thoroughly in the future (Milgram, 1963; 1965).

Regardless of the apparent lack of permanent harm, Milgram's obedience studies were disconcerting to many psychologists, such as Baumrind (1964). Baumrind in particular felt that Milgram never should have deceived his participants, that he put undue pressure on them to stay (i.e., coercion), and that the results were not very generalizable because they took place in a laboratory setting (Baumrind, 1964). Milgram responded by saying that Baumrind and others likely would not have been concerned with the ethics of his study if the results had been different (in other words, if the participants had been more disobedient) (Milgram, 1965). Milgram had no preconception of these results and it is clear that he was concerned for the participants, but he also decided that it was critical to continue because the results were so groundbreaking and insightful (Milgram, 1963). His decision was supported by the participants, as only 1.3% expressed negative feelings about their experiences (Milgram, 1965).

Regarding Baumrind's second criticism, Milgram noted that participants were paid even before the session began (see Milgram, 1963; 1965). They all gave their informed consent as well. Milgram chose not to reveal the true hypothesis because it could have biased the results; indeed, of all the people he explained the study to outside of the laboratory, not one

imagined himself or herself giving the 450 volt shock (Milgram, 1963). Because he was unable to conceive of any harm occurring to the participants, he believed that his methods were ethically sound (Milgram, 1965). Finally, Milgram countered Baumrind's third claim by saying that his results are all the more surprising because they were gathered within a controlled setting; this effect is likely even stronger in "real world" situations (Milgram, 1965).

Because of the controversy instilled by Milgram's studies, the APA formulated its own ethical guidelines and formalized them in 1973 (Gravetter & Forzano, 2006; Kimmel, 1996), although the ethics committee had been established over 10 years prior to Milgram's first obedience study in 1963 (Gravetter & Forzano, 2006). The guidelines (which are continuously being revised [APA, 2002]) consisted of 10 broad standards that attempted to cover most types of research that a psychologist may be interested in (Kimmel, 1996). The major themes in these guidelines includes: a) reducing harm to participants, b) obtaining informed consent, c) if deception is used, it must be justified, d) ensuring confidentiality, and e) debriefing participants (Gravetter & Forzano, 2006). Additionally, psychologists must be aware of any federal, state and local regulations that may impact their research.

HUMPHREYS ANALYSES OF TEAROOMS

Like Milgram, Humphreys used deception in his research on "tearooms" (public restrooms where men meet to engage in sexual activities; see Humphreys, 1975), but rather than actively lie to participants, he simply withheld information from them. Because the study was so intensive and required lengthy and detailed background information, Humphreys

published a book on his tearoom research in 1970 (and republished it in 1975 with comments on ethics). Essentially, his goal was to secretly observe men engaging in sexual acts in tearooms and gather information about them to determine what their home lives were like. Since tearoom activities are illegal, the first ethical problem with this study involved Humphreys' refusal to contact the police about the evidence he collected. However, whether or not his decisions were right or wrong depends largely on individual interpretations; he was attempting to study this unique population and had no intentions of having his participants arrested (Humphreys, 1975).

After carefully choosing adequate locations to perform his observations, Humphreys (1975) discovered that it was difficult to be accepted in the tearooms; eventually, though, he was able to settle into his role as "watchqueen" and would stand by the windows of the bathrooms and alert the others of approaching people or police cars. He spent many hours in the tearooms observing behaviors (sexual and otherwise) and noting descriptions of the individuals he encountered, the make and model of each of their vehicles (if they drove), and their license plate numbers. One year later, he changed his appearance so he would not be recognized, located each participant's residence, and conducted personal interviews with many of them. The interviews revolved around a survey that was devised for a separate study by one of Humphreys' colleagues and included questions on family history, socioeconomic status, personal health, political views, marital relationships, sex, and other topics (Humphreys, 1975).

Humphreys (1975) discovered that many of the men involved in the "tearoom trade" were happily married yet would be horrified if they realized

someone knew of their secret. This highlights the second major ethical concern of this study – voluntary participation along with confidentiality. Even though Humphreys kept all participants' personal information in a locked safe-deposit box and destroyed the information after the interviews (Humphreys, 1975), the fact that he used this data to track many of them down several months later sparked considerable concern (Kimmel, 1996). It is likely, though, that the concern was based less on withholding participant data for a length of time than on the fact that the study focused on such a controversial topic with legal ramifications.

STANFORD'S PRISON EXPERIMENTS

Zimbardo and his research assistants, Haney and Banks, were also curious about criminal behavior, but their study revolved around the prison experience (Haney et al., 1973). Similar to Milgram's obedience hypothesis (1963), Zimbardo and his assistants wanted to know if the actions of prisoners and guards are the result of personality flaws or the prison situation itself. The experimenters created a simulated prison environment at Stanford University and carefully selected 24 middle-class, white male college students who scored high on assessments of stability and maturity and low on crime involvement. They randomly assigned each person to be a prisoner or a guard and gave them considerable monetary compensation for each day they stayed in the study (it was originally planned to last for two weeks). Participants were given no guidelines for how to behave, with the exception of the guards who were told not to use physical abuse under any circumstances (Haney et al., 1973).

Within only two days, the participants literally became the roles they were assigned. The guards began

to act very harshly and sometimes even cruel to the prisoners, forcing them to earn their privileges with good behavior; the definitions of good behavior varied from guard to guard, however. This strenuous situation led the prisoners to become very passive, quiet and obedient, and they also exhibited learned-helplessness. While no one was physically hurt during the study, a few of the prisoners displayed extreme emotional reactions that warranted termination of the study after only six days. After the experiment was over, the researchers went to great lengths to inform participants of the nature of the study as well as allow them to communicate with each other about their feelings (Zimbardo, 1973). As in Milgram's studies, none of the participants considered, in retrospect, that the study adversely impacted their lives, and many found the experience to be beneficial. Furthermore, after psychological examinations, none were found to have suffered lingering psychological problems (Zimbardo, 1973).

The Stanford prison experiment elicited a large amount of concern over ethics in psychology (Kimmel, 1996), mainly because many of the participants became severely distressed during the study. Some argue that the study should have been terminated sooner or even not conducted at all, but as in Milgram's case, Zimbardo was unaware that such striking results would be observed (Haney et al., 1973; Zimbardo, 1973). Interestingly, although the prisoners were told that they could withdraw from the study at any time, only one purposefully attempted to do so. Out of fascination, or perhaps because they too began to accept their temporary roles, Haney and colleagues chose to continue the study until it became absolutely necessary to stop. Zimbardo later responded to criticisms by urging IRBs

to not ban this type of research, but rather to monitor it closely (such as by utilizing a "metaresearcher" to oversee the researchers and ensure that everyone in the study remains comfortable) (Zimbardo, 1973). Zimbardo further noted that they were given permission to conduct the study and that the APA later found it to be within its ethical guidelines.

CONCLUSION

The arguments used against the methods employed by Watson and Rayner (1920), Milgram (1963; 1964), Humphreys (1975), Haney et al., (1973) and others are valid because everyone has different ways of evaluating benefits and risks. This discrepancy of opinions has consistently been a problem with formal psychological research ethics because the APA ethics board cannot possibly predict and outline the potential risks of every situation an experimenter may devise. Instead, IRBs rely on general rules and their own subjective "best judgment" of each unique research project (APA, 2002; Gravetter & Forzano, 2006; Kimmel, 1996). Therefore, IRBs are very cautious when evaluating projects that utilize deception as a central concept (Gravetter & Forzano, 2006). These cases are even more deeply scrutinized if participants may become traumatized or otherwise harmed psychologically, or if the research would dissuade them from participating in future studies (Gravetter & Forzano, 2006).

Presently, researchers of psychology in the United States must adhere to the guidelines outlined by the APA (see APA, 2002). The basic principles of these guidelines have changed very little since their creation in the 1960s, but much detailed information has been added (APA, 2002). But, because they are technically only guidelines and not strict rules, researchers are encouraged to read these principles

thoroughly and use them in tandem with their own judgment to determine if their study designs are ethical. If a researcher can properly justify a potentially unethical study, it may pass the IRB review (Gravetter & Forzano, 2006). Justifications for such designs can be highly variable, but they generally involve the potential for attaining important knowledge to help a large number of people in the population or yield groundbreaking information to be used in future studies (APA, 2002). Further, researchers must always be cautious when studying special populations, such as infants, prisoners, or individuals with medical problems because these persons may be particularly sensitive to certain treatments (such as drugs) or controversial methods such as deception (Gravetter & Forzano, 2006).

Clearly, the development of psychological research ethics has been difficult because on the one hand, participants should be treated with respect, dignity and kindness; conversely, it is sometimes necessary to use tactics such as deception or to instill mild psychological harm in order to gather useful information about the nature of cognitive processes and how they influence behavior. This task is made especially difficult when considering infants, special populations, individuals with psychological disorders, and animals. Even though the four aforementioned studies had particularly significant impacts on the development of formal ethical guidelines, there are nonetheless hundreds (or possibly thousands) of other studies that have been important in the development of ethics (Fischer, 2005; Gravetter & Forzano, 2006; Kimmel, 1996). Nonetheless, the studies by Watson and Rayner (1920), Milgram (1963; 1964), Humphreys (1975), and Haney et al (1973) outline how controversial

methods can yield unexpected yet useful and insightful results.

REFERENCES

- American Psychological Association (2002). *Ethical principles of psychologists and code of conduct*. Retrieved March 18, 2008, from <http://www.apa.org/ethics/code2002.html>
- Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's "Behavioral study of obedience." *American Psychologist*, 19(6), 421-423.
- Fischer, B. A. (2005). A summary of important documents in the field of research ethics. *Schizophrenia Bulletin*, 32(1), 69-80.
- Gravetter, F. J. & Forzano, L. B. (2006). *Research methods for the behavioral sciences*. (2nd Ed.). Mason, OH: Thompson.
- Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1, 69-97.
- Hergenhahn, B. R. (2001). *An introduction to the history of psychology* (4th Ed.). Belmont, CA: Wadsworth.
- Humphreys, L. (1975). *Tearoom trade: Impersonal sex in public places: Enlarged edition with respect on ethical issues*. Chicago: Aldine.
- Kimmel, A. J. (1996). *Ethical issues in behavioral research*. Cambridge: Blackwell.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371-378.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, 19(11), 848-852.
- Milgram, S. (1965). Liberating effects of group pressure. *Journal of Personality and Social Psychology*, 1(2), 127-134.

- Pope, K. S. (1992). Ethical dilemmas encountered by members of the American Psychological Association: A national survey. *American Psychologist*, 47(3), 397-411.
- Skinner, B. F. (1960). John Broadus Watson, behaviorist. *Science*, 129, 197-198.
- Watson, J. B. & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, 3, 1-14.
- Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition*, 2(2), 243-256.