A critical review of replicating milgrim

With the present day's strict ethical guidelines and rules in psychology, it would appear almost impossible that one of the most valuable experiments ever carried out in social psychology would be approved to be replicated. Reason being the amount of deception used and the emotional stress inflicted on participants. Nonetheless, one of Stanley Milgrim's (1974) most famous experiments on obedience to authority has been replicated by Jerry Burger (2009).

To give a brief description of the experiment, Milgrim deceived participants into believing they were administering electric shocks to another participant (a confederate). Milgrim found that with the encouragement of the experimenter (authority figure), participants (65%) were willing to administer shocks right up to 450-volts. However the confederate did not actually received any electric shocks. This startling discovery gave rise to the question, what makes people obey orders from authority figures, at the cost of willingly inflicting pain on others? The answer to this question is especially useful to allow us to understand real world occurrences of such obedience, for example the actions of German Soldiers killing nearly six million Jews during the Second World War under the Nazi regime. Milgrim (1974) himself described this phenomenon as the " agentic shift" in which individuals characterise obedience amid the transference of personal responsibility to the authority figure giving orders. It also stresses the importance of situational factors in influencing behaviour. Forty-five years later, Burger (2009) carried out a partial replication of Milgrim's (1974) study, interested in one particular question, " Will people still obey today?".

Obedience studies have raised a great deal of discussion regarding ethical issues. It has been argued that the potential harm and stress this type of experiment can inflict on participants is not a justified (Burger, 2009). Although Milgrim (1974) study administered follow up questionnaires to his participants, which indicated that the majority of participants were happy to have taken part and believed the study was justified. Today's ethical standards for the treatment of participants are in set in stone, making similar procedures out of bounds (Elms, 1995). Therefore Burger made some crucial changes to Milgram's (1974) procedure, making it ethically possible to replicate, partially, the classic obedience study.

Burger noted that a crucial pivoting point in Milgrim's study in which the participant presses the switch labelled 150-volts causing the learner to scream grinding the study to a halt. He described this as the point of no return, for the reason that 79% of participants who passed this point carried on to the 450-volt switch. Burger therefore believed that stopping the experiment at the 150-volt switch would be the solution to combat ethical issues, as this would undoubtedly expose participants to less stress. On the other hand, this factor may have been a huge limitation of the study. Twenge (2009) argued that it is possible that today, participants would not administer shocks above 150-volts. Yet, as the level of electric shocks was capped at 150-volts, we will never know if this is the case due to ethical restrictions. In addition, Twenge (2009) also argued that today's generation has shifted compared to that of the 1960's, American culture assert more emphasis on individualism and assertiveness, placing the rights of the self above the importance of duty. It is therefore possible that in the present day,

obedience has declined, and that participants would not continue to administer shocks up to 450-volts.

The author of the article used additional “ safeguards” to ensure the participants were treated ethically. Care was taken not to use too much deception, although some was necessary. Participants were screened to remove individuals who may be familiar with the study, those who had previously studied psychology (more than two classes) were excluded. Participants were also screened for those who may be more susceptible to stress and emotional distress. Those who answered yes to a number of questions regarding mental health and stress were excluded from the study. Participants were also administered a number of scales which assessed; empathy, the severity of anxiety, the desirability of control and severity of depression. After completion, participants were interviewed by a clinical psychologist, who was instructed to further identify anyone that may possibly have a negative reaction to the experiment. Out of the 123 participants, 47 (38. 2%) were excluded for reasons unknown (due to confidentiality agreement between the interviewer and interviewee). The remaining 76 was then further reduced to 70, due to participants expressing awareness of Milgram’s obedience research. It is evident that the thorough screening process is a disadvantage, as it has dramatically reduced the sample. The original size of the unscreened participants is unknown, it can only be estimated that over half of the participants were excluded after screening. An obvious weakness of this small sample is that it becomes less generalizable to a wider population. It is questionable whether, in an attempt to make the experiment more ethically sound, Burger has in reality sacrificed

the validity of his findings. Even though it remains unknown whether this study would have passed the Institututional Review Board (IRB) without these procedures in place. In addition participants were told on multiple occasions they could withdraw from the experiment at any point and still received payment ($50). Although this was contradicted buy the use of the experimenter's prods, ordering the participant to carry on when they showed concern. Another safeguard was to only administer the participant with a 15-volt sample shock (Milgrim administered a 45-volt shock) as proof that the generator was real. Furthermore, the participants were informed the shock generator was not real almost immediately after the experiment ended.

Although it seems Burger has taken some extraordinary measures to prevent participants being exposed to emotional stress, it is necessary to discuss the limitations and resulting hindrance on the findings. The screening process removed potential participants who are familiar with Milgrim's studies (30% of the sample were excluded), which the reasoning is understandable, however this could have removed a potentially disobedient population. As a result, Burger's rate of obedience may be higher than if such a screening process had not taken place (Elms, 2009). Additionally, those participants excluded due to their susceptibility to emotional stress may have added to the disobedient population. In giving only a mild electric shock to the participants and not the 45-volt shock as used by Milgram may have led them to believe that the shocks were generally not painful, this may have influenced the participants decision to continue to the final 150-volt shock. Therefore the large differences between populations and the change in

voltage administered to the participants, have caused Milgrim’s and Burger’s experiments to be incomparable.

The methodology used by Burger was similar to the template laid out by Milgrim’s study. Burger’s experiment had two conditions. The first ‘ base’ condition was similar to Milgrim’s original experiment, the second ‘ modeled refusal’ condition included a confederate who refused to administer shocks . Participants were assigned randomly to each condition but an attempt was made to keep the gender ratios equal. Those assigned to the base condition were taken a room and introduced to the experimenter (a White Caucasian man in his mid 30’s) and the confederate (also a White Caucasian Man, who was in his mid 50’s). The experimenter and confederate were chosen due to their resemblance of those used in Milgrim’s study. The experimenter informed the participants that the research was regarding the effects of punishment on learning. After signing consent forms and being paid $50, the participants were escorted to an adjacent room. The confederate was secured to a chair, and an electrode connected the shock generator was then placed on his wrist. The experimenter placed a list of 25 word pairs in front of the confederate to memorise, and told him that the teacher would read the first word and then four possible answers. The confederate was told to answer by choosing one of four buttons. For every wrong answer, the confederate was told that he would be given a shock. The confederate then informed the experimenter of his slight heart condition and expressed concern about the danger of the shocks. After reassuring the confederate the experimenter and participant returned to the adjacent room and the participant was sat in front of the shock generator. The participant was

informed how to use the machine and handed the list of word pairs, identical to Milgrim’s stimuli. For every word pair the confederate answered incorrectly, the participant had to administer and an electric shock, increasing by 15-volts every time. If the participant expressed concern, the experimenter used prods, such as “ Please go on” and “ It is absolutely essential you continue”. After the 75-volt shock, a pre-recorded grunt noise was played to give the illusion to the participant that the learner was in pain. The grunt became louder as the shocks increased, until the 150-volt shock was pressed, in which the participants hear the confederate yell and complain about his heart condition. The experiment ended after this point and the participant was immediately told the shock generator was fake . To assure the participant the confederate then entered the room to confirm he was not harmed. The participant was then debriefed.

In the procedure for the modelled refusal condition was the same as the base condition with a few differences. Two confederates were used, one played the role of the male learner and the other played the role of a second teacher, posing as a participant. Both participants were sat in front of the shock generator in this incident. It was fixed so that the confederate was always Teacher 1 and the participant was always Teacher 2. Teacher 1 began the procedure, by reading the list of words to the learner and giving the electric shocks. After the 90-volt shock the pre-recorded grunt was played and the confederate refused to continue. The real participant was then asked by the experimenter to continue from where the confederate had finished.

Burger’s procedure appears very thoroughly controlled and almost identical to Milgrim’s procedure. There are important strengths to be noted, but also weaknesses to be pointed out. The limiting of shock level to 150-volts is probably one of the key factors which allowed the replication to pass ethical approval. Although Burger stated that this solution would cause less stress to participants, he also carried out intense screening cause less stress. Miller (2009) argued that the 150-volt solution may have been enough to combat ethical issues, and that extreme screening would not have been necessary. The base condition was highly controlled to prevent confounds influencing the results and was almost identical to Milgram’s study. Thus making it comparable. The modeled refusal condition was a reasonable variation, as a number of studies on obedience and conformity have seen the effects of witnessing disobedience to social pressure (Asch, 1956; Milgrim 1974). However, the lack of significant findings can only suggest the manipulation was not properly applied.

The results of Burger’s study are somewhat confusing. The results demonstrated that in the baseline condition, 70% of participants would continue after pressing the 150-volt switch, until stopped by the experimenter, compared to Milgrim’s 82. 5%. However this was not statistically significant. In the modelled refusal condition, 63. 3% would continue beyond this point. This also was not significant. Burger also states there were no real gender differences, although slightly more women than men continued to the end in both conditions. Once again, this finding was not significant. Burger also measured participants’ rates of desire for control and empathetic concern, there was no significant difference found. However

it was found that within the base condition, those participants who stopped before the 150-volt shock, scored significantly higher on the desire for control measure compared to those participants who continued, but this was not found for the modelled refusal condition. This shows little difference between the two conditions, only 6. 7% fewer participants disobeyed in the modelled refusal condition, which Burger put down to the powerful force of situational factors. This finding is similar to a manipulation Milgrim carried out (1974) in Experiment 17. In this manipulation three teachers were used, Teacher 1 administered the word pairs, Teacher 2 stated whether the learner was wrong or right and Teacher 3 was the real participant who had to administer the shocks. After the 150-volt shock Teacher 1 refused to continue (in a dramatic fashion), as did Teacher 2 at the 210-volt shock, until Teacher 3 was told to continue alone. Milgrim found that 7. 5% refused to carry on when Teacher 1 quit, 30% refused after Teacher 2 quit, and a surprising 10% continued to the final switch. Although this finding was similar to Burger's modelled refusal condition, it is difficult to compare, not only due to the voltage difference but the number of confederates used. It would have been quite feasible for Burger to replicate Milgrim's Experiment 17 using extra confederates, to truly be able to infer that situational factors are an important aspect. Miller (2009) stated that there is no possible way to compare Milgrim and Burger's experiments because only some aspects of Milgram's study were duplicated, and at the same time vital changes were made. This may be the reason for the lack of statistical significance.

In comparing Burger's sample to Milgrim's sample, it is necessary to note that Milgrim's participants were all male, whereas Burger used more female

than male participants (29 men and 41 women). Realistically this shows; 82. 5% of Milgram's sample obeyed, whereas only 62. 7% of Burger's male sample obeyed (Twenge, 2009). Twenge (2009) also noted that if the percentages are reversed: 17. 5% of Milgrim's participants disobeyed, compared to 33. 3% of Burger's male participants disobeyed, almost twice as many participant disobeyed in Burger's sample. Therefore, this suggests that obedience has decreased by half in the past 45 years. The sample also differed in terms of ethnicity. Burger's sample was ethnically diverse, 12. 9% were Hispanic, 27. 2% were Asian Americans and over half were White. It is unlikely that Milgram used an ethnically diverse sample, as according to Twenge (2009) The U. S. Census did not collect data from ethnic groups until 1980. This is also clear from Milgrim's film recordings of the experiment. The importance of culture differences must be noted. Asian Americans score lower on certain individualistic personality traits such as self-esteem and narcissism. This may be due to the fact that Asian cultures place more emphasis on collectivism and less emphasis on individualism. This may cause Asian Americans to obey the experimenter at a higher rate compared to other ethnic groups, therefore this may have raised the level of obedience.

To conclude, Burger's partial replication of the classic Milgrim study seems in essence a well thought out, well controlled attempt to recapture one of the most ethically and methodologically controversial studies of obedience in social psychology. However due to the amount of variation in the methodologies and samples of both studies, a direct comparison is difficult to make. Burger must be praised for his efforts as he has undoubtedly paved

that way for psychologists to revisit the possibility of conducting research in this area using Burger’s design. In addition, Burger has also examined a previously unexplored contributory variable, personality. Both individual differences in personality and situational factors can affect obedience. This experiment has given us a glimpse of the amount of, or lack of changes in behaviour since the 1960’s. However due to strict ethical guidelines today compared with recent decades, we will never truly be able to replicate Milgrim’s experiments, and therefore never truly know how obedient individuals are today.