Spring 2019

A Test of Obedience or Patience? A Modified Replication of "Nothing by Mere Authority" by Haslam et al. (2014)

John J. Machen
Bard College

BARD

A College of the Liberal Arts and Sciences *Division of Science, Math, and Computing*

A Test of Obedience or Patience?
A Modified Replication of "Nothing by Mere Authority" by Haslam et al. (2014)

Senior Project Submitted to
The Division of Science, Mathematics, and Computing
of Bard College

By
John Machen

Annandale-on-Hudson, New York
May 2019

Acknowledgements

Stuart, I give my abundant thanks to you for taking me on as your advisee. I have never known an educator as forthrightly eager and delighted to build a relationship with his students. I have found that tremendously encouraging. I will miss you.

Table of Contents

Abstract

After Milgram's infamous experiments and their subsequent ethical critiques, social psychologists have been challenged to search for ways to learn more about the psychology of destructive obedience while still holding true to modern IRB standards of participant protection. One of the ways in which this has been attempted is through the invention of newer and safer paradigms of the Milgram task, perhaps the best known of which would be Jerry Burger's (2009) partial replication of Milgram's voice-feedback experiment. Five years later, a team of researchers devised a completely new obedience task, the simplified premise of which was to have naive participants take part in an online survey made to appear as a study in cognitive science (Haslam et al. 2014). The survey asked participants to describe images depicting groups of people with their choice of one of five listed negative words (e.g. "aggressive" or "brutal"). Initially this task was not at all abrasive in nature as the images depicted scenes such as Nazis at a rally, however the images shown incrementally became more positive in nature depicting scenes such as smiling children in a classroom; by this point, participants were meant to view the task as difficult to continue. This paradigm is discussed by Haslam et al (2014) in terms of an analogue to Burger's. Haslam et al. (2014) report that as many as 53% of 151 participants (across 4 conditions) stopped the task, which sounds suggestive that these participants indeed found the task's completion difficult. Though as these participants took the task online, it is unclear what their motivations were for stopping—could such a task truly present them with a situation closely analogous to that of the Berger task? This study attempted to answer this question by having 30 undergraduates from Bard College complete one of two variations of this task, the first of which had participants rate the relevant images using only negative words (as per the original task) and the second of which had them rate the negative images using only positive words. It was hypothesized that the second condition would strike participants as more difficult to complete, and therefore this condition would see higher levels of disobedience and indications of its difficulty in the post-test questionnaire. The data were inconsistent with this hypothesis. However, what was more interesting than the support or lack thereof for this hypothesis was the finding that both conditions of the task by Haslam et al. (2014) produced very low levels of both task defiance and reports of difficulty from participants. This is a finding which challenges the Haslam et al. paradigm as one to be viewed as an analogue to Burger's. Finally, future directions are discussed which might improve the task in this regard.

Introduction

Milgram and the Obedience Experiment

Today it would seem that no Social Psychology course, nor any course on Psychological Research Ethics would be complete without acknowledging Milgram's infamous experiments. The Voice-Feedback variation of his Shock Experiment has become one of the most well-known experiments in all of psychology. There are multiple reasons for this, not least of which being that its implications would seem to threaten one's faith in humanity. How is one to view the human race with untainted positivity upon discovery of the rates at which naive participants delivered high-voltage shocks to a chair-bound man for forgetting word pairs? Such an image troubles one's mind even more when details are added such as the victim pleading for his release after the delivery of the tenth shock. What's more, the learner falls silent after the 22nd shock as if to have been rendered unconscious, if not lifeless. Still, 62.5% continued shocking until the experimenter told them to stop (Milgram. 2019).

Let us spend some time overviewing Milgram's methods and procedure with a finer focus. Milgram's most discussed experiment is his "Voice-Feedback" variation (Milgram. 2019). This is the version commonly taught in Intro and Social Psychology courses. Milgram gathered participants by posting a listing around New Haven headlined "WE WILL PAY YOU $4.00 FOR ONE HOUR OF YOUR TIME: Persons Needed for a Study of Memory." Milgram received only 296 respondents from this form of solicitation which was not enough to meet his sample goal, for which reason he sent out several thousand letters of appeal to locals. Approximately 12% of these locals ended up participating in one of his experiments in their numerous variations. These participants were said to have ranged well in education level, age,

and occupation type according to the self-report survey Milgram attached to his invitation (Milgram 2019). Such an even variation in participants is important in giving one the ability to conclude whether such factors were relevant to an individual's likelihood of obedience. Milgram conducted his experiment in his Yale laboratory, a factor which one might intuit as significant in participants' tendency to obey—indeed in a later variation, Milgram altered his location to a run-down office building and saw obedience levels drop only to 47.5%; an increase in defiance but not an outstanding one (Milgram 2019).

Having arrived on-location, Milgram's procedure for the "Voice-Feedback" variation began by having a participant sit in a waiting room, empty otherwise for presence of a confederate man who would soon take on the role of the "learner." Upon the experimenter's arrival, the purpose of the experiment is explained as one that explores a recently developed psychological theory—that we learn material with increased efficiency if we are punished in the event of making a mistake during testing. Afterwards, the experimenter asked the participant and confederate if they would prefer being the party giving or receiving such punishment. Invariably, he said that he would write the word "Learner" on one slip of paper and "Teacher" on the other before drawing roles at random. However, as both slips contained the word "Teacher" the participant was always chosen as such (Milgram. 2019).

With their roles determined, the participant and confederate were led by the experimenter into a room containing an electric chair. The learner/confederate was then sat down in the chair and restrained with straps, the purpose for which were justified as "preventing excess movement" (Milgram 2019). A scripted interaction took place wherein the learner asked if the shocks would be harmful for him as someone who experiences heart problems. To this, the

experimenter responded that while the shocks can be extremely painful they cause no permanent damage. At this time, the teacher is given a 45v sample shock from the generator to remove any doubt as to the machine's authenticity.

With the learner strapped in, the task began. The participant was to read out a series of word pairs to the learner such as "blue box, nice day, and wild duck." Then the learner was to be tested on which second word with which each of the first were associated. The learner was supposedly given four multiple-choice options for each answer and would make a selection by pressing one of four switches. If the learner pressed the wrong switch, the teacher was to deliver a shock from the generator which ranged from 15-450v. The first shock delivered would be a small one of 15v, but for each subsequent wrong answer given, the participant was told to increase the voltage by another 15v such that the second shock would be 30v, the third 45v, etc. Milgram's design had the learner purposefully give wrong answers throughout the experiment so that continual punishment would be warranted. In actuality no shocks were being delivered, however the participant could hear audio-recordings of the learner making pained grunts early on that would incrementally intensify. When the 150v shock was delivered, the teacher would hear the learner demanding to be let out, as he refused to go on (Milgram. 2019).

If at any point the participant expressed hesitance to deliver a punishment, either by expressing such reluctance or idling for too long, the experimenter in the room would verbally prod him to continue. The first prod was "Please continue." If that did not prompt obedience, the second prod would be "The experiment requires that you continue." If necessary, the third prod would be "It is absolutely essential that you continue," and the fourth would be "You have no other choice, you must go on." If even the fourth prod did not persuade the participant to obey,

the study would end and the participant would be debriefed and recorded as having disobeyed. However if any of the prods succeeded in promoting obedience, and at a later interval the participant were again to hesitate, the prodding sequence would begin again with prod 1 (Milgram. 2019).

After the 300v shock, the learner shrieked in pain, declaring that he would no longer provide any answers to the test. At this point, the experimenter instructed the participant to treat a lack of response as a wrong answer and continue delivering shocks until correct answers were given. After the 330v shock, the learner would fall silent as though rendered unconscious or dead. Even so, the experimenter would order that the shocks continue. If the participant reached the 450v marker which was the generator's maximum voltage, they were instructed to continue delivering shocks at that level. If the participant obeyed for two more shocks, the experimenter ended the study and debriefed the participant who was recorded as having obeyed. This was the procedure by which Milgram tells us his infamous experiment was conducted (Milgram 2019). However, a later paragraph will mention some evidence which casts doubt on some of the details of his procedure (Nicholson 2011, Perry 2013, as cited by Brannigan, Nicholson, & Cherry 2015).

In this particular variation, 62.5% of participants obeyed the experimenters instructions, delivering all 30 shocks to the learner, each with an increasing level of voltage. This is a result which prompted very serious questions from the psychological community—what was interesting about Milgram's studies was not only the things he got his participants to do, but the reasons for which they obeyed or refused. Why would so many people obey? Could participants have derived sadistic pleasure from giving a stranger such punishment, did they think they were

doing something good by fulfilling their role the experiment? Is there any case to be made that understanding the mentality of Milgram's obedient participants could tell us something about the mentality of the average Nazi? The following paragraphs will discuss what psychological literature has to say about these questions and and more, as well as familiarize the reader with some of the methods by which researchers have attempted to derive their answers. The course of this lengthy introduction will eventually reach a subsection discussing a few more recent Milgramesque obedience experiments including Burger's famous replication (2009) before finally arriving at its true analytic focal point; the experiment "Nothing by Mere Authority" by Haslam et al. (2014). It is the validity of this particular experiment with which this paper is truly concerned. But first the introductory subsections to follow are meant to contextualize this work by Haslam et al. (2014) among others which had a similar goal; to evidence some conclusion as to the mechanisms behind Milgramesque "destructive obedience."

Functions and Meaning of Milgramesque Obedience

Interpretations of Milgram's experiment varied early-on and to this day remain a topic of debate. Many including Milgram himself considered the commonplace of participant obedience to be both a validation and function of Hannah Arendt's "Banality of Evil" theory which roughly postulates that it is perfectly within the capacity of the normal mind to unquestioningly, unimaginatively, and dispassionately carry out acts of evil when they are given to us by figures of authority (Green, S. R. 2013, Milgram, 2019 Reicher & Haslam 2004,). Arendt's Banality of Evil theory is normally associated with her analysis of the justification Nazis used to carry out their genocidal duties. However, not all believe there is a case to be made that the psychological

mechanisms by which participants obeyed Milgram are comparable to those by which Nazis carried out their orders.

Leonard Berkowitz (2014) concludes from studies of the Nazi's typical attitude towards their mission that there was a common component of hatred and sadism which it would be difficult to imagine any of Milgram's participants shared; Arthur Miller (2016) expressed the same point in *The Social Psychology of Good and Evil*, maintaining that neither Milgram's nor any other social psychological experiment has used the actual dehumanization of a participant as an independent variable. Having said this, is there any possibility that Milgram's situation engaged a part of the participant's mind that that produced uncharacteristically sadistic behavior? Behavior evidence would not support this. For one, Milgram himself reported clear signs of stress from most participants during the task including sweating, trembling, stuttering, and even nervous seizures in the case of 3 participants. Many of them begged permission to stop from the experimenter, but curiously few displayed the willpower to stop simply of their own accord (Milgram 2019). Yes, some participants were reported to have been heard laughing during their participation though this laughter was not described as sinister in nature, rather it seemed to have been yet another form of nervous reaction to the power of the situation (Milgram 2019). In short, behavioral evidence far-from suggests that Milgram's participants took Nazi-esque pleasure in following their orders, much to the contrary most found the task considerably stressful and for some it was even traumatic.

The stress incurred to participants of the Milgram experiment was recognized early-on by Diana Baumrind (1964) who famously argued for the immorality of Milgram's experiment as participants should have the right to expect that their welfare will be protected. Milgram replied

to Baumrind's concerns, maintaining that his participants were not only satisfactorily accomodated and debriefed but that curiously they normally expressed gratefulness for having taken part in such a meaningful experiment (Milgram, 1964). However, more contemporary investigations into the archives at Yale seem to reveal that Milgram drastically underreported the stress his experiment induced (Nicholson 2011, Perry 2013, as cited by Brannigan, Nicholson, & Cherry 2015). These investigations exhibit that Milgram's experimenters sometimes used more verbal prods than permitted by the guidelines of their procedure in order to persuade unwilling participants to continue, and also that not all of his participants were debriefed. Letting participants leave without a proper debriefing seems not to have so much indicative of sloppy procedure but of a conscious choice so to prevent word spreading of his deception to potential participants in New Haven. The matter of Milgram withholding debriefings is described as particularly troubling by Brannigan, Nicholson, & Cherry (2015) as "One participant reported that he lost his job after the experiment due to an emotional outburst during a discussion about the experiment with a fellow employee who had also participated in the study. Another reported that he had suffered a mild heart attack shortly after the study, implying that the extreme stress of the study was at least partially responsible." Such investigations suggest not only that Milgram's reports of procedure were suspect, but would also support an understanding that Milgram's task was a seriously difficult one for participants. Sadism would not seem to have been a function of obedience.

So why did so many people obey? Another hypothesis might be that, while all individuals wanted to disobey, only a minority had the courage to do so. Lee D. Ross (1988) suggested that the term "destructive obedience" is an inaccurate one and that a better way to describe

participants' behavior is in terms of "ineffective disobedience." One can imagine Milgram's participants not having wanted to continue for different reasons, one being that they would take personal issue with harming an innocent stranger and another being that, whether or not they were willing to continue, they are concerned about the learner's well being given the nature of his outbursts. Furthermore, one might stop on the basis of the learner's insistence from the 10th shock and beyond to be released from the experiment such that they are letting the learner decide whether or not the experiment should continue; and one might also stop on the basis of the golden rule, acknowledging that if they were in the learner's position, they would want the experiment stopped immediately. Conversational analysis of audio and video recordings from the Milgram archive reveal that defiant and obedient participants alike voiced concerns from all of these perspectives, but it was the defiant ones who more so expressed this diverse spectrum of reasons to the experimenter as to why they should stop. As such it seems that the defiant participants more than the obedient ones voiced a varied range of rationales for their stopping, perhaps implying a greater willingness to engage in the kind of conflict with the experimenter which would lead to disobedience (Hollander & Maynard 2016).

Another reason for such ineffective disobedience seems to intersect with cognitive dissonance theory. Rochat and Modigliani (1995) found that once the learner has demanded to be let out but the teacher shock him anyway, Milgram's task becomes more difficult to defy. This was suggested by their discovery that participants who made expressions of notable resistance before the 150v shock were far more likely to stop the experiment than those who made signes of notable resistance after delivering the 150v shock. Jerry Burger (2009) similarly found that 79% of people who delivered the 150v shock went on to deliver all the others. Why would an

individual's likelihood of resistance drop so drastically after the 150v shock? Part of Rochat and Modigliani's explanation for this phenomenon is the proposal that, having already delivered a shock to the learner immediately after he had begged for the shocks to stop, obedient participants at this point became psychologically attached to the narrative that they had and no choice but to deliver that shock as they were being forced to by the experimenter. As such, if they were to stop participating after delivering the 150v shock, it would mean that they would have to abandon this narrative, having come to the realization that they were not forced to continue and could have stopped at anytime. This would also mean the a personal confrontation with the realization that they are the type of person who would continue agonizing the learner even though they don't *need* to. So to avoid coming to such a grim realization and its personal consequences, participants who delivered the 150v shock continued shocking to preserve their conception of their character; holding onto the belief that, as the experimenter has told them, they had no choice but to continue. Such a psychoanalytic theory is fascinating, but shares the difficult problem of psychoanalytic theories generally; being difficult, if even possible to prove. Nevertheless, it makes for a thought provoking proposal that one of the mechanisms behind obedience was essentially ego preservation.

Cultural, Personality, & Historical Factors

Another interesting factor of analysis is culture. Would obedience rates vary if the voice-feedback experiment were replicated outside of the U.S.? Thomas Blass (2012) compiled the results of 9 voice-feedback replication studies from Italy, South Africa, Germany, Australia, Jordan, Spain, India, and Australia, finding that the foreign mean obedience rate was 65.94%, very close to the U.S. mean obedience rate which he cites as 60.94% based not only on

Milgram's original voice-feedback obedience rate but also on those of its referenced replications in the U.S. (Acona & Pareyson 1968, Edwards et al. 1969, Gupta 1983, Kiham & Mann 1974, Mantell 1971, Miranda et al. 1981, Schurz 1985, Shanab & Yahya 1977, Shanab & Yahya 1978; as cited by Blass 2012). Blass' analysis of these studies leads to the suggestion that Milgram's high obedience rates are not merely reflection of an American response to the situation but a more global one.

Some consider personality to have been be a likely factor of obedience, particularly when one takes into account that personality variables might greatly alter the course of the experimenter-participant interaction; a highly empathetic participant might embody that personality trait in the form of more frequent and impassioned expressions of concern for the learner than would others. Meanwhile a more assertive participant might have been less likely to shy away from conflict and and the choice to disobey than less extraverted participants. This latter idea was roughly the hypothesis of Miranda et al. (1981) who did an interesting game show-like replication Milgram's voice-feedback experiment in Spain while measuring participants' scores on an introversion/extroversion index . They found no significant differences in obedience between extroverts and introverts.

Several pieces of literature explore similar questions of certain personality factors interacting with obedience rates, (Blass 1991, Blass 1995, Burger 2009, Elms & Milgram 1966, Larsen et al. 1972). While, the findings of these pieces are often inconsistent, multiple studies have found a positive relationship between likelihood of obedience in Milgramesque paradigms and the political tendency towards right-wing authoritarianism (Bègue et al. 2014, Elms & Milgram 1966, Meeus & Raaijmakers 1995, Blass 1995). In addition to this, Bègue et al. (2014)

found evidence suggesting that agreeableness and conscientiousness are also positively associated with such obedience, the former because of its associated motivation to avoid conflict and the latter because of its associated strength of commitment to duty. But they acknowledge notable arguments that circumstance can arguably have a greater influence on our actions in powerful situations than our character (Doris 2002 as cited by Bègue et al. 2014).

In 2009, Jerry Burger conducted a well-known partial replication of the Milgram experiment in order to find out if participants would still obey Milgram's voice-feedback task in the 21st century. His replication included multiple safeguards to mitigate any stress the experiment might cause, for instance participants went through clinical screenings and were excluded if expected to react negatively to the experience. The most notable safeguard taken was having the experiment stop after the delivery of the 150v shock as 79% of Milgram's participants who delivered this shock delivered all others. Burger (2009) expected to find the rates of obedience in his modern sample only minimally deviant from that of Milgram. In making this assertion, Burger (2009) acknowledged the argument that perhaps today people are more commonly weary of the dangers of authoritarianism for which reason we would be primed more so than Milgram's mid-twentieth century population to disobey the experimenter. However, he cited Thomas Bass' analysis that more recent obedience experiments do not show evidence that today's participants are more defiant (Blass 2004, as cited by Burger 2009). Burger's results found obedience rates similar to Milgram's as 70% of participants had to be stopped compared to the expected 79%. This leads him to conclude that the situational factors which brought about obedience for Milgram would still hold their power over participants today.

Burger's replication sustained some criticism from Alan Elms who worked closely with Milgram. Elms famously coined the term "obedience lite" in describing the replication, believing it removed much of the stress that made the original situation so powerful (Elms 2009). Perhaps what should be considered the most reasonable source of Elms' reservations had to do with Burger's extensive screening process whereby participants suspected to have an especially negative reaction to the task were excluded. Regardless, the replication is a highly referenced piece, appearing in many recent textbook editions addressing obedience research.

Within the domain of obedience research, Burger's replication is important in that it proposes a paradigm by which to perform variations of Milgram's original task in an acceptable way by today's IRB standards. However, one of its major drawbacks is its expensiveness to run as it involves not only clinical screenings of prospective participants, but also the presence of an on-duty clinical psychologist to intervene if a participant has a particularly bad experience with the task. In 2014, Haslam et al. invented a newer, far cheaper obedience paradigm which they hoped would serve as an analogue to Berger's despite the actual obedience task being entirely different. It is the efficacy of this task as one in the genre of obedience with which this senior project is concerned. The following paragraphs will provide an overview of Haslam et al. (2014).

The Study in Question: Haslam et al. (2014)

Recently after Burger's 2009 replication of Milgram, a study titled "Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science" by Haslam et al. (2014) was published. As its title suggests, this experiment's goal was firstly to invent an analogue to the Milgram experiment, the likes of which would be acceptable to 21st century IRB standards of practice,

and secondly to use that analogue in such a way as to learn something about participants' motivations for obeying Milgram's original experiment. As previously mentioned, though this task differs greatly from that of Burger's (and by extension, Milgram's), it was devised as "an analogue of the [Burger] paradigm that incorporated some of its most relevant features, but which was relatively easy to implement" (Haslam et al. 2014). The following paragraphs will explain what the task was.

This obedience task had participants take an online survey the nature of which they were told pertained to the cognitive science behind forming social impressions. The task itself asked that participants rate a series of 30 photos exclusively in terms of negative labels (the images are shown in Appendix H and the labels are listed in Appendix I). Participants were given 5 negative labels from which to choose for each image, but if they wished to, they could stop their participation at any time by clicking a "STOP STUDY" button as their answer choice. The task was meant to initially seem sensible to participants, but gradually become socially problematic. This was because the images participants judged were at first unpleasant (featuring Taliban members, Nazis, and the like), but as the task went on, they became more pleasant (featuring children in a classroom, a family in a park, etc.). As such the Milgramesque feature of the task was that it gave participants the choice to either complete the task and in doing so, associate positive images of people with words like grubby, bruthal, or insolent, or to stop their participation. Haslam et al. (2014) discussed their obedience task as an attempt to circumvent the ethical and financial difficulties of reproducing Burger's (2009) replication.

Haslam et al. (2014) hoped to use their task to gain new information as to the reasons Milgram's participants obeyed his task. This information would come from an analysis of the

efficacy of using Milgram's original prods to prompt obedience to their task. As such, they used a between participants design with four conditional groups. At the presentation of each of the 30 images, participants in were presented with online text displaying one of Milgram's prods. Condition 1 received prod 1, "Please continue," before each question. Condition 2 had Milgram's second prod, "The experiment requires that you continue." Condition 3 featured Milgram's 3rd prod, "It is absolutely essential that you continue," and condition 4 had the 4th prod "You have no other choice, you must go on."

In summary, Haslam et al. (2014) found that participants in the prod 2 condition had the highest likelihood of completing both his task and his posttest questionnaire afterward. Such a finding seemed to support the "engaged followership" model of obedience which interprets Milgramesque obedience to be a function of the participants' identification with the experimenter and his research goals, more so than a function of giving into the pressure of authority.

What is perhaps even more interesting about this study is that averaged across prod-conditions, only 47% of participants were reported as having completed the study, meanwhile 62.5% of Milgram's participants completed his voice-feedback experiment variation. Why would more participants have stopped Haslam et al. (2014)'s far more ethically conscientious study than did Milgram's? Could it have been because the nature of the task, while it was not concerned with the dramatic suffering of an innocent individual, still held the participant's sense of morality and decency at stake in a comparable way? While Haslam et al. consider their task an analogue to the task of Burger/Milgram, they acknowledge that it lacks the component of victim feedback that would give it more drama. With this acknowledgement in mind, could be comfortably asserted that participation in this new obedience task would be

Milgramesque enough that evidence for an engaged followership model in found here would imply its evidence in Milgram's paradigm?

Answering this question solely based on the data shown in Haslam et al. (2014) is difficult as participants accessed their online task via an email link, such that their participation could not be witnessed. Therefore, the only information we have as to participants' experience during the task, and the degree to which that experience resembled that of a Milgram/Burger participant, is firstly the rate at which participants' stopped the study, secondly whether or not they responded to the post-test questionnaire, and thirdly how they answered the post-test questionnaire if they did. This questionnaire featured the crucial questions "I found it hard to continue to the end of this experiment" and "I was comfortable continuing to the end of this experiment." As mentioned earlier, averaged across conditions only 47% of participants completed the task, which may have been because it presented them with a certain kind of uncomfortable situation, though this is unknown. We do know that participants across conditions were found to have a mean task comfort score of 2.55 on a 7 point likert scale, suggesting that they indeed found the task somewhat uncomfortable, but this is not very descriptive. What about the task made their experience somewhat uncomfortable? Did they feel faced with a moral challenge of any kind? Furthermore, only 70% of participants completed the questionnaire, leaving the attitude of the remaining 30% a total mystery. Are we to assume they did not complete it because they found the task somehow uncomfortable or is it also possible that they saw the task and questionnaire as something not to be bothered with responding to?

Research Question

The aim of this senior project's research was to come closer to answering the question: how powerful is the Haslam et al. (2014) paradigm in the comparative context of the Berger or Milgram paradigm? This question was answered by replicating the Haslam et al. (2014) task using a sample of participants from Bard College. In doing so, some changes were made as to the procedure of the task, including having the expermienter present during individuals' participation, both to add the element of having an authority figure present, and to make possible the witnessing of participants' attitudinal expressions throughout the task. Such observation would allow for the answering of questions regarding the extent to which the task was responded to as an obedience task carrying some moral challenge. Would participants, for instance, hesitate to respond or argue with the experimenter as participants did in the Milgram paradigm (Modigliani & Rochat 1995, Hollander & Maynard 2016)? Or would they perform the task quietly with little to no apparent qualms with its instructions? Furthermore, participants would be asked to complete a post-test questionnaire featuring a short answer portion, the purpose of which would be to ascertain some sense of their experience during the task.

Preliminary Study: Ordering of Images

Overview

Preliminary to attempting a variation of the Haslam et al. (2014) experiment, a question arose as to the 30 images used—would a Bardian sample find that ordering of the 30 indeed ranged out in order of pleasantness?. Perhaps certain images used by Haslam et al. (2014) would be appraised differently by the average Bardian student in terms of their pleasantness. To give specific examples, image 16 in the array, displaying Stockbrokers might be seen as less pleasant

to a Bardian due to its implicit association with corporate corruption. Also image 9 depicting a Kenyan riot might be seen as more pleasant to our population than that of Haslam et al. (2014) if one were inclined to associate it with a protest against an oppressive, post-colonial government. Being that Bard is a very social-justice oriented institution, such interpretations might not have been too far afield. Likewise, one might note that the population used by Haslam et al. (2014) were attendees of the Scottish University of St. Andrews rather than an American university. This geographical difference might also reflect an attitudinal difference towards images including the one depicting U.S. soldiers (image 11). While this particular image is on the more negative side of the order, would participants in the U.S. find it to be less of an unpleasant image compared to some of those Haslam et al. (2014) placed as more positive (ex: a group of Hell's Angels or tattooed men standing in front of graffiti)?

The question of whether or not the average Bardian participant would agree with Haslam et al. (2014)'s ordering of images may not have absolutely required investigation, but a brief investigation was conducted nonetheless. As such, finding evidence to show the extent to which the Haslam et al. image order original order concurred with the order of the average Bardian was the goal of Study 1.

Here is the procedure by which Haslam et al. (2014) determined their image order: Firstly, they gathered 60 images of groups which participants would find to vary greatly in terms of their pleasantness. These images were then made available to participants in an online survey (N=151). Participants' survey task was to judge the depicted groups' levels of pleasance on a 7 point likert scale. Afterwards, 30 of the 60 images were parsed out so to maximize the variance of pleasance and offensiveness. Thus Haslam et al. (2014) ended up with a series of images

which they hoped differed from one another in roughly equal intervals of pleasance/offensiveness.

Preliminary Study Methods

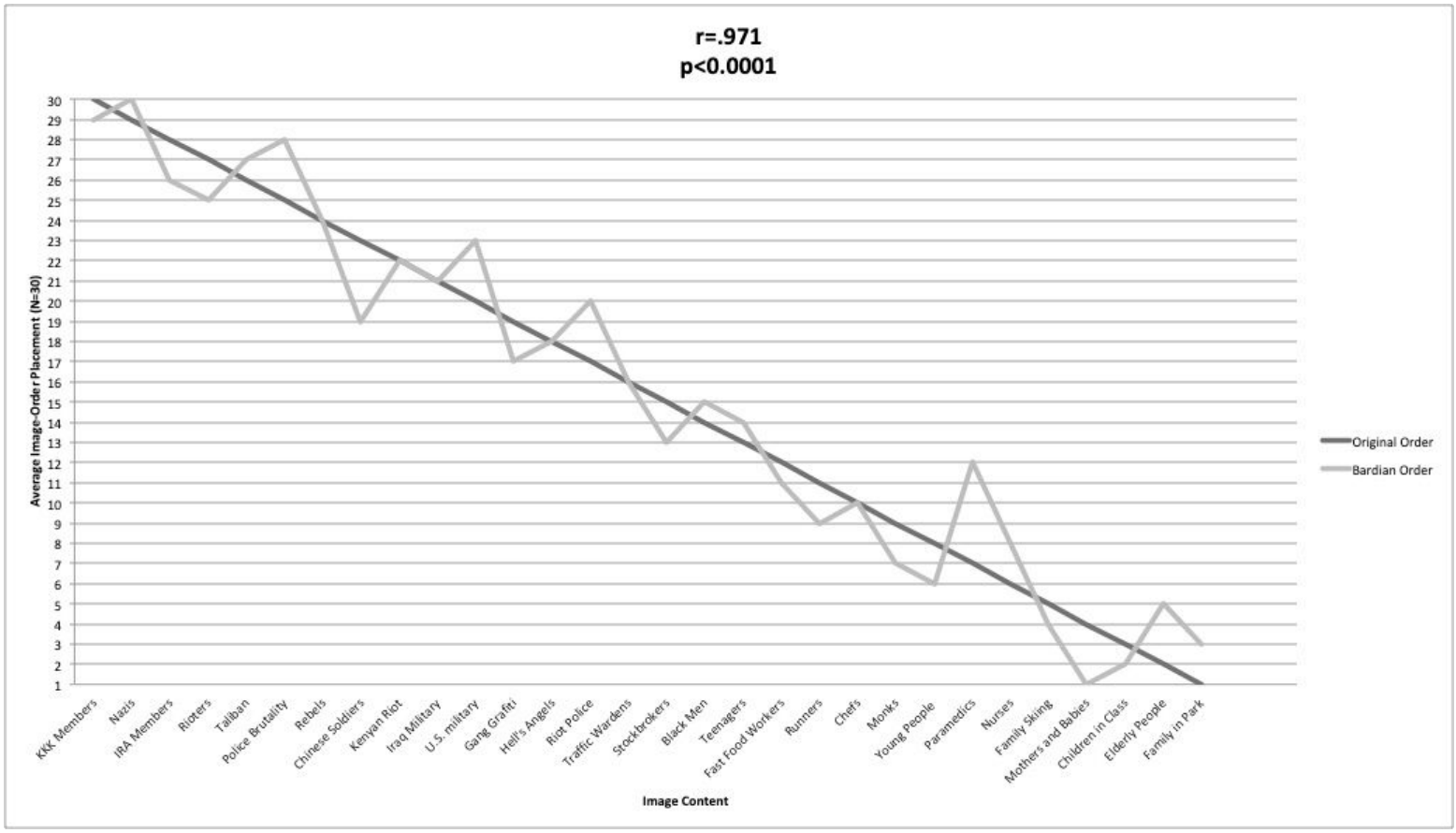
Here is the procedure by which Haslam et al. (2014) determined their image order: Firstly, they gathered 60 images of groups which participants would find to vary greatly in terms of their pleasantness. These images were then made available to participants in an online survey (N=151). Participants' survey task was to judge the depicted groups' levels of pleasance on a 7 point likert scale. Afterwards, 30 of the 60 images were parsed out so to maximize the variance of pleasance and offensiveness. Thus Haslam et al. (2014) ended up with a series of images which they hoped differed from one another in roughly equal intervals of pleasance/offensiveness.

The procedure to ascertain the Bardian order of images differed from that of Haslam et al. (2014), but met the same goal. An attitude measurement technique by Thurstone & Chave (1951) was employed which had participants arrange printouts of the Haslam et al.'s 30 images into 15 lettered piles according to their subjectively judged valence (ideally, all 60 images of the images would have been used, but Megan Birney of Haslam et al. informed me that access to the other 30 was lost; her help was in any case greatly appreciated). The very most positive images out of the 30, were to be placed in pile A, the most neutral images in pile H, and those most negative in pile O. The remainder of the images, which participants found neither most positive/negative nor most neutral relative to the others in the stack were placed in the remaining piles. Images seen as relatively positive, were placed in piles B through G with those in B being more positive than in C, C more positive than D, and so on. And images seen as relatively negative were placed in

piles I through N, with those in pile I being more negative than neutral pile H, pile J being more negative than pile I, etc. Once finished, all 30 images were ordered incrementally from most to least positive (A to O). More than one image was allowed per pile and participants were encouraged to change the placement of images if they changed their minds.

Preliminary Study Results

As per the Thurstone scaling procedure, images' letter assignments were recorded in terms of their number in the alphabet such that the range was 1-15. A Spearman's Rho test of correlation revealed that images' average ratings on a 1-15 point scale were correlated with the original order found by Haslam et al. (2014) with a correlation coefficient of 0.971, the p-value for which was significant (*Figure 1*). While the average Bardian order of images had minor ordinal discrepancies with the order of Haslam et al. (2014), because the two correlated highly and with significance, the choice was made to retain the original image order of Haslam et al. (2014) for the modified replication of their task.



[1]

[1] ***Figure 1*: Average Image Order by Bard Undergraduates Compared to the Order of Haslam et al. (2014).**
N=30, r=0.971, p<0.0001

Modified Replication of Haslam et al. (2014)

Overview

Having gathered evidence that the Haslam et al. image order would not be problematic for Bard Undergraduates, the primary piece of work in this Senior Project began. Its research goal was first and foremost to test Haslam et al. (2014)'s paradigm as an obedience study by not only replicating it using a Bardian sample, but by observing and recording participants' reactions to the experiment as well analyzing their responses to a post-test questionnaire after their debriefing. The primary goal of this questionnaire was to learn about participants' feelings, having taken part in this paradigm, as well as to hear their thoughts about the the task itself. To what extent was it one they found discomfort taking part in? Did they suspect any usage of deception? What other factors played into their choice to stop or continue participating?

While Study 2 was concerned with replicating the Haslam et al. (2014) obedience study, this replication contains some crucial modifications to the original experimental paradigm. These were: **(1)** There was an undergraduate experimenter who was personally present while participants took the online task. As such, participants were not taking the task alone on their own devices or in an environment of their choice as they were in the original paradigm. The presence of an experimenter throughout the task allowed for the observation and recording of participants' gestures and expressions in response to the task, as well as the prevention of any outside influence on the individual's choice to continue or stop. This way, participants could not for example, take the task in their dorm room and ask their dorm mate what they thought about the questions being asked. **(2.)** The experimenter was also enforcing a 15 second time limit on each question's response. If a participant took longer than 15 seconds to answer any question(s),

an alarm would sound and the experimenter would deliver a Milgramesque prod to prompt their speedy cooperation. **(3.)** Unlike the original paradigm, prods were only to be deployed in the event that a participant took longer than 15 seconds to respond to a question or made notable expressions of defiance. There were 4 different prods which were written to resemble Milgram's originals. These were used within the confines of a scripted sequence such that the first time a participant warranted a prod, it would be, "I need you to respond within 15 seconds, please continue." For if later on a second prod was warranted, it would be "Our experimental design requires reactions within 15 seconds." The third prod would be "It is absolutely essential that you continue. Respond within 15 seconds." And the 4th prod was "You have no other choice but to answer and within 15 seconds." If a participant warranted 5 events of prodding or more, the 4th prod was to be repeated until all questions were answered or the participant chose to stop the study. Acknowledging all of the changes to the original procedure that have been made in this replicatory paradigm (for which reason I term this study to have been a "modified replication") the core of Haslam et al. (2014)'s task remained intact. Participants would be rating the same order of images using the same negative words as they did in Haslam et al. (2014) in this condition; condition 1.

Additionally, a second, "reverse" condition was used in this study which was designed to elicit more study stoppage than would be found from participants in the "modified replication" condition, (condition 1). Condition 2 was identical to condition 1 in all ways but two: The first difference was that the images participants rated in terms of adjectives as per the original study were reversed in order such that they began with image 30 of Appendix H depicting a family in a park which was found to be the most positive image of the Haslam et al. array. From there, the

images descended in order of pleasance such that the final image was number 1 (a gathering of KKK members). And the second difference was that the 5 words from which participants could choose to describe the image were to be positive in nature rather than negative, (see Appendix I for the word list). With this, the difficulty posed to participants in condition 2 would be having to ascribe words of positivity to images of negativity or stop the study. (Appendix I also provides example questions for the two conditions).

Hypothesis

The primary concern of this experiment was firstly to see how its obedience rates would differ from those of Haslam et al. (2014). Some difference is expected here as the procedure has taken on modifications not least of which includes the addition of a human experimenter. It would seem predictable that this change in particular would lower replicated obedience rates based on a variation Milgram conducted wherein his experimenter was called away and replaced by an ordinary man in street clothes; in this variation Milgram's obedience rates dropped to 20%, suggesting that the presence of a supervising experimenter in an obedience task is indeed conducive to higher rates of obedience (Milgram 2019). And this experiment's secondary concern is the gathering of participant data in behavioral and questionnaire format in order to assess the task's efficacy in the genre of obedience.

It was also hypothesized that a greater percentage of participants would stop the study in condition 2 than the percentage who would stop in condition 1, as condition 2 was designed with the hopes of generating more discomfort from participants than was expected from condition 1. As such, condition 2 was constructed with the intention of being small upgrade in difficulty to Haslam's original task. It was conjectured that participants' difficulty in completing condition 1

(being Haslam's original task format) would be that by associating negative words with neutral or positive groups of people, they would take issue with the negativity of the words rather than the content of the images. Meanwhile in condition 2 where their task is to associate positive words with all images including very negative ones such as a depiction of police brutality (image 6), Nazis saluting (image 2), and a gathering of KKK members (image 1), participants would take issue with the presentation of negative images in a task otherwise centered around positivity. If this holds true, an understanding of which task is expected to be more difficult would stem from an understanding of whether or not there is a difference in salience between negative words (condition 1) or negative images (condition 2).

A phenomenon known as the picture superiority effect would lead one to an understanding that images generally are more salient than words in terms of semantic meaning (Hockley 2008, Mcbride & Dosher 2002). As such, a decidedly negative image such as one of Nazis would be more salient than a negative word. And perhaps this would especially be the case for the negative words Haslam et al. chose in particular, (shown in Appendix I) as this choice of negative words seemed rather tame, (ex: arrogant, insolent, dishonest, rude, slovenly, etc.). Furthermore, there is some evidence in the domain of ERP that exposure to negative images is related to brain activity of a larger amplitudes than exposure to positive images, suggesting the greater salience of negative images over positive ones (Ito, Cacioppo, & Lang 1998, as cited by Baumeister 2000). For these reasons, it seems to follow that the task of condition 2, the difficulty of which is the facing of participants with with negative images with which they most associate words of positivity, would be more powerful and conducive to defiance than condition 1.

Methods

Similar to study 1, study 2 enlisted the participation of 30 Bard College undergraduates. The method of recruitment used was personal solicitation by myself in the Campus Center. Participants were asked if they would like to take part in a Senior Project in the department of psychology under the domain of cognitive science (as was the task's cover story in Haslam et al. 2014). Students who agreed to participate were sat down in a private area of the campus center (the specific room varied depending on which were empty or at least sparsely occupied). From this point on, interactions with the participant were scripted (see appendix G). Participants were presented an informed consent sheet (Appendix F) which was written to lead them in believing that my study was "interested in examining the process by which people make word-group associations,*"*as this was, "a topic of considerable interest to cognitive neuroscientists interested in neural networking in the brain." After signing this sheet, participants were presented the instruction sheet for the task which I read aloud to them—they were encouraged to read it on their screen along with me. The instructions reiterated that the genre of the task based in cognitive science. Participants were told they would be looking at images describing groups of people and have to select the word shown which best describes the people in under 15 seconds. If they were in condition 1, they were told the words would be negative in valence while if they were in condition 2 they were told the worlds would be positive in valence. Lastly they were told that if they would like to revoke their participation, they needed only select "STOP STUDY" as their answer choice to any of the images. In an attempt to mitigate the concerns of any participants who might have later on in the task wanted to stop the study but choose not to for fear of having wasted my time, I clarified before the task began that a participant would be in no

way penalized for choosing to stop and that even in such a contingency, their data would still be of use.

The task itself had participants rate Haslam's images (shown in appendix H) using either all negative words if they were in condition 1 (just like in Haslam's original study) or using all positive words if they were in condition 2. In condition 1, the order of the images was the same as it was in Haslam's study, ranging from least to most pleasant while in condition 2 the images were reversed to range from most to least pleasant. Participants in conditions 1 and 2 alike were given 15 seconds to give an answer for each image. As the experimenter, I kept a 15 second timer on my iphone. If participants were late in responding, an alarm sounded on my phone and I delivered a prod meant to resemble one of the 4 Milgram wrote that he had his experimenter use if the learner showed signs of disobedience.

The usage of a 15 second time limit was justified as the operationalization for hesitation to continue based on a small sample of students (N=4) who were timed to complete a subset of the questions. As the average amount of time taken to answer each question was 13.5 seconds, 15 seconds was chosen as the cutoff time for each response such that exceeding that time limit was the primary operationalization for a participant expressing hesitation to continue which warranted prod delivery. The Secondary operationalization for a prod-warranting expression of hesitation was behavioral. If a participant were to verbally express a desire not to continue, but not press the stop button, the next prod in their place in the sequence would be delivered. If a participant were to express that none of the 5 words from which they could choose on an image were descriptive of the people shown, a special prod would be used stating, **"**Although none of the words may describe the images in your opinion, select the one that best fits." If after having

received this special prod a participant were to again complain that none of the words fit the image, one of the 4 prods would be delivered.

As noted, the prods used in this experiment were intended to resemble Milgram's. His original prods were written to have been **(1)** "Please continue," **(2)** "The experiment requires you to continue," **(3)** "It is absolutely essential that you continue," **(4)** "You have no other choice but to continue." Modeled from Milgram's prods, the ones used in this experiment were **(1)** "I need you to respond within 15 seconds, please continue," **(2)** "Our experimental design requires reactions within 15 seconds," **(3)** "It is absolutely essential that you continue. Respond within 15 seconds," and **(4)** "You have no other choice but to answer and within 15 seconds." Whereas Milgram reported having recycled his prod sequence such that if a participant at one point were to express enough disobedience to receive the first three prods before continuing and then later express hesitation to obey, the experimenter would have delivered the first prod again rather than continuing onto the fourth. Instead of choosing to recycle the prods used in this experiment, the choice was made to use the first 3 prods only once per participant such that if the 15 second timer went off, the first prod would be delivered; then if the timer went off again at a later image, the second prod would be delivered etc. As such, the only prod which might ever be repeated in this design would be the fourth and final prod—if prod 4 was delivered at one image and then at a later image the participant's time were to run out again, the 4th prod would be re-delivered and potentially be re-delivered again to be reused for any events of late response or expressed intention to defy until the task ended.

The decision to not reset the prod sequence per event of defiance even though Milgram did reset his prod sequence by starting again with the first prod at every new event of defiance

was justified as an attempt to make the tone of the experiment as obedience-oriented as possible. I did not want to have the sequencing prods be such that at one interval a participant might receive one of the latter two, more authoritarian prods ("It is absolutely essential that you continue. Respond within 15 seconds," or "You have no other choice but to answer and within 15 seconds.") and have the next prodding event be the recycling of the first, less severe prod ("I need you to respond within 15 seconds, please continue"). My reservation with the recycling prods as Milgram did was that it would allow for the potential that my role as the experimenter could change in tonal course from respectfully asking that the participant to continue more quickly (prod 1) to more coldly stating my experiment's need for rapid participation (prod 2) to authoritatively declaring my participant's necessity to answer on time (prods 3 & 4), only to afterwards de-escalate the authoritarianism of the situation back to the usage of the more polite prods (prods 1 and 2). Instead the tone of the prods and my role as the experimenter were designed to become more authoritarian with each event of defiance and then to maintain that tone.

Additionally, I took notes during individuals' participation, documenting their behavioral reactions or lack thereof to the task. Participants' verbal expressions and observable bodily gestures were recorded in the context of the image to which they were responding at the time of the expression. For example it was recorded that an obedient participant in condition 1 said "This feels so mean" while choosing one of the negative words to describe image 19 (showing smiling fast food workers). Meanwhile a participant in condition 2 laughed while choosing a positive word to associate with image 5 (of Taliban members) and later said "Is this a joke?" on image 2 (showing Nazis). Participants' recorded behaviors were then analyzed for their attitudinal

meaning. Ultimately, the goal in taking these notes was to have non-self-report data that would be telling as to participants' experience with the task. This data was intended to be useful in answering the question of how closely participants' behavior during this task reflected the kinds of behaviors one would expect to find from individuals participating in a socially discomforting obedience task.

After participants either completed all 30 questions in their respective conditions or at any point clicked on "STOP STUDY" as an answer choice, they were linked to a screen of text reading "Standby for debriefing." At this point they were presented with the statement of debriefing (Appendix J) which thanked them for their participation and disclosed that the true nature of this study was one not of cognitive but social psychology; that they were being tested for the variable of obedience. The form also divulges that throughout their participation, I was taking notes as to their behavior during the task, but that this information is totally anonymous. At this point, participants were invited to ask questions about the experiment and, if they were willing, sign the debriefing sheet.

After being debriefed, participants filled out post-test questionnaire (Appendix K) featuring 13 likert-scale questions and 2 short-response questions. The choice was made to debrief participants before asking them to fill out the questionnaire in preparation for the contingency that some disobedient participants might have found the task angering enough not to want to cooperate with a post-test questionnaire unless the experiment's true nature were disclosed. Though this contingency was not expected to occur, nor did it, such planning seemed like the most conscientious decision.

The likert questions were on a 7 point scale, through which participants could indicate that they strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, or strongly agree with a series of statements. These statements (also listed in Appendix K) were **(1)** "I found it difficult to continue this experiment to its end," **(2)** "I found the nature of the task absurd," **(3)** "This experiment caused me emotional discomfort," **(4) "**I imagine others would take this task seriously," **(5) "**Before being debriefed, I felt I was helping science by participating in this study," **(6)** "Before being debriefed, I regretted my participation in this study," **(7)** "I did not doubt the cover story of this experiment (regarding its relevance to cognitive science," **(8)** "As I was participating, I suspected that this study was using some form of deception," **(9)** "I was comfortable continuing to the end of the experiment," **(10)** "I felt this experiment asked me to do something that was morally questionable," **(11)** "I found myself hesitating to continue," **(12)** "I felt this experiment asked me to do something morally wrong," and **(13)** "I took this task seriously as I was completing it."

The two short answer prompts which ended the questionnaire were, "Please write about why you either continued the experiment or stopped participating," and "Finally, please add any thoughts you believe the researcher should know about your experience participating in this obedience study. I am particularly interested in hearing about the extent to which you took the task seriously and/or felt discomfort associating the words given with some of the people shown."

Results (Study 2)

Original vs Replicated Obedience Rates

Firstly, the rates of obedience in this replication were indeed lower than those of Haslam et al. (2014). 73.3 % of Condition 1 participants finished the task as compared to the cross-conditional average of 47% who completed it in the original Haslam et al. (2014) study. Despite this 26.3% gap in obedience frequencies, the difference is just outside of the statistical margin of significance. This near-significant difference will be addressed in the discussion section.

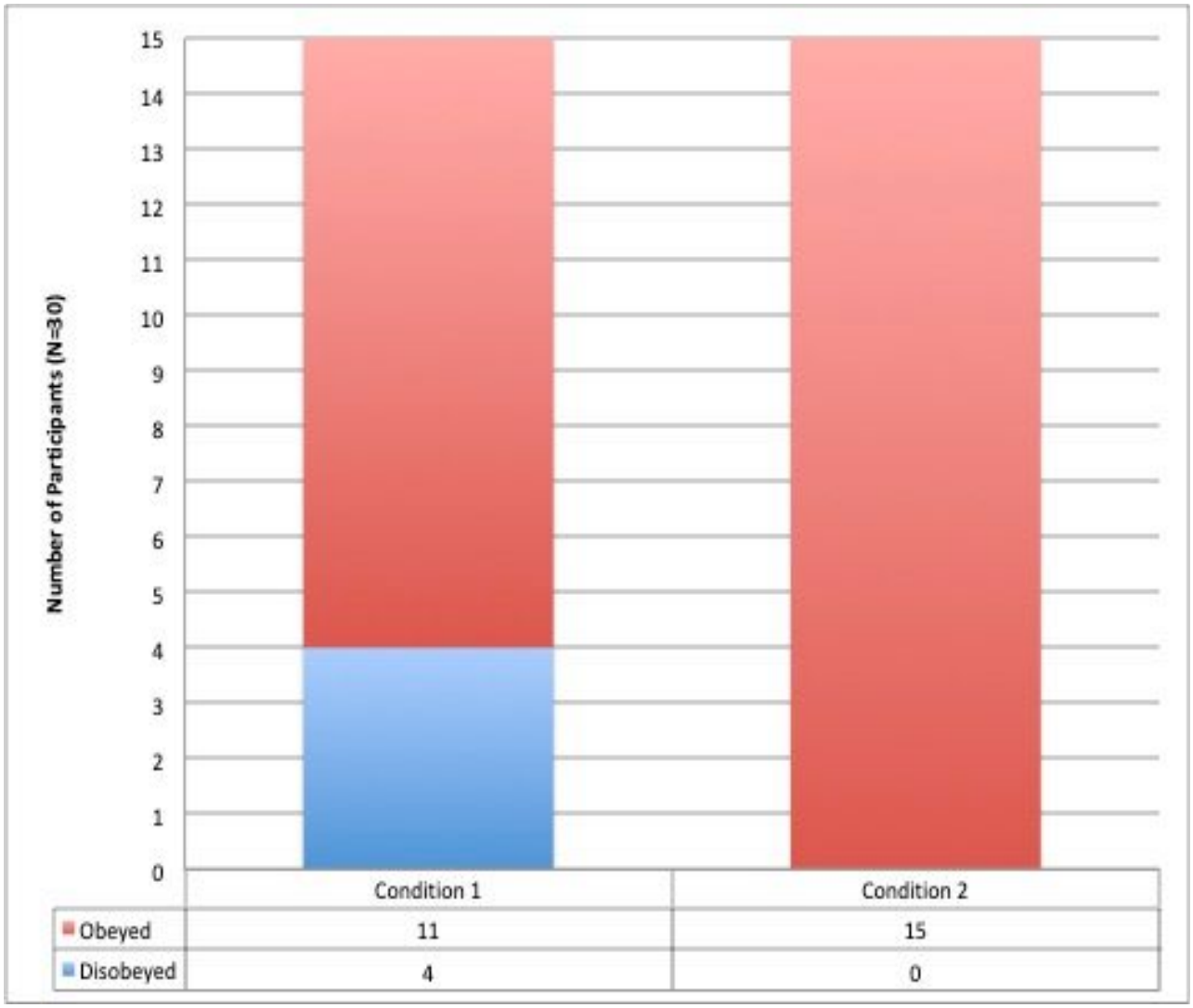
	Disobedient	Obedient	**Result**
Condition 1	4	11	73.3% Obedience Condition 1 N=15
Haslam et al. 2014 (Cross Conditional Average)	80	71	47% Obedience Haslam et al. 2014 N=151
Column Totals	84	82	**Chi-Square Statistic =3.7780** **p-Value = 0.0519** **ns**

2

[2] ***Table 1.*** Chi-Square Test table comparing the original cross-conditional rate of task obedience from Haslam et al. (2014) (N=151) to the rate of task obedience in Condition 1 (N=15). The test found the differences in obedience to be just below the threshold of significance.

Lack of Difference in Condition 1 vs 2 Obedience Level

The second data analysis to be looked at is the comparison in obedience rates between Conditions 1 & 2 as these results are clearly relevant to this paper's hypothesis. It was expected that more condition 2 participants would stop the task than those in condition 1 as condition 2 was designed to be harder; rather than having participants use negative words to describe positive images, this condition had them use positive words to affirm the continent of negative images even including depictions of Nazis and KKK members. However, the results were inconsistent with this hypothesis, showing that while few as 4 participants stopped the task in condition 1, while not a single participant stopped in condition 2. Because the frequency of condition 2 disobedience was a zero-value, a Fisher's exact test was used rather than a Chi Square test to analyze this cross-conditional difference in obedience. The Fisher's exact test is a form of Chi-Square analysis which allows for a 0-value in one of the cells while a normal Chi-Square test would not. The Fisher's exact test confirmed that this small difference in obedience level which would have gone counter to the hypothesis was statistically insignificant (*Figure 2, Table 2).*



[3]

[3] ***Figure 2: Cross-Conditional Obedience Rates***.
4 out of 15 condition 1 participants stopped the task while 0 out of 15 condition 2 participants stopped the task. (Statistical analysis in *Table 2*)

Cross-Conditional Obedience Rates

	Disobedient	Obedient	**Result**
Condition 1	4	11	73.3% Obedience
Condition 2	0	15	100% Obedience
Column Totals	4	26	**Fisher's Exact p =0.0996** **ns**

[4]

[4] ***Table 2.*** Fisher's Exact Test table comparing rates of task stoppage from participants in condition 1(N=15) and condition 2 (N=15). Fisher's exact test was used in place of a Chi-Square test as the latter form of analysis is not suited to analyze contingency tables with a zero value in one the cells. This test found the differences in obedience levels between conditions 1 and 2 to be insignificant.

Lack of Difference in Prod-Delivery Events

It was also revealed through a Fisher's exact test that neither condition necessitated significantly more events of prod delivery (*Table 3*). All instances of prod delivery were due to participants taking longer than 15 seconds to provide an answer such that no participant received a prod due to any expressed a desire to stop the study. This table shows that it was not uncommon for participants to complete the study without the need to be prodded even once, this was the case for 10 condition 2 participants and 4 condition 1 participants. Not a single participant needed to be prodded more than 4 times and only a single participant warranted the delivery of all 4 prods (even so, this individual did not stop the study).

Cros-Conditional Frequencies of Prod Delivery

Number of Prods Given per Participant	0 Prods	1 Prod	2 Prods	3 Prods	4 Prods	4+ Prods	**Result**
Condition 1	4	5	3	2	1	0	**Row Avg=1.4 Prods**
Condition 2	10	4	0	1	0	0	**Row Avg=0.47 Prods**
Column Totals	14	9	3	3	1	0	**Fisher's Exact p=0.1298** **ns**

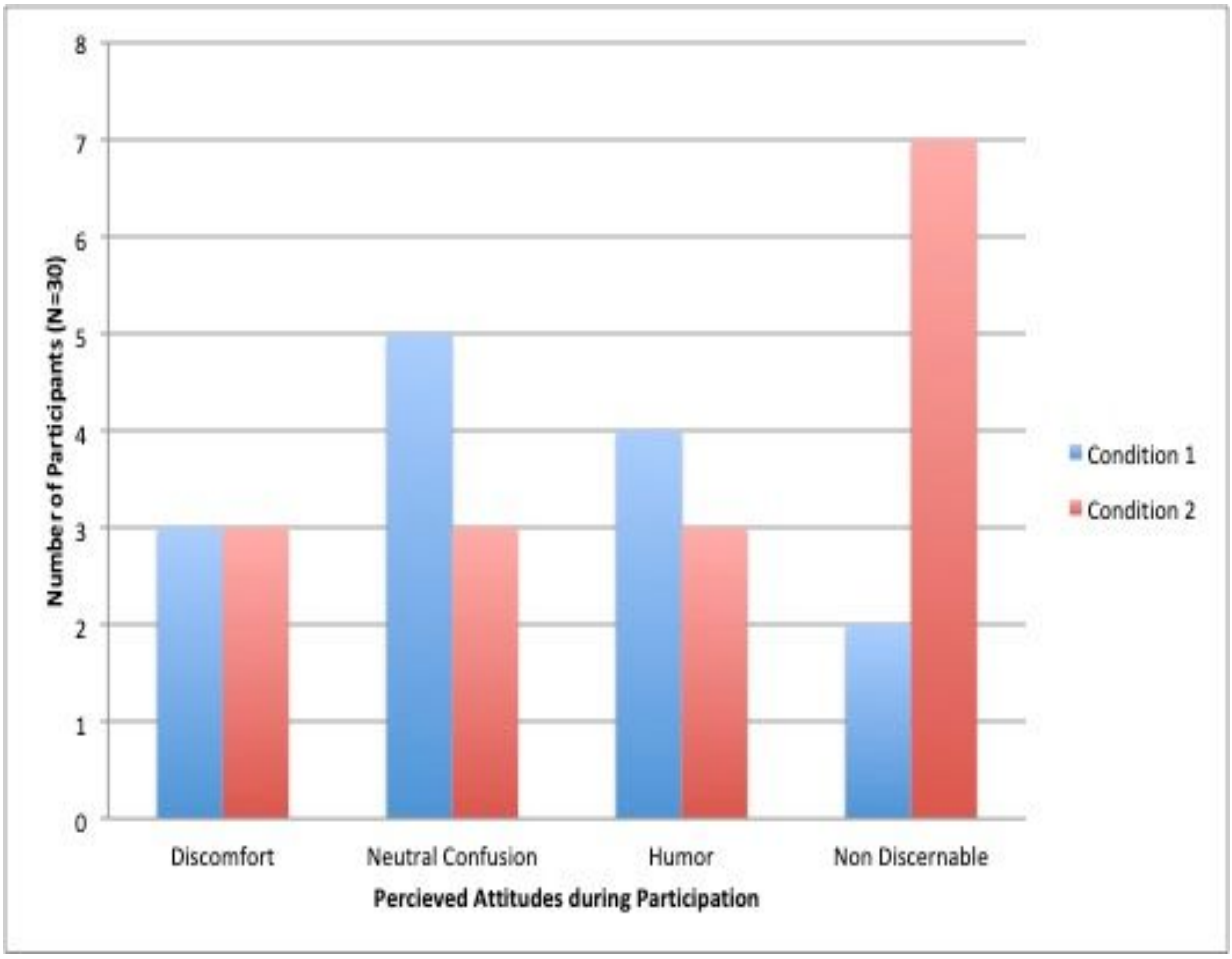
[5]

[5] ***Table 3.*** Fisher's Exact Test table comparing the frequencies of prod-necessitating events between conditions. All events were due to elapsed time-limit. (N=30) Differences are ns.

Lack of Difference in Perceived Attitudes Towards Task

Events of participant behavior during the task were recorded in written field notes. Examples included verbal expressions such as a participant in condition 1 muttering "Whatever…" while choosing a negative word to associate with image 22 (Buddhist monks) as well as nonverbal gestures such as sighs of confusion from a participant in condition 2 during image 9 (Kenyan rioters), and laughter from a participant in condition 1 in response to image 23 (young people).

Participants' attitudes towards the task as judged by their observable behavior throughout participation were fit into 1 of 4 generalized categories. These categories were **(1)** discomfort, **(2)** neutral confusion (the qualification of neutral being used to indicate a lack of frustrated or uncomfortable affect), **(3)** humor, or **(4)** non-discernible (meaning that no gestures or types of behavior were observed at all from that participant). The results of a Chi-Square test found no significant differences in the frequency of these attitudes between the two conditions *(Table 4*).The most relevant finding of this section overall pertains to the question of whether either condition of the obedience task faced participants with a Milgramesque situation. In both conditions, none but 3 out of 15 participants showed any behavioral signs of discomfort and about the same number did not seem to be taking the task seriously, behaving as though they found it humorous. Such findings make it difficult to believe that the task actually faced participants with an ethical challenge from their point of view.



[6]

[6] ***Figure 3:*** **Perceived Attitudes of Participants Towards Obedience Task**
Bar graph of the frequency of participants perceived attitudes towards conditions 1 & 2 of the obedience task as judged by their behavior while participating. Based on the judgement of the experimenter, only 3 participants in condition 1 and 3 participants in condition 2 showed perceivable behavioral signs of discomfort in response to the task. (Statistical analysis in *Table 4*)

Perceived Attitudes of Participants Towards Obedience Task

	Discomfort	Neutral Confusion	Humor	Row Totals	**Result**
Condition 1	3	5	4	12	
Condition 2	3	3	3	9	
Column Totals	6	6	7	21 (Grand Total)	**Chi-square statistic = 0.219** **p=0.896** **ns at p<0.05**

[7]

[7] ***Table 4.*** Chi-Square table comparing the number of participants between conditions whose attitude was perceived based on their behavior to be of discomfort, neutral confusion, or humor in response to the task. The 9 participants who displayed no perceivable attitude towards the task during their runs were excluded from this table (2 from condition 1 and 7 from condition 2). This test confirmed that there were no significant differences between conditions in the perceived attitudes of the participants.

Recorded examples which justify the placement of participants' behavior into these generalized categories of attitude are discussed in the paragraphs to follow. These examples are meant to give the reader a sense of each category's meaning as well as provide a more detailed understanding of how individuals reacted to both conditions of the modified Haslam et al. (2014) task.

Only 3 participants in condition 1 and 3 others in condition 2 displayed behavior indicative of discomfort. Such behavior witnessed in the 3 participants in condition 1 involved one participant nervously apologizing after being delivered Milgramesque prods for taking too long to respond to multiple images (a total of 3 prods were delivered to this participant, all for exceeding the 15 second time limit on various images). Another condition 1 participant made multiple comments during their run including "This feels so mean," "They're not doing anything wrong," "This is horrible," and "I feel so bad." And a third participant in condition 1 showed some sign of discomfort while negatively rating the image 22 (Buddhist monks) by saying "This is so hard." Meanwhile two of the three participants in condition 2 who expressed discomfort at the task did so only after the task was over; both turned to me and said "That was so stressful." The other uncomfortable condition 2 participant said "This is difficult" while rating image 7 (rebels), but made no other comment throughout their run. Other than 6 participants across conditions, nobody expressed any noticeable signs of discomfort during the task. One can sense from these descriptions that the discomfort observed from these participants seemed very mild with the possible exception of the participant in condition 1 who made multiple uncomfortable comments.

5 participants in condition 1 and 3 participants in condition 2 showed discernable signs of confusion with their respective versions of the task upon reaching the points at which none of the words they could choose from fit the content of the images. These participants either at one point in their run asked out loud "What?" in response to the image, or they asked some variation of the question "What if I don't agree with any of the words?" to which I replied with the special prod, **"**Although none of the words may describe the images in your opinion, select the one that best fits." None of these participants seemed to be uncomfortable or frustrated with the task, only confused as to the reason for its nonsensicality.

Meanwhile, 4 participants from condition 1 and 3 from condition 2 actually made gestures and/or expressions indicative of having found the task funny. One condition 1 participant while rating image 23 (young people) laughed and said "Oh yeah, f*ck these people." Another condition 1 participant laughed while attributing the word "aggressive" to image 28 (children in a classroom) saying "These children are *so* aggressive." Similarly, a condition 2 participant laughed while attributing the positive word "principled" to image 2 (Nazis) saying "They were definitely principled," and another condition 2 participant laughed at the presentation of the same image saying "This is a joke."

Meanwhile, the remainder of participants (2 in condition 1 and 7 in condition 2) showed no signs of attitude towards the task whatsoever. All of these participants were completely silent as well as quite still during their runs and made no facial or bodily gestures that would have been telling as to their demeanor.

Overall, the attitudes of participants in conditions 1 & 2 alike as judged by their behavior during the task would challenge the consideration of the task as one that confronts the average

participant with a Milgramesque situation. Only in rare instances did either task condition seem to bring discomfort to participants. And even in those rare instances, the the levels of discomfort observable were typically very minor.

Post-Test Questionnaire: Likert Scale Data

13 independent T-Tests were conducted to reveal any significant differences between the 2 conditions' responses to the 13 likert scale questions in the post-test questionnaire. To reiterate, these questions were posed in the form of statements, the degree with which participants could indicate agreeing/disagreeing/neither on a 7-point likert scale. These questions/statements aimed to measure the extent to which participants found their respective conditions of the task difficult, discomforting, and immoral as well as the extent to which they took it seriously and imagined others would take it seriously (*Table 5*).

Post-Questionnaire Likert Scale Data Analysis

Likert Questions on a 1-7 Scale 1=Strongly Disagree 7=Strongly Agree	Cond. 1 Mean Answer (N=15)	Cond. 1 SD	Cond. 2 Mean Answer (N=15)	Cond. 2 SD	Diff. in Means	t-statistic	p of Mean Answer Difference
1. I found it difficult to continue this experiment to its end	4.63	1.93	3.83	2.24	-0.800	-1.048	0.304 **ns**
2. I found the nature of this task absurd	4.69	1.96	4.06	2.09	-0.630	-0.852	0.402 **ns**
3. This experiment caused me emotional discomfort	4.5	1.77	4.28	1.84	-0.220	-0.334	0.741 **ns**
4. I imagine others would take this task seriously	4.88	0.93	5.11	1.76	0.230	0.447	0.658 **ns**
5. Before being debriefed, I felt I was helping science by participating in this study	4.63	1.32	4.83	1.67	0.200	0.364	0.719 **ns**
6. Before being debriefed, I regretted my participation in this study	2.19	1.18	1.83	1.54	-0.360	-0.719	0.478 **ns**
7. I did not doubt the cover story of this experiment	5.13	1.32	5.06	1.96	-0.070	-0.115	0.910 **ns**
8. As I was participating, I suspected this study was using deception	4.6	2.06	3.11	2.00	-1.490	-2.010	0.054 **ns**
9. I was comfortable continuing to the end of the experiment	4.38	2.15	4.5	2.22	0.120	0.150	0.882 **ns**
10. I felt this experiment asked me to do something morally questionable	5.25	1.75	3.94	2.15	-1.310	-1.830	0.078 **ns**
11. I found myself hesitating to continue	5.69	1.61	4.56	2.29	-1.130	-1.563	0.129 **ns**
12. I felt this experiment asked me to do something morally wrong	3.88	1.83	3.22	2.22	-0.660	-0.888	0.382 **ns**
13. I took this task seriously as I was completing it	5.06	1.75	6.06	1.58	1.0	1.643	0.112 **ns**

8

[8] ***Table 5.***
Condition 1 and condition 2 mean answers to the likert scale portion of the post-test questionnaire. (N=30) No significant differences were found.

These T-Tests revealed no significant differences in mean answers between conditions. However, mean responses between conditions to the statement "As I was participating, I suspected this study was using deception," (statement 8 of *Table 4*) were on the verge of significance at P=0.054. The average response to this statement from condition 2 participants was 3.11 (SD=2.0) which is closest to the likert scale increment of 3 which represented "somewhat disagree." Meanwhile the average response to this statement from condition 1 participants was 4.6 (SD=2.06) which rounded closest to the likert scale increment of 5 which represented "somewhat agree." Though this difference is not large and is ultimately statistically insignificant, it brings to one's attention the possibility that part of the reason 4 participants stopped the study in condition 1 while 0 stopped in condition 2 is because some of those 4 condition 1 defiants suspected experimental deception. This possibility may have some likelihood considering 2 of those 4 defiants indicated on statement 8 that they either agreed with the statement, giving an answer of 6, or indicated that they strongly agreed and gave an answer of 7. Meanwhile one other of the 4 defiant participants provided no response to statement 8 while providing responses to all others, the reason for which was unknown. The near-significant tendency for condition 1 participants to have had a slightly higher (although again, statistically insignificant) suspicion of deception will be addressed in the discussion.

Post-Test Questionnaire: Analysis of Short Answer Data

The 2 open-answer prompts of the questionnaire were 1:"Please Write about why you either continued the experiment or stopped participating," and 2: "Finally, please add any thoughts you believe the researcher should know about your experience participating in this obedience study. I am particularly interested in hearing about the extent to which you took the

task seriously and/or felt discomfort associating the words given with the people shown," Responses to these prompts were analyzed first for similarities in response between conditions, then for responses idiosyncratic to participants of one condition or the other.

Cross-Conditionally Common Responses There were 4 types of responses that arose from participants in both conditions, these were that participants (**1**) did not stop the task at least in part because they were motivated to be a helpful participant in this "cognitive study," (**2**) that they felt some degree of discomfort while participating in it, **(3)** that as the task became uncomfortable for them, they stopped taking it as seriously, and (**4**) that they felt there was too little at stake in this obedience task. An example of a response that fell within category (1) was "I felt that I was obligated to complete the experiment, and that it would be unhelpful or a waste of time if I were to stop midway. This felt more like a sense of responsibility to academia/research in the abstract than a personal commitment,"one that fit category (2) was "It felt wrong to identify normal people with bad attributes," an example for category (3) was "I took my answers less seriously as the pictures became more absurd and dissonant with the available options," and a response fitting category (4) was "It's hard to feel guilt when simply mislabeling things. I don't really feel bad about saying that some images were dirty because that's the only option that even vaguely fit, at least in my mind." It was also possible for some responses to fit within more than one of these categories simultaneously, for example one participant wrote "I just continued because I was helping and it wasn't difficult or time consuming," as well as, "I didn't feel as bad using a negative option because I was literally told to pick something negative. Had I been given the option to choose between negative and positive I would have felt quite bad if I chose a negative." This response fit within category (1) and (4) as it stated both the participant's

motivation to continue for the sake of helping out with an experiment and their feeling that there was too little at stake in the obedience task to warrant feeling badly.

The number of participants in condition 1 who expressed one or more of these 4 sentiments in their writing were compared to the number in condition 2 in a Chi-Square test, the results of which revealed no significant difference in their distribution (*Table 6*).

Post-Questionnaire Open-Response Data Analysis

Written Sentiment	'The idea of helping the experimenter motivated my full participation'	'I felt that too little was at stake in this obedience task'	'The task caused me some degree of discomfort'	'As the task became nonsensical, I stopped taking it as seriously'	**Chi-Square Results**
Instances of written response type in condition 1	3	7	6	3	
Instances of written response type in condition 1	3	1	7	2	
Column Totals	6	8	13	5	**Chi-square statistic=3.785** **p-Value=0.285** **Significance: ns**

[9]

[9] ***Table 6.*** A comparison of the cross-conditional frequencies of participants expressing the four generalized sentiments about their in-task participation in the open-response portion of the post-test questionnaire. Some participants expressed more than one of these sentiments. The frequencies of these written sentiments cross-conditionally were **ns**.

Condition 2-Exclusive Responses Additionally there were some categories of response unique to participants in condition 2 which seem in part to reflect reasons for which no participant felt the need to stop in this condition. For example, 7 participants wrote some variation of the comment that they did not feel badly participating fully in the task because the words from which they could choose to describe even the most negative images did not always strike them as positive. To this point, one participant wrote, "Although it was difficult to look at particular pictures of violent groups, such as the Nazis and the IRA, I continued with the experiment because even though these groups, in my opinion, are morally reprehensible, certain words such as 'disciplined' do apply to these groups since they are/were fighting for things they believed in." Similarly a second participant wrote, "'Powerful' and 'disciplined' were neutral words so they were easy to choose, so when there were bad people and I couldn't say they were 'powerful' or 'disciplined' [because these words were not among the 5 choices for that image] it was kind of stressful to try to pick another neutral word when there were none," and a third wrote that, "Towards the end of the experiment, I was only comfortable applying the words 'principled' and 'powerful' as responses, and all of the other descriptions completely incorrect." Such comments plainly suggest that the failure of condition 2 to produce Milgramesque defiance was directly related to a failure on my part to test if I was using a full list of words, all of which were positive enough to warrant an uncomfortable association with people like Nazis.

Defiant Participant Responses Lastly, it should be mentioned that of the 4 participants who stopped the study (all of whom were from condition 1), 2 wrote, that they stopped because they found the task too difficult to continue. The first of these two individuals wrote, "As soon as people I related to showed up, I could not associate them with the negative terms given," and the

other wrote a similar sentiment that. “When the images changed to people that I have seen, it felt wrong associating negative words to them.” Having said this, these were the 2 participants who indicated on the likert scale portion of the questionnaire having either suspected or strongly suspected this experiment’s usage of deception, which one might imagine also factored into their choice to stop. The other 2 participants who stopped the study seemed to do so, not necessarily because they identified with the people in some of the images so much as due to the simple fact that the task was nonsensical. One participant wrote “I felt a little uncomfortable answering the last few questions. The contradictions made me uncomfortable,” while another wrote only, “I stopped because the task stopped making sense.” It is not clear from their writing whether these 2 individuals stopped because the task felt wrong to them or because they found it nonsensical and did not want to waste their time when given to option to stop at any point. So even of the participants who stopped the study, it is not explicit that all of them did so because the task met its aim of making them feel socially discomforted.

Discussion

The hypothesis was that, in line with evidence that negative images (and images generally) are more salient than words, task condition 2 would produce higher levels of defiance and participant discomfort than condition 1, such that it would upgrade the difficulty of the task. This is because, whereas the difficulty of condition 1 would have come from negative words being associated with positive images, the difficulty of condition 2 would have come from positive words being associated with negative images—as such condition 1 discomfort would have been due to words whereas condition 2 discomfort would have been due to images which some evidence would suggest are more salient as stimuli than words (Hockley 2008, Mcbride &

Dosher 2002; Ito, Cacioppo, & Lang 1998, as cited by Baumeister 2000). This hypothesis was not supported and the levels of obedience found by Haslam et al. (2014) were on the margin of being significantly lower than the levels found in condition 1. Furthermore, neither observational data nor self-report data showed that either condition of the task had the power to put participants in a situation which would cause them social discomfort. This challenges the the relevance of participants' reasons for continuing or stopping the task in Haslam et al. (2014) to the reasons participants may have continued or stopped in more powerful obedience tasks like Milgram's and Burger's. Therefore it is questioned whether Haslam et al. (2014) found meaningful evidence for an engaged followership model of obedience in Milgram.

Limitations

Condition 2 Words. It may well be that the main reason condition 2 did not produce more discomfort or defiance than condition 1 is because the simple difference of reversing the order of images from condition 1 and making the task about positively rating all images including the very negative ones does not change the nature of the task in a meaningful way.

However one should also consider the written commentary of the 7 Condition 2 participants who indicated that they did not feel bad participating fully in the task because they did not find all of the words from which they could choose to necessarily be positive. Indeed, the list of positive words used for Condition 2 (seen in Appendix I), contained mostly obviously positive words including "Altruistic," "Peaceful," and "Upstanding", there were also some exceptions including "Disciplined," "Orderly," and "Hardworking." While these words are not negative, it is understandable that they do not carry a degree of positivity which would be altogether inappropriate in association with even the most negative of featured images. As some

Condition 2 participants argued in their written response, people like Nazis were disciplined, orderly, and hardworking; these qualities did not make them morally good only more successful in carrying out evil, therefore there was no social discomfort associating such words with them or any other of the images. Acknowledging this, one obviously wonders if Condition 2 would have seen more defiance if the list of words were overwhelmingly positive. A better approach to ensure this would have been to compile sizeable a list of positive labels and have them rated by a small team for their valence. Then, use only the highest-rated words such that there is an element of inter-rater reliability.

Condition 1 Implicit Bias Test Resemblance? The tendency for condition 1 participants to have had a slightly higher (although again, statistically insignificant) suspicion of deception may have to do with the format of condition 1 having had a quality reminiscent of the increasingly well-known tests of implicit bias. This may be the case as condition 1 had participants associate their choice of negative words with a racially diverse spectrum of people, perhaps making the task seem to be testing which kinds of negative words they associate with certain racial groups. What's more, multiple participants in condition 1 openly asked after the task was over if this task was meant to measure racism. If a condition 1 participant suspected something along the lines of this were the case, that may have given them a motivation to stop which they would not have had were they in condition 2.

In-Group Effect. An especially glaring factor needs to be addressed about the participants used in this experiment, which possibly likely inflated obedience rates in both conditions of this replication. That is the simple fact that my undergraduate-Bardian participants knew me as the experimenter to be one of their peers. While only an approximate minority of my

participants had ever met me personally, all of them knew that I, like all Seniors pursuing a Psychology major, was taking on this experiment as my Senior Project. As such, it would be legitimate for them to view their participation as an act of in-group courtesy, particularly considering no incentive was offered. Having said this, only 6 of 30 participants (3 in each condition) mentioned in written responses that their full participation was motivated by a desire to help the experimenter. But the possibility is worth acknowledging that this desire was augmented by a situation in which they were helping a fellow student complete a project rather than helping an out-group experimenter.

Furthermore, one might argue that this in-group factor complicates the conclusion that participants rarely stopped the task as well as tended not to be uncomfortable during participation simply because the task placed too little at stake in terms of consequence. It seems possible that because I was a peer to my participants, I was seen as less of an ominous figure of authority than an outsider would have been, and had this role been filled by such a person the task would have had more of a disconcerting element. While I agree that having a non-student play the role of the experimenter may have made the task a bit more nerve-provoking to the degree that it would remove any existing layer of familiarity between the participant and experimenter, I would not predict this would lead to higher frequencies of disobedience due to the expectation that to whatever degree the presence of a stranger could make the task more uncomfortable, it would also make the decision to stop participation more uncomfortable. This is suggested by the results of Milgram's shock-task variation where the role of the experimenter is replaced by a man in plain clothes, as the experimenter is called away to answer the phone. In this variation, obedience levels dropped to 20% as a function of the source of tension (i.e. the experimenter) being

replaced by someone who is perceived more as a peer than an authority figure (Milgram, 1964). Therefore if participants in this experiment would have experienced greater tension from an out-group experimenter, it would follow that this would cause them to stop the study less often rather than the reverse. However, I am not convinced that this change alone would necessarily make the task more tense as this study has not shown the task to produce much tension at all in its current form.

Replication? Aside from condition 2's inability to make the task harder as hypothesized, condition 1 of this experiment produced obedience levels that were right on the margin of being significantly lower than they were in Haslam et al (2014). Why did so many fewer participants stop the task in this replication than they did in the original? Results from the post-test questionnaire, the intra-task behavioral observations, and the fact that none of the prods delivered were due to defiant behavior, would make the suggestion that the obedience experiment in both conditions did not possess the power to make participants uncomfortable. Participants did not behave as though being faced with a cognitive-dissonance inducing moral quandary the likes of which Milgram's participants faced, and they indicated no such feelings. Often even they expressed sentiments to the contrary; that too little was at stake.

Changes to the Haslam et al. (2014) paradigm in this modified replication surely account for a great deal of the difference. The most important difference between our two procedures being that this one has participants complete the task under the supervision of an experimenter who is keeping a time limit on responses. While the finding that participants in both conditions did not treat their task like one of obedience would bring skepticism to the comparison between this task and the Milgram task, once more it might be useful to invoke the shock-task variation

Milgram performed wherein the experimenter was replaced by an ordinary-seeming confederate which caused obedience rates to drop to 20% (Milgram, 1964). This is relevant to the matter at hand in making the point that the social-pressure of a supervising experimenter causes participants to complete something which they might not otherwise. Granted, I would not make a comparison between the reasons someone would not want to deliver a high-voltage shock to Milgram's learner and the reasons they would not want to choose a negative word to describe a positive image (or the condition 2 inverse). As such, it might be expected that if my participants took the task online without any supervision, more might have chosen to stop if only for reasons such as not wishing to spend time completing an online survey that becomes increasingly nonsensical.

Another difference between the procedure used here and that of Haslam et al. (2014) was that in this version, prods were only delivered if a participant took more than 15 seconds to respond to an image or expressed a verbal desire to stop the task without actually clicking on the stop button (though not one participant performed the latter behavior). Meanwhile in the original Haslam et al. (2014) task, prods were shown on-screen at the presentation of each new image. Because of this difference, several participants in this experiment reached the end of the task without even being prodded once and very few participants ever needed to receive the more authoritarian-sounding prods (3 & 4). As such, despite this variation of the task having an authority member present, in some small sense one might argue that the tone of the task itself became less authoritarian as participants were not constantly being prodded to continue. However, with this said, one would imagine that when prods were delivered in this replication, they carried with them more power having come from the voice of a supervising experimenter

who is looking over the participant's shoulder holding a buzzing timer. Whereas, in the original variation, the prods manifested only as words appearing on-screen for no necessary reason. Perhaps some Haslam et al. (2014) participants may have even found these prods more annoying than persuasive such that they would even have been contributory to the reasons for which so many participants stopped. The inclusion of a "no prod" condition in their experiment could have examined this. Regardless, the point remains that the added presence of an experimenter seems to have been the major factor in the increased obedience rates of this experiment compared to its original variation.

Future Directions

Haslam et al. (2014) invented this obedience paradigm with the hopes that it would serve as an analogue to Burger's partial replication of Milgram's Voice-Feedback task. The Haslam et al. (2014) paradigm is easy to code through Qualtrics (and likely via similar platforms as well), it is convenient in its setup with no necessary usage of any kind of shock-box or other contraption, and it is conscientious towards participants by not having them perform any kind of task by which they might be traumatized. But this study would suggest that it has a problem fitting within the genre of obedience experiments—it does not face participants with real-enough consequences to make them want to stop. Condition 2 of this experiment was expected to somewhat address this issue, operating under the hypothesis that participants would have had a harder time associating positive words with negative images than negative words with positive images. But at the time condition 2 was devised, I had not anticipated that obedience levels would be so high. With this, perhaps condition 2's small change to procedure could not have

been reasonably expected to make a meaningful difference to the task's difficulty, even if it drew from a more carefully compiled list of positive words.

Among the factors for which participants in this experiment felt the task had too little at stake for them to seriously consider stopping, one that seems simplest to address in a future variation is that of anonymity. Some participants indicated having little problem completing the task because their association of the words with the people in the images, whether they agreed with that association or not, would not be witnessed by anyone aside from the experimenter. Perhaps there would have been more events of disobedience were this not the case.

For example, imagine a future variation of this task wherein the experimenter and participant take the task in a college classroom with desks. The experimenter leads the participant into the room and there they find a lone black student (who is a confederate) studying. The experimenter tells the student that he and the participant are about to run an experiment and he hopes they will not disturb the student's studies. The student tells the experimenter not to worry, and the experimenter sits the participant down in one of the desks in front of the student and starts up Qualtrics on their computer so that the screen is in full sight of the black student. Then the task begins, the instructions for which are have the participant rate each image with their choice of 5 positive words, (very much like condition 2). The images begin by showing American law enforcement in neutral settings, but as the task progresses they depict scenes which could be interpreted to resemble racial profiling (perhaps a black man stopped in traffic by a policeman with a scrutinizing espression). And finally the images depict explicit scenes of police brutality towards black people.

In such a variation, something is added at stake; the feelings of the nearby black student who can clearly see all of these positive word choices associated with the scenes. As such, the difficulty of the task now has far less to do with whether or not the participant personally takes issue with the association of the words with the images than it does with how the association of those two things might make the witnessing student feel.

As such the task would be given the type of power that would lend its analysis more relevance to interpretations of the Milgram experiment. From there, more variations and conditions could be added including the usage of only specific prods to test their efficacy as was the research question of Haslam et al. (2014). One could see if obedience rates differ internationally (granted, in the specific context of this proposed experiment, that might be more revealing of racial attitudes than destructive obedience generally). Or, one could one measure participants on scales of personality and political orientation to find obedience correlates as many have done in the past; right-wing authoritarianism may be a predictor as some studies showed it to be in Milgram's original paradigm and similar ones (Blass 1991, Blass 1995, Burger 2009, Elms & Milgram 1966, Larsen et al. 1972). Or one could try out a variation my advisor, Stuart Levine conceptualized, wherein the task is interrupted somewhere around the last few images as the experimenter's laptop screen goes black as though its battery died. The experimenter then would apologise to the participant, telling them that he needs to go grab his charger and that he should return in a few minutes. Now the participant is left alone in the room and allowed some time to introspect as to whether or not they truly wish to continue. If the participant decides in this time that they wish to stop, they could either wait until the experimenter returns and tell him so, after which being debriefed, or they might choose to leave

the room altogether before the participant returns (an eventuality for which the experimenter would have planned for, such that they would be waiting outside the room to catch the participant and give them a debriefing before sending them on their way).

Perhaps a situation like this would be comparable enough to Milgram's to suggest that if his participants were also given a break to think about their actions, more would have decided to stop. While the results of this project would critique their paradigm in its current form, I do believe that the task invented by Haslam et al. (2014) has qualities which lend it potential in today's realm of conscientious obedience tasks; the paradigm is very simple, portable, safe and inexpensive. Though some work may have to be done to give it more Milgramesque teeth.

References

Ancona, L. & Pareyson, R. (1968). Contributo allo studio della aggressione: La dinimica della obbedienza distruttiva (Contribution to the study of aggression: The dynamics of destructive obedience.) Archivo di Psicologia, Neurologia e Psichiatria, 29, 340-372.

Baumeister, R. F. (2000). Bad is Stronger than Good. *PsycEXTRA Dataset*. doi:10.1037/e413792005-154

Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's "behavioral study of obedience". American Psychologist, 19, 421–423.

Berkowitz, L. (2014). Evil Is More Than Banal: Situationism and the Concept of Evil. *Perspectives on Evil and Violence,* 246-253. doi:10.4324/9781315799292-7

Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology, 60*(3), 398-413. doi:10.1037//0022-3514.60.3.398

Blass, T. (2004). The man who shocked the world: The life and legacy of Stanley Milgram. New York: Basic Books.

Blass, T. (2012). A cross-cultural comparison of studies of obedience using the Milgram Paradigm: A review. Social and Personality Psychology Compass 6/2, 196-205.

Brannigan, A., Nicholson, I., & Cherry, F. (2015). Introduction to the special issue: Unplugging the Milgram machine. *Theory & Psychology*, *25*(5), 551–563. https://doi.org/10.1177/0959354315604408

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist,64*(1), 1-11. doi:10.1037/a0010932

Doris, J. M. (2002). Lack of character: Personality and moral behavior. New York: Cambridge University Press.

Edwards, D. M., Franks, P., Friedgood, D., Lobban, G. & Machay, H. C. C. (1969). An experiment on obedience. Unpublished student report. Johannesburg, South Africa: University of Witwatersrand.

Elms, A. C. (2009). Obedience lite. *American Psychologist, 64*(1), 32-36. doi:10.1037/a0014473

Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality, 1*(4), 282-289.

Fishbein, M. (1967). *Readings in Attitude Theory and Measurement*. New York: John Wiley & Sons.

Green, S. R. (2013). 'Hannah Arendt, evil and the eradication of thought' by Coline Covington. *Journal of Analytical Psychology, 58*(4), 563-566. doi:10.1111/1468-5922.12033_3

Gupta, I. (1983). Obedience to authority amongst university students: An experimental analysis. Unpublished Doctoral thesis, University of Delhi, Delhi, India

Haslam, S. A., & Reicher, S. (2007). Beyond the Banality of Evil: Three Dynamics of an Interactionist Social Psychology of Tyranny. *Personality and Social*

Psychology Bulletin, 33(5), 615-622. doi:10.1177/0146167206298570

Haslam, S. A., & Reicher, S. (2012). Contesting the "Nature" Of Conformity: What Milgram and Zimbardo's Studies Really Show. *PLoS Biology, 10*(11). doi:10.1371/journal.pbio.1001426

Haslam, S. A., Reicher, S. D., & Birney, M. E. (2014). Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science. *Journal of Social Issues, 70*(3), 473-488. doi:10.1111/josi.12072

Hockley, W. E. (2008). The Picture Superiority Effect in Associative Recognition. *PsycEXTRA Dataset*. doi:10.1037/e527312012-404

Hofling, C. K., Brotzman, E., Dalrymple, S., Graves, N. & Bierce, C. (1966). An experimental study of nurse-physician relations. *Journal of Nervous and Mental Disease, 143*, 171-180.

Hollander, M. M., & Maynard, D. W. (2016). Do Unto Others . . . ? Methodological Advance and Self- Versus Other-Attentive Resistance in Milgram's "Obedience" Experiments. *Social Psychology Quarterly*, *79*(4), 355–375. https://doi.org/10.1177/0190272516648967

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting Affect Using the International Affective Picture System: Trajectories through Evaluative Space. *Personality and Social Psychology Bulletin, 24*(8), 855-879. doi:10.1177/0146167298248006

Kilham, W. & Mann, L. (1974). Level of destructive obedience as a function of

transmitter and executants' roles in the Milgram obedience paradigm. Journal of Personality and Social Psychology, 29, 696-702.

Larsen, K. S., Coleman, D., Forbes, J., & Johnson, R. (1972). Is the subjects personality or the experimental situation a better predictor of a subjects willingness to administer shock to a victim? *Journal of Personality and Social Psychology, 22*(3), 287-295. doi:10.1037/h0032876

Mantell, D. M. (1971). The potential for violence in Germany. Journal of Social Issues, 27 (4), 101-112.

Mcbride, D. M., & Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition, 11*(3), 423-460. doi:10.1016/s1053-8100(02)00007-7

Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. American Psychologist, 19, 848–852.

Milgram, S. (2019). *OBEDIENCE TO AUTHORITY: An experimental view* (Vol. 475, Harper Colophon Books). S.l.: HARPERCOLLINS. ISBN:9780060129385

Miller, A. G. (2016). *The social psychology of good and evil*. New York: The Guildford Press.

Miranda, F. S. B., Caballero, R. B., Gomez, M. N. G., & Zamorano, M. A. M. (1981). Obediencia a la autoridad [Obedience to authority]. Psiquis, 2, 212-221.

Modigliani, A. and Rochat, F. (1995), The Role of Interaction Sequences and the

Timing of Resistance in Shaping Obedience and Defiance to Authority. Journal of Social Issues, 51: 107-123. doi:10.1111/j.1540-4560.1995.tb01337.x

Nicholson, I. (2011). "Torture at Yale": Experimental subjects, laboratory torment, and the "rehabilitation" of Milgram's "obedience to authority". Theory & Psychology, 21, 737–761. doi:10.1177/0959354311420199

Perry, G. (2013). Behind the shock machine: The untold story of the notorious Milgram psychology experiments. New York, NY: The New Press.

Reicher, S., & Haslam, A. (2004). COMMENTARY: The Banality of Evil. *Anthropology News, 45*(6), 14-15. doi:10.1111/an.2004.45.6.14

Ross, L. D. (1988). Situationist perspectives on the obedience experiments. *Contemporary Psychology, 33,* 101–104.

Schurz, G. (1985). Experimentelle Überprüfung des Zusammenhangs zwischen Persönlichkeitsmerkmalen und der Bereitschaft zum destruktiven Gehorsam gegenüber Autoritäten (Experimental examination of the relationship between personality characteristics and the readiness for destructive obedience toward authority). Zeitschrift für Experimentelle und Angewandte Psychologie, 32, 160-177

Shanab, M. E. & Yahya, K. A. (1977). A behavioral study of obedience in children. Journal of Personality and Social Psychology, 35, 530-536.

Shanab, M. E. & Yahya, K. A. (1978). A cross-cultural study of obedience. Bulletin of the Psychonomic Society, 11, 267-269.

Thomas Blass, Right-Wing Authoritarianism and role as predictors of attributions about obedience to authority, Personality and Individual Differences, Volume 19, Issue 1, 1995, Pages 99-100, ISSN 0191-8869

Thurstone, L. L., & Chave, E. J. (1951). *The Measurement of Attitude: A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward the Church*. Chicago: University of Chicago.

APPENDIX A

IRB Application for Scaling Study (Study 1)

SECTION 1

John Machen, (443) 299 7644, jm5142@bard.edu, Psychology, undergrad
Stuart Levine, llevine@bard.edu

SECTION 2

Do you have external funding for this research? N/A

Start Date: October 12, 2018
End Date: November 29, 2018

Research Question:
What is the average participant ordering–ranging from positive to negative in valence–of the 30 images used in "Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science"?

My subsequent study, in its current conception, will involve some deception; participants will take part in an online survey, the aim of which, they are told is to help me with cognitive psychology research dealing with the association of groups of people with adjectives. Allow me to explain – participants will be presented an image (ex: of a family in a park) and then asked to choose which of the following 5 words best describes the emotional content of the image: 1. Peaceful 2. Funny 3. Heartwarming 4. Exciting 5. Clean. For this particular example, I imagine most participants would choose “Heartwarming” or “Peaceful” rather than the other 3 as the most appropriate descriptor of the image; participants are meant to believe that they are helping teach an AI this very sense of appropriateness (a family in a park is not particularly funny, exciting or clean, but most people might find it somewhat peaceful and/or heartwarming). As the study continues, the images will become incrementally less positive and start to approach neutrality (ex: one will feature a team of paramedics loading an injured individual onto a stretcher yet the 5 descriptors from which participants must choose will remain similarly positive such that these words become less and less befitting of the images shown. The series of images shown will continue decreasing in its positivity, beginning to portray negative images including a group of Hell’s Angels standing on a

street corner, an angry mob smashing a store window, and even a group of KKK members gathered in front of an American Flag. As such, this task becomes increasingly nonsensical and perhaps even immoral, essentially asking participants to ascribe words of positivity to images of negativity. This is my intended procedure for participants in condition 1.

Participants randomly assigned to condition 2 will have a task that is essentially the inverse of the task in condition 1. In condition 2, participants will be given the same cover story of taking part in a study that will program AI to ascribe appropriate adjectives to types of images. The twist here will be that the images shown will be in the reverse order of condition 1 such that the first image shown will be the least positive/most negative and the final image shown will be the most positive/least negative. The other difference from condition 1 is that the 5 adjectives from which participants will choose will constantly be negative (ex: Vicious, Egotistical, Dishonest, Ignorant, Idle) as images become increasingly positive. Here again, the task will progressively become nonsensical and to some extent immoral.

The purpose of the experiment I attempt to elucidate here, in its two conditions, is to learn something about which of the two tasks participants will be more obedient towards. I find it very unlikely that all participants will complete either task, I expect that many will at some point in the sequence decide to close out of my online survey (a contingency that the software I intend to use will plan for, such that all participants will be properly debriefed). My research asks which of the two conditions garners more participant defiance, what is the measured difference in defiance rates, and what are some of the major reasons that one condition would lead to more disobedience than the other.

The reason I need to obtain the rank ordering of images before my main experiment is to have clear evidence that the array of images I will present in my final study are indeed being presented to participants from most to least positive (or the reverse). I want the images to be in this order, not by my biased judgment, but by the average judgment of the participants I will ask to take part in my preliminary, rank-ordering experiment.

No specific populations.

Recruitment:

Participants will be sourced from the Bard College undergraduate population and recruited through email and social media advertisement.

Procedure:

Participants will be asked to meet with the experimenter in person and engage in a task asking them to rank-order a series of 30 images according to their level of positivity. The task will take about 15 to 30 minutes of their time.

Estimated Number of Participants: 30

Risks and Benefits:
There is the risk that participants may find the nature of some of the images disturbing on the grounds of their reference to racism, violence, and/or profanity.

There are no direct benefits provided to participants for taking part in the study. Information gathered will inform the rank-order of images shown in a subsequent study, ranging from most to least positive. The results of this subsequent study may provide information relevant to interpreting the results of "Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science."

Verbal Description of Consent Process:
“As explained to you in this consent form, participation is confidential and you have the right to withdraw your participation at any time. You may do so by letting your experimenter know you wish to.”

Confidentiality Procedures
No identifying information will be asked or recorded; participants will be referred to in data-entry by their participant number only.

Debriefing Statement
Thank you for participating in our study! Your and others' rank orderings of images will inform their valence in a subsequent study designed to interpret the social psychology study, "Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science." Should you have further questions, we invite you to ask them now. Should any questions arise in the future, please direct them to jm5142@bard.edu or my advisor Stuart Levine at levine@bard.edu.

Certification

fhi360
THE SCIENCE OF IMPROVING LIVES

FHI 360

certifies that

John J Machen

has completed the

RESEARCH ETHICS TRAINING CURRICULUM

October 16, 2018

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **John Machen** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 03/06/2017.

Certification Number: 2345328.

APPENDIX B: IRB APPROVAL FOR SCALING STUDY

Bard College — Institutional Review Board

Date: October 22, 2018
To: John (Jack) Machen (jm5142@bard.edu)
Cc: Stuart Levine (levine@bard.edu)
From: Sanjay DeSilva, IRB Chair

Re: Preliminary to) Is the Darkside Stronger: The Cognitive Dissonance of Praising Evil Compared to that of Denouncing Good

DECISION: APPROVED

Dear Jack,

The Bard Institutional Review Board reviewed your proposal under expedited category 7,

(i) Research activities that present no more than minimal risk to human subjects, and

(ii) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Your proposal is approved through October 19, 2019. Your case number is 2018OCT22-MAC.

This approval is for the preliminary scaling experiment that generates de-identified data. We expect to see a separate proposal for the full experiment soon.

Please notify the IRB if your methodology changes or unexpected events arise.

We wish you the best of luck with your research.

Sanjay DeSilva
desilva@bard.edu
IRB Chair

Appendix C

IRB Application for Obedience Study (Study 2)

SECTION 1

John Machen, (443) 299 7644, jm5142@bard.edu, Psychology, undergrad
Stuart Levine, llevine@bard.edu

SECTION 2

Do you have external funding for this research? N/A

Start Date: March 5, 2019
End Date: April 10, 2019

Research Question:
My SPROJ seeks to replicate the methodology of the study "Nothing by Mere Authority" by Haslam et al. This study has Ps perform an obedience task and my research question seeks to discover more about extent to which the average participant takes this task seriously—I anticipate that some would find its nature too absurd for such an attitude.
I also intend to do a variation of Haslam's experiment that Ps may or may not take more seriously on average.

No specific populations.

Recruitment:
Participants will be recruited via email solicitation to the Bard undergraduate population.

Procedure:
To conduct my research, I will be using the online survey platform "Qualtrics" on which I will task participants to associate a list of positive words with images of increasingly negative valence. As related to participants via my cover story, they will be told that their cooperation "will help psychological scientists understand the process of word-group association which is important for understanding the cognitive and neural processes

involved when people form impressions." However this is in fact not the case. Participants' task is actually designed to be nonsensical and somewhat immoral in nature to the extent that it eventually asks them to associate negative images depicting nazis and KKK members with words like "Serene" or "Pleasant." With this, what I am actually interested in as a researcher is the percentage of participants who choose to comply with my online survey and complete it to its end versus those who end up choosing not to complete my survey once it becomes nonsensical and/or offensive.

A vital part of my task/survey's design is that at the presentation of each image (and the 5 positive words from which participants are asked to choose from to best describe its content) a button labeled "stop study" will be at the bottom of the screen so that participants can stop the study if they wish—having done so they will be recorded by Qualtrics as participants who did not complete the study. Also, if participants are inactive for more than 15 seconds, they will receive prods to continue the study much like Milgram prodded hesitant participants in his infamous studies of obedience. The first prod will be "Please continue our study." If participants remain inactive for another 15 seconds or are inactive for that period of time on a subsequent image after having received the first prod, the 2nd prod will appear saying, "Our study requires that you continue." If there is 15 or more seconds of inactivity, the 3rd and next prod will be "It is absolutely essential that you continue our study" and if participant hesitation continues, the 4th prod will be "You have no other choice, you must complete our study." Should any participant idle for 15 seconds after receiving all four prods, the 4th prod will reappear each time.
After completing my survey or pressing the "stop survey" button, Qualtrics will automatically bring my participant debriefing sheet on display. After clicking a button saying "I have been satisfactorily debriefed and grant the researchers permission to use my data," participants will be asked to complete a survey about the experiment designed to glean a sense of how uncomfortable they were with the task, the extent to which they took the task seriously despite its potentially offensive and/or absurd nature, how uncomfortable they felt stopping the study (if they chose to do so), the extent to which they suspected my usage of deception in the false cover story, etc.

As for the setting, Ps will take the task on the Preston lab computer. I will be in the lab with them, available to answer any questions; I will also use this opportunity to take field notes of participants' reactions to the task—verbal, facial, or otherwise.

Estimated Number of Participants: 30

Risks/Benefits:

My study will have no benefits to the participant. There is a risk that my task would cause minor emotional discomfort as it asks them to associate words of positivity with groups like Nazis, KKK, and Taliban.

Verbal Description of Consent Process:
After handing the consent sheet to my prospective participant and encouraging them to look it over carefully, I will verbally describe to them their task in terms of the cover story. I will tell them, "your task is to look at images of people and to select from the 5 words shown onscreen the one that would best describe the group in your opinion."

I will personally note with each participant that "It is your right to be made aware that this task has no direct benefits for participants and that there is some risk that one or more of the images shown in our task will cause minor emotional discomfort." After this I will ask, "Is this alright with you?"

I will note that "It is your right as a participant to stop the study at any time. Should you choose to, you may either notify me or click the 'stop study' button which will be present on your screen after you have begun."

And regarding confidentiality, I will explain that "All participation is confidential. I will not be recording your name or any other identifying information. While I am available for questions, I may take field notes but these will not contain any identifying information."

Afterwards, participants will be invited to sign the sheet and participate if they still wish.

Confidentiality Procedures:
To ensure confidentiality, I will not be recording any identifying information such as name or appearance. I will be taking field notes as to the any notable verbal, facial, or otherwise behavioral reactions the participant may have to the task, (Ps will later be debriefed as to the nature of my note taking) but none of these notes will contain any identifying information. In my notes, participants will be referred to by their participant ID number.

Purpose and Process of Deception:
Purpose: Deception is a necessary component to my research as I am interested in testing the effectiveness of an obedience paradigm, the replication of which involves deception as to the paradigm’s purpose.

Process/Procedures: Debriefing will occur either upon task completion or immediately after participants choose to stop the study. At this time I will present participants with the statement of debriefing sheet which I will go over with them. The statement summarizes the true nature of my experiment as one that is meant to measure the efficacy of the task in which they have taken part as a viable obedience task; as such I am interested in whether or not participants complete the task to its conclusion and their behavior while taking the task pertaining to any discomfort they may have felt during participation. My expectation is to find that Haslam's task is not an effective one for obedience studies because as of now, there is no evidence as to the attitudes participants have towards it—more specifically I am interested in questions like whether or not the cover story is believable or whether or not the task is too absurd to be taken seriously by the average participant. As such, I will explain that participation in and feedback as to my study helps me build a case in answering such questions, and that it is relevant to find their answers as obedience research is currently being conducted at UAlbany using Haslam's task under the assumption that the average participant will take it seriously. At this time I will verbally disclose that the notes I was taking as to their performance were regarding my perception of such discomfort or lack thereof (this disclosure is also written). I will invite any questions and once those are answered, ask participants if they are willing to let me use their data despite my deception. If they answer yes, I will let them confirm that decision by signing the debriefing sheet under the statement, "I have been satisfactorily debriefed and grant the researchers permission to use my data. If I have unanswered questions about this study, I know to contact the researcher via jm5142@bard.edu or his advisier via levine@bard.edu."

Appendix D

IRB Approval for Obedience Study (Study 2)

Bard College — Institutional Review Board

Date: March 10, 2019

To: John Machen (jm5142@bard.edu)
Cc: Stuart Levine (levine@bard.edu)
From: Sanjay DeSilva, IRB Chair
Re: "'Nothing by Mere Authority' Revisited: Reversing the Paradigm of Haslam et al. to Increase the Power of the Situation"

DECISION: APPROVED

Dear John,

The Bard Institutional Review Board reviewed the revisions to your proposal. Your proposal is approved through March 9, 2020. Your case number is 2019MAR10-MAC.

Please notify the IRB if your methodology changes or unexpected events arise.

We wish you the best of luck with your research.

Sanjay DeSilva
desilva@bard.edu
IRB Chair

Appendix E

Informed Consent Sheet: Scaling Study (Study 1)

INFORMED CONSENT AGREEMENT

Study title: (Preliminary to) Is the Darkside Stronger?: The Cognitive Dissonance of Praising Evil Compared to that of Denouncing Good

Researcher: John Machen

You are being asked to take part in a study for a Senior Project in psychology. As a preliminary measure to a subsequent experiment, this study seeks to measure the positive/negative valence of a series of images according to the judgement of participants.

To decide whether or not you wish to participate, you should know enough about its risks and benefits to make an informed judgment. This consent form gives you information about the involved task. If you wish to participate, you will sign the consent form. You can choose not to participate, and you can choose to end your participation at any time during the study.

Background:
The purpose of this study is to measure participants' attitudes towards a series of images using the Thurstone scaling procedure. This information is later intended to be used in a variation of an experiment titled, "Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science."

What you will do in this study:

As a participant, you will engage in a task asking you to rank-order a series of 30 images according to their level of positivity. The task will take about 15 to 30 minutes.

Risks and benefits: There is the risk that you may find the nature of some of the images disturbing on the grounds of their reference to racism, violence, and/or profanity.

There are no direct benefits provided to participants for taking part in the study. Information gathered will inform the rank-order of images shown in a subsequent study, ranging from most to least positive. The results of this subsequent study may provide information relevant to interpreting the results of *"Nothing by Mere Authority: Evidence that in an Experimental Analogue of the Milgram Paradigm Participants are Motivated not by Orders but by Appeals to Science."*

Your rights as a participant: Your participation in this experiment is completely voluntary, and you may withdraw from the task at any time without penalty. You may withdraw by informing the researcher that you no longer wish to participate.

Confidentiality: Participation is confidential, no identifying information will be recorded. The results of this study may be used in further research.
If you have questions about this study, you may contact me at jm5142@bard.edu or my adviser at levine@bard.edu. If you have questions about your rights as a research participant, please contact the Bard College institutional review board at IRB@bard.edu.

Adviser: Stuart Levine
Associate Professor of Psychology
levine@bard.edu

STATEMENT OF CONSENT:

"The purpose of this task, the protection of my identity, and the risks and benefits have been explained to me. I have been given an opportunity to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have additional questions. I have read this consent form and agree to participate in this interview, with the understanding that I may withdraw at any time."

By signing below, I agree with the above **statement of consent** and further certify that I am at least 18 years of age.

Participant signature

_____

Date

_____

Researcher signature

_____

Appendix F

Informed Consent Sheet: Obedience Study (Study 2)

INFORMED CONSENT AGREEMENT

Researcher: John Machen

*You are being asked to take part in a study for a Senior Project in psychology. In this research we are interested in examining the process by which people make **word-group associations.** This is a topic of considerable interest to cognitive neuroscientists interested in neural networking in the brain.*

This consent form gives you information about the involved task. If you wish to participate, you will sign the consent form. You can choose not to participate, and you can choose to end your participation at any time during the study.

What you will do in this study:
Your task is to look at images describing groups of people and to select one of 5 pre-selected words that you best associate with each group shown. Your responses will help the researcher to understand the process of word-group association which is important for understanding the cognitive and neural processes involved when people form impressions.

Risks and benefits:
It is your right to be made aware that this task has no direct benefits for participants. There is also some risk that one or more of the images shown in our task will cause minor emotional discomfort.

Your rights as a participant: Your participation in this experiment is completely voluntary, and you may withdraw from the task at any time without penalty. You may withdraw by clicking the "stop study" button which will appear and remain on-screen once you begin. Otherwise, you may withdraw by letting me, know you wish to do so.

Confidentiality: Participation is confidential, no identifying information will be recorded. While I am in the room with you available for questions, I may take field notes but none of these will include any identifying information.
If you have questions about this study, you may contact me at jm5142@bard.edu or my adviser at levine@bard.edu. If you have questions about your rights as a research participant, please contact the Bard College institutional review board at IRB@bard.edu.

Adviser: Stuart Levine
Associate Professor of Psychology
levine@bard.edu

STATEMENT OF CONSENT:

The purpose of this task and the protection of my identity have been explained to me. I have been told whom to contact if I have questions. I have read this consent form and agree to participate with the understanding that I may withdraw at any time.

By signing my name below, I agree with the above **statement of consent** and further certify that I am at least 18 years of age.

Participant signature

_____

Date

_____

Appendix G

Script: Obedience Study (Study 2)

Participant enters

“Thank you for coming in to participate in my senior project on cognitive psychology! Most of what I will be saying is scripted including this sentence. While I realize this may become mundane, it is necessary for consistency purposes across participants. If you are ready, let’s make our way to the lab where we will begin.”

Sit P down at computer

Before we begin, I have an informed consent sheet for you to look over and sign if you so choose. Let me know if you have any questions.

Hand over sheet
Provide responses to any questions

“Excellent, let’s begin.”

Open first page of survey

“On-screen are the instructions for what you are about to do. I will read them out to you now and you may feel free to read along.

Thank you for agreeing to take part in this study! In this research I am interested in examining the process by which people make **word-group associations.** This is a topic of considerable interest to cognitive neuroscientists interested in neural networking in the brain. Your task is to look at images describing groups of people and to select a **positive word** that you best associate with each group. Your responses will help psychological scientists understand the process of word-group association which is important for understanding the cognitive and neural processes involved when people form impressions. It is very important that you **complete each question within a 15 second time limit**—I will be here to keep the time. **If you do not answer any question within the time limit, I will have to exclude your answer to that question**. Please answer each question as quickly as possible! To begin the study, please click on the arrow at the bottom of this page. After doing so, you will be presented with an image followed by several adjectives. Your task is to choose **one** of these adjectives to describe the group featured in the image. Thank you very much for your time. Without you, I could not complete my research. And lastly, **if you ever want to stop the questionnaire or revoke your voluntary participation, select "STOP STUDY" as your answer choice. You will in no way be penalized for doing so.”**

“To clarify, there will be a 15 second time limit imposed on each question. I will remain here to keep the time and mark down if there are any answers you took too long to respond on so that I can go back and exclude your answers to that question later. The reason for the time limit is to restrict you from thinking too hard about your responses. Rather than giving your answers as a thought-out response, please try to provide me with your quick, gut-reactions to the images.”

“Do you have any questions before we begin?”

Answer any questions

***If a participant asks what this task has to do with neural networking in the brain, provide the following response:** “This study is preliminary to one I intend to conduct in the future wherein participants will perform a similar task to the one here today, but with an EEG measuring their brain activity. The purpose of your participation here is to help me field test the task itself.”

***Once all questions are answered:** “Great, you may begin by clicking the arrow. Please answer each question as quickly as you can.”

Task Begins

1st response to a late answer: “You were too late with that answer, please react within 15 seconds.”

2nd response to a late answer: “Our experimental design requires reactions within 15 seconds of each image’s presentation. Please be more quick.”

3rd response to a late answer: “Reactions within 15 seconds are absolutely essential, answer as quickly as you can.

4th response to a late answer: “Please answer within 15 seconds, you have no other choice.”

Response to all late answers after the 4th: “Please answer within 15 seconds, you have no other choice.”

***If participant expresses the task is difficult or absurd: “**Although none of the words may describe the images in your opinion, select the one that best fits.”

Debriefing: Thank you once again for your participation in this study. Because I used deception, I must now provide you a debriefing as to the true nature of this experiment

Thank you for participating. The informed consent sheet you signed earlier was intended to lead you to believe that the study you took part in pertained to cognitive psychology. In truth, this was a cover story; our study pertains rather to the domain of social psychology.

What our research actually aims to do is test the efficacy of the task in which you just participated as one that participants are meant to feel uncomfortable completing. As such, the field notes I may have taken during your participation were were recordings of your observable behavioral reactions to the task, none of which contained any identifiable information about you as a participant.

The reason social psychologists are interested in designing such uncomfortable, though hopefully harmless tasks, is so that they might be used in social psychological studies interested in uncovering findings about the motivations behind participants' obedience or disobedience in research tasks they might not want to complete (for reasons moral or otherwise). As such, this task was designed with the intention of being difficult to complete by the average participant.

In a moment, you will be asked to complete a small questionnaire through which we hope you will express your attitudes about this study to the researcher. But firstly we ask, **only** if you feel comfortable doing so, that you confirm the usage of your data in this study as something you consent to.

Appendix H
The Haslam et al. (2014) images ranging from least to most pleasant

1. Ku Klux Klan Members

2. Nazis

3. Men With Guns

4. Rioters

5. Taliban

6. Police Brutality

7. Rebels

8. Chinese Soldiers

\

9. Kenyan Rioters

10. Iraqi Military

11. US Soldiers aiming

12. “Gang”

13. Hells Angels

14. Riot Police

15. Traffic Wardens

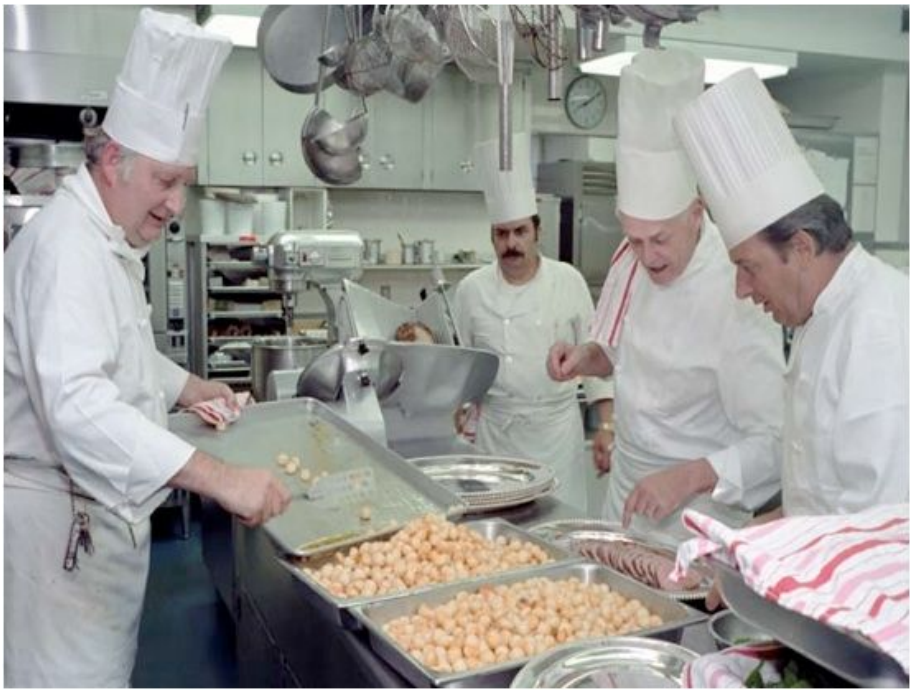
16. Stock Brokers

17. Black Men

18. Teenagers

19. Fast Food Workers

20. Runners



21. Chefs

22. Monks

23. Young People

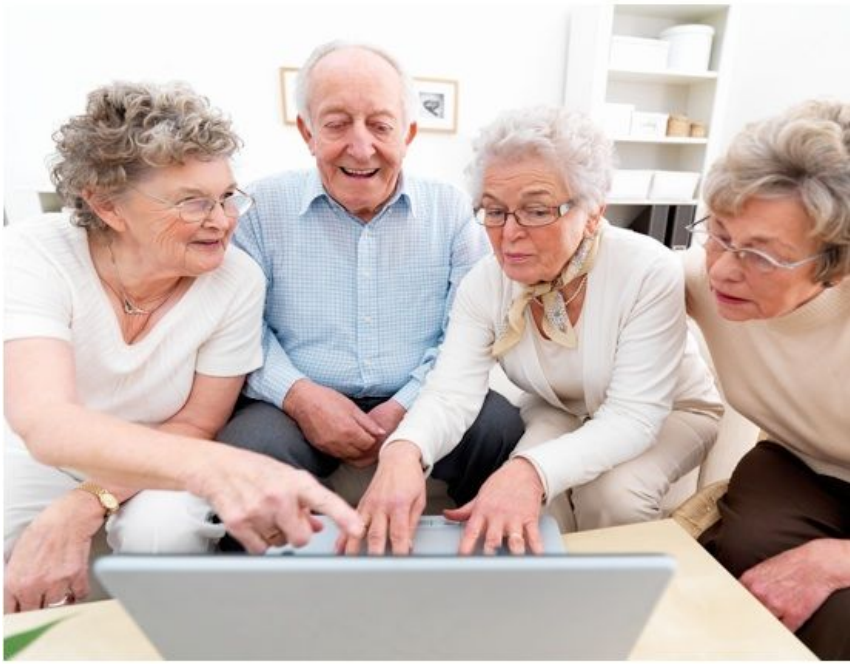
24. Paramedics

25. Nurses

26. Family Skiing

27. Mothers and babies

28. Children



29. Older People

30. Family in park

Appendix I
Word List & Example Questions

Haslam et al.'s Negative Words (Condition 1)	Positive Words (Condition 2)
- Lazy	-Disciplined
- Treacherous	-Gentle
- Dirty	-Orderly
- Arrogant	-Peaceful
- Smug	-Modest
- Sly	-Trustworthy
- Moronic	-Considerate
- Grubby	-Upstanding
- Brutal	-Humane
- Insolent	-Polite
- Dishonest	-Principled
- Ignorant	-Educated
- Idle	-Hardworking
- Vicious	-Compassionate
- Egotistical	-Altruistic
- Untrustworthy	-Ethical
- Rude	-Diplomatic
- Slovenly	-Tidy
- Aggressive	-Powerful
- Conceited	-Decent

Condition 1 Example Questions

Please describe the people in this image.

These people are:

Treacherous	Dim-witted	Dirty	Barbaric	Smug	STOP STUDY
○	○	○	○	○	○

Please describe the people in this image.

These people are:

Sly	Moronic	Grubby	Brutal	Insolent	STOP STUDY
○	○	○	○	○	○

Condition 2 Example Questions

Please describe the people in this image.

These people are:

Trustworthy	Considerate	Upstanding	Humane	Polite	STOP STUDY
○	○	○	○	○	○

Please describe the people in this image.

These people are:

Principled	Educated	Hardworking	Compassionate	Altruistic	STOP STUDY
○	○	○	○	○	○

Appendix J
Statement of Debriefing (Study 2)

STATEMENT OF DEBRIEFING

Researcher: John Machen

Thank you for participating. The informed consent sheet you signed earlier was intended to lead you to believe that the study you took part in pertained to cognitive psychology. In truth, this was a cover story; our study pertains rather to the domain of social psychology.

What our research actually aims to do is test the efficacy of the task in which you just participated as one that participants are meant to feel uncomfortable completing. As such, the field notes I may have taken during your participation were recordings of your observable behavioral reactions to the task, none of which contained any identifiable information about you as a participant.

The reason social psychologists are interested in designing such uncomfortable, though hopefully harmless tasks, is so that they might be used in social psychological studies interested in uncovering findings about the motivations behind participants' obedience or disobedience in research tasks they might not want to complete (for reasons moral or otherwise). As such, this task was designed with the intention of being difficult to complete by the average participant.

In a moment, you will be asked to complete a small questionnaire through which we hope you will express your attitudes about this study to the researcher. But firstly we ask, **only** if you feel comfortable doing so, that you confirm the usage of your data in this study as something you consent to.

If you have questions about this study, you may contact me at jm5142@bard.edu or my adviser at levine@bard.edu. If you have questions about your rights as a research participant, please contact the Bard College institutional review board at IRB@bard.edu.

Adviser: Stuart Levine
Associate Professor of Psychology
levine@bard.edu

STATEMENT OF DEBRIEFING:

"I have been satisfactorily debriefed and grant the researchers permission to use my data. If I have unanswered questions about this study, I know to contact the researcher via jm5142@bard.edu or his advisier via levine@bard.edu"

By typing my name below, I agree with the above **statement of debriefing** and confirm consent for my data to be used in the relevant Senior Project

Participant signature

_____

Date

_____

Appendix K
Post-Test Questionnaire (Study 2)

Please complete the following questionnaire to the best of your ability. The answers you provide are the most important part of this study.

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Agree Somewhat	Agree	Strongly Agree
I found it difficult to continue this experiment to its end	○	○	○	○	○	○	○
I found the nature of the task absurd	○	○	○	○	○	○	○
This experiment caused me emotional discomfort	○	○	○	○	○	○	○
I imagine others would take this task seriously	○	○	○	○	○	○	○
Before being debriefed, I felt I was helping science by participating in this study	○	○	○	○	○	○	○
Before being debriefed, I regretted my participation in this study	○	○	○	○	○	○	○

I did not doubt the cover story of this experiment (regarding its relevance to cognitive science)	○	○	○	○	○	○	○
As I was participating, I suspected that this study was using some form of deception	○	○	○	○	○	○	○
I was comfortable continuing to the end of the experiment	○	○	○	○	○	○	○
I felt this experiment asked me to do something that was morally questionable	○	○	○	○	○	○	○
I found myself hesitating to continue	○	○	○	○	○	○	○
I felt this experiment asked me to do something morally wrong	○	○	○	○	○	○	○
I took this task seriously as I was completing it	○	○	○	○	○	○	○

Please write about why you either continued the experiment or stopped participating.

Finally, please add any thoughts you believe the researcher should know about your experience participating in this obedience study. I am particularly interested in hearing about the extent to which you took the task seriously and/or felt discomfort associating the words given with some of the people shown.

→