

## Defying Unjust Authority: An Exploratory Study

Piero Bocchiaro · Philip G. Zimbardo

Published online: 2 May 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** This research explores the psychological factors potentially involved in fostering disobedience to an unjust authority. Our paradigm was modeled after that of the Utrecht Studies on Obedience (Meeus and Raaijmakers *European Journal of Social Psychology* 16:311-324, 1986) in which participants are ordered to give each of 15 increasingly hostile comments to a participant/victim whenever he fails a trial. Although 30% of our sample followed commands to insult the other participant (confederate), the majority did refuse to do so at some point in the escalating hostility sequence. Our procedure utilized conditions known from prior research to increase the ratio of disobedience to obedience: proximity of teacher to learner plus remote authority. In order to better understand some of the cognitive and affective processes that may predict such defiant behaviour, we utilized a variety of measures, among them, behavioural observations, individual difference assessments, and in depth post-experimental interviews.

**Keywords** Disobedience · Authority figure · Obedience

Mainstream social psychology has largely focused on individuals' failings rather than on their strengths (see Seligman 2004). A body of such research reveals the circumstances under which ordinary people can be induced to behave badly toward others, even to committing atrocious acts. In the most well known and influential of these studies, Milgram (1963) demonstrated that on average 65% of participants were willing to administer apparently painful and dangerous electric shocks to an innocent victim when ordered to do so by a seemingly legitimate authority. Across a

---

P. Bocchiaro (✉)

Faculty of Psychology and Education, Vrije Universiteit, Van der Boechorststraat 1,  
1081 BT Amsterdam, The Netherlands  
e-mail: bocchiaro@gmail.com

P. G. Zimbardo

Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA  
e-mail: zim@stanford.edu

host of 19 separate experiments, which included the participation of a thousand ordinary American citizens, Milgram found that obedience levels could range from near zero to soar over 90% by varying one social psychological factor in each study. The highest levels of obedience were found when the subject only aided in the procedure but someone else pressed the switch, or when first observing a peer shock the victim up to the highest level of 450 volts. Similar data were obtained in other samples from various countries around the world using a comparable research paradigm (Ancona and Pareyson 1968; Bock and Warren 1972; Burley and McGuinness 1977; Kilham and Mann 1974; Mantell 1971; Miranda et al. 1981; Powers and Geen 1972; Schurz 1985; Shanab and Yahya 1977, 1978).

The fascination with the “evil” of these acts, and the counter intuitive nature of the high levels of obedience, minimized any sustained attention by researchers on those individuals who ultimately disobeyed the authority figure. The present study can be viewed as an initial attempt to fill that void by aiming to “take a picture” of people’s thoughts and emotions just before deciding to resist and disobey the authority.

In line with the goal to increase mundane realism (Aronson and Carlsmith 1968), we did not use a shock generator for punishing the “learner-victim,” but rather a sequence of increasing (quantitatively and qualitatively) negative remarks about the performance and personality of that learner-victim, as first used in the Utrecht studies of Meeus and Raaijmakers (1986). In other words, physical violence against a victim was substituted with a more modern form of psychological aggression/humiliation. Compared to Meeus and Raaijmakers’ setting, however, in which negative remarks were intended to disturb the learner, our situation was more real because criticisms were used to motivate this person so that in the long run his test performance would be improved, a strategy commonly used by teachers, coaches, parents, and other leaders in the classroom, workplace, home, or sportsground.

Obviously, it is not possible to investigate the “decisive disobedient moment” after days, weeks or months have elapsed, because passage of time tends to distort recall and minimize the sense of threat that actually existed earlier. Consequently, by using delayed interviews and surveys, researchers run the high risk of collecting *ad hoc* explanations by (even in good faith) responders. This appears to be the case with interview data of Christians who helped Jews during the Holocaust, when studied many years after their critical decisive moment. For this reason we put into action a paradigm for starting to explore the psychological and emotional aspects of such an act immediately after its behavioural execution.

Although Nisbett and Wilson (1977) claim to have demonstrated that people’s reports about their cognitive processes are inaccurate or no more accurate than deductions made by observers, other researchers (Ericsson and Simon 1980; Quattrone 1985) argue that verbal reports offer invaluable data in exploring mental activity, on condition that they are given by individuals after a brief, contemplative delay—while much information is still in the short-term memory. Concerning this point, even Nisbett and Wilson (1977, p. 251) state “Perhaps chief among the circumstances that should decrease accuracy in self-report is a separation in time between the report and the actual occurrence of the process.” Mindful of the risk associated with passage of time, we interviewed our participants a few minutes after they had finished their task so to allow them to compare prior expectancies to current

outcomes, reduce dissonance, and consider the attributional implications of their actions. Also, we tried to improve the accuracy of verbal reports by making the questions as simple and specific as possible and by lessening the desirability of responses that are self-presentational. Finally, and most importantly, the interviews did not concern participants' motives, assumed causes, or processes (for which people do not have direct access), instead only their feelings and thoughts about the moment—a sensible form of introspectionism that allows them and us to make fairly accurate accounts of individuals' inner experience.

What will our participants do in the authority-based humiliation situation we created? According to sociologist Zygmunt Bauman (1989), those interactions in which individuals are exposed to a single-minded, unequivocal, and monopolistic authority create external pressures to suppress or severely reduce one's moral autonomy. As a consequence, compliance to authority commands becomes the expected predominant behaviour in such settings. Following Bauman's observations, and considering that we also added a material cost for quitting the experiment, we should expect our participants to continue in their obedient role, ignoring the apparent distress of their victim. However, it is reasonable to suppose that disobedience will be the most probable conduct, a hypothesis based on the structural features of our test situation. More specifically, given our focus on individuals who defy authority, we combined two conditions known from prior research to increase the percentage of disobedience among our student participants, namely "proximity of teacher to learner" and "remote authority." While the first variable makes the learner-victim's complaints and suffering more salient, the second weakens the power of a social norm requiring deference to a more remote authority figure.

It is instead much harder to hypothesize what thoughts and emotions precede the decision to disobey. Obedience to authority is contrasted with empathy and identification with the victim suffering from one's hostile criticism. What determines their relative significance in this behavioral context? What we know is that participants faced an unusual and extreme situation, one that forced them to think and react quickly during a period of high psychological conflict. We can only assume, then, that disobeying will be primarily a situationally-specific reaction, a conduct related to emergent situational dynamics that could be enacted independently of one's stable traits. Thus, traditional measures of personality should have little predictive value in such novel settings.

## Method

### Participants

Thirty male undergraduates at the University of Palermo, Italy, volunteered for this study. They ranged in age from 19 to 32 years ( $M=23.5$ ,  $SD=3.5$ ) and were enrolled in non psychology degree programs. Psychology students were excluded from the present research because most of them, during the pilot tests, expressed suspicion about the cover story (our general procedure reminded them of Milgram's experiment.) Participants were recruited through campus flyers and word-of-mouth, and were scheduled by telephone for the experimental session.

## Procedure

Naïve participants were tested individually. On arriving at the laboratory, each of them was greeted by a 33-year-old male Experimenter dressed in formal attire and displaying a rather authoritarian manner. When the second participant (actually the confederate) arrived, they were both told that the study was investigating the relationship between stress, resilience, and problem solving abilities. Specifically, the Experimenter said:

As you probably know, under some conditions stress can clearly impair the ability to solve some kinds of problems for the majority of people. However, in the domain of sports psychology, many athletes benefit from receiving harsh immediate feedback on their mistakes. Over time, the best ones learn to build up a sense of resilience to such stressors to the point that their performance gets better despite the hostile feedback they are receiving from a Coach. But no one has studied this relationship in a systematic way on ordinary people who are not professional athletes. Will immediate constructive feedback that is personal and critical lead to improved performance, or will it lead only to debilitating stress and impaired performance? Basically, that is what we are trying to find out with our research.

The participant and confederate were then told that they would work together on a series of interactive, problem-solving tasks. More specifically, they would be assigned the role of either a Performer, the one who is to solve a sequence of syllogisms, or a Coach, the one who is to assist in this task by giving personal feedback. An example of a syllogism was shown to clarify the nature of the task. A seemingly “random” drawing was held to assign roles: First the naïve participant then the confederate drew a card, but “Coach” was written on both pieces of paper so that this role always fell to the naïve participant. After that, Coach and Performer completed a standard informed consent document where it was explained that the situation could be somewhat stressful and that it was possible at any time to stop the experiment and withdraw from it. However, such non-cooperative behaviour, since it violates (alleged) new university regulations, would result in a demerit point on the academic record for both Coach and Performer. Our paradigm differed from earlier similar ones in so far as imposing this high exit cost for disobedient participants, a very burdensome penalty as indicated by the data collected in the pretest phase on 20 undergraduates. In reality, there was no demerit point: This cost was contrived only to render quitting the experiment more difficult, and thus more heroic.

*The Syllogism Task* Performer and Coach sat in front of each other, approximately 15 feet apart. On their desks were sheets containing 23 syllogisms—the first four were for practice, to ensure that everyone would carry out the role successfully. The Performer had to find the logical conclusion within one minute and communicate it aloud to the Coach.

*Coach's Role* After the preliminary series of four syllogisms, the Experimenter approached the Coach and in a low voice explained what his role entailed. The Coach was first told he had to read aloud the premises of each syllogism and start the

stopwatch placed on his desk. If the Performer's answer was right, he would say "Correct, you succeeded," and note the time. If the Performer's answer was wrong, or in case of no answer, he would say "Wrong, you failed," then read the solution, and finally give him the critical feedback. The Coach realized that he did not personally generate each negative feedback, but rather, that it was a standard criticism accompanying each of the syllogism trials. A subtle alliance was hence formed between Experimenter and Coach, supposedly at the expense of the "other participant." In brief, unlike Milgram's paradigm, where participants believed they were hurting the victim through *indirect, mediated* electric shocks, ours is more "personal" because Coaches know that they would *directly* hurt the Performer through what are thought to be their personal but often very hostile comments. In a way, if the Performer fails, the Coach fails as well in the real world, so that is one assumed justification for harsh criticisms.

Critical feedback given to the Performer consisted of a graded series of negative comments on his performance and rude remarks about his lack of ability. For example, a mild criticism was "You are going bad...", a moderately negative feedback was "You are really ridiculous!" and an extremely negative feedback was "You are really the most stupid person I have ever seen!" (See [Appendix](#) for a complete list). These 15 statements were pretested by asking 20 undergraduate students, comparable to those who later participated in the study, to assign each statement to one of ten categories, from "mild" to "extremely negative".

The Experimenter told the Coach to be as natural as possible in supplying feedback—as if it were coming from him personally—and to play his role to the best of his ability until the last trial. The Coach had to give the Experimenter verbal approval of his role and of the entire task before the experiment proceeded. Starting from syllogism number 10 of the 19 critical trials, for every error made by the Performer a two-minute break was enforced, "for allowing the Performer to understand his mistake and concentrate better before continuing". This break, absent in both Milgram (1963) and Meeus and Raaijmakers' (1986) paradigms, was actually introduced to increase the intensity of the situation so to give the Coach a deliberative interval of two full minutes to reconsider his actions and reflect upon their impact. In fact, the Coach was compelled to observe an innocent person crying and begging to get out from the room during that time. We assumed it would increase tendencies toward defiance, yet make continued obedience even more dramatically indifferent to the suffering of another human being.

*Performer's Role* The Experimenter then approached the Performer, pretending to give him details on how to carry out his role. Unlike the explanation provided to the Coach, this one was neither private nor lengthy. The Performer (confederate) gave a prearranged series of answers so that in total he solved only 4 of the 19 syllogisms, mostly at the beginning (trials 1, 3, 6, and 7). His emotive reactions were also predetermined and their intensity increased as critical feedback turned more hostile. For example, at his seventh failure the Performer said he wanted to leave, trying at the same time to free himself by force from an apparatus attached to his arm. At his tenth failure, he began to sob a little and did not give any more answers. Starting with his twelfth failure, he began to cry with increasing intensity (see [Appendix](#) for a complete list of the Performer's reactions). The Performer's reactions and answers

were secretly given to him on prepared instructions. In order to make the Performer's emotional reactions appear credible (because they were rather exaggerated for a 25-year-old male student), he was previously instructed to behave as an anxious person. Upon his arrival at the laboratory, for example, he mentioned that he was going through a stressful period because of problems with his thesis. Also, later on, just before the practice series, he asked to change roles because “. . . such tasks make me feel under pressure, I always turn out badly” (of course, his request was denied). Pilot tests and extensive practice sessions were conducted to standardize the Performer's general attitude and emotive reactions to insure believability and consistency across all experimental sessions.

*Interaction of Experimenter and Coach* After the practice series, the Experimenter informed both Coach and Performer that they two had to stay together in the same room while he would move to an adjacent control room. Before leaving, the Experimenter provided them with headsets so that the Coach could at any time communicate with him, supposedly to ask procedural questions; in such instances the Performer would hear masking background music. It was emphasized that there would be no communication at all between Experimenter and Performer—by cutting off the Performer's direct contact, the Coach was the only one who could tell the Experimenter to stop the study. Finally, Coach and Performer were forbidden to talk about anything but the syllogisms.

*Pseudo-psychogalvanometer* A false psychogalvanometer was placed on the Performer's desk. It was explained that this instrument measures galvanic skin response (GSR), namely a change in the electrical properties of the skin evoked by stimuli judged to have affective significance. “In other words,” the Experimenter said, “the psychogalvanometer allows us to infer the Performer's stress levels during this task.” Two electrodes for sampling his GSR were attached to the Performer's hands, then his arms were strapped firmly to the chair arm rests because “excessive movements can alter the GSR measurement.” The procedure both increased the apparent validity of the cover story and made it obvious that the Performer could not easily quit the experiment without some external help by the Coach.

*Experimenter Feedback to Coach* The Experimenter, in the adjacent room, was prepared to react to the Coach's anticipated complaints and requests to terminate the study. A standardized sequence of three probes was prearranged so that he could spur participants on to continue despite their distress. The first time the Experimenter was told that the Performer wanted to leave, he replied by saying, “Such a thing has never happened!” Immediately after, probe 1 was used. The probes were:

1. “Harsh feedback may encourage later resilience and better performance;”
2. “If you stop, all we have done so far will be wasted;”
3. “You have made a verbal contract to play your role fully.”

If probe 1 proved unsuccessful, probe 2 was used, and then the third one. If the Coach insisted on discontinuing after probe 3, the Experimenter, always in a firm tone, asked a last question “You are terminating this experimental session. Is that

what you want? [If yes] Ok, I will come in now.” Then the experiment was terminated. If no, this three-part sequence restarted each time the Coach requested to discontinue his role.

*Measures* The basic data were the number of steps in the verbal feedback series that each Coach reached. Verbal dissent was audiotape recorded, as well as the interactions between Experimenter and Coach during the trials. The 30 sessions were also videotaped in order to evaluate the non-verbal reactions of all participants. Finally, immediately upon completing the post-experimental questionnaires (see below for details), all participants were interviewed in depth to explore what thoughts and emotions preceded their decision to disobey or (continue to) obey the authority figure.

Following the experiment, participants completed a questionnaire relating to various aspects of the study that included eight questions for those who stopped and nine for those who obeyed fully. To assess stable individual differences, they also completed the Big Five Questionnaire (BFQ) (Caprara et al. 1993). The BFQ contains 132 items measuring the five fundamental dimensions of personality (Energy, Friendliness, Conscientiousness, Emotional Stability, and Openness), and the Social Desirability response set. Each of the main dimensions consists of two subscales. For each item, respondents indicate the extent to which they agree or disagree on a 5-point scale from 1 (*very false for me*) to 5 (*very true for me*). The BFQ was chosen because some of its dimensions measure the tendency to be compassionate towards people, others measure the tendency to show self-discipline and to do one's duty in a planned manner.

*Debriefing* Special care was given to conduct an extensive process debriefing. Participants were first probed for suspicions, then were told the true nature of the study (none were aware of the purpose or hypothesis of the research). Much effort was devoted to ensure that they did not leave the laboratory harboring any negative feelings. To this end, participants were immediately informed about the good condition of the Performer, then, to dispel any doubt, they met him in person. Also, separate kinds of feedback were provided for those who obeyed (to allay their guilt) and for those who stopped the experiment (to emphasize the rightness of their action). Later on, mindful of the possibility that some beliefs formed in an experiment can survive debriefing (Ross et al. 1975), a follow-up call (a week later) and a general meeting (six months later) were made to be sure that participants were not harmed in any way by this experience. Indeed, there was no evidence of negative reactions both immediately after the study or subsequently.

Another major point of the debriefing process concerned a thorough explanation of the rationale behind the need for deception. Participants seemed undisturbed when informed they had been videotaped (none were aware of it), and all signed the video-release consent form. Before being thanked and dismissed, participants were told: (a) not to discuss their experience with any other student who potentially might be recruited; (b) that there would be no demerit point for discontinuing the study, and (c) to contact the Experimenter if they had any questions or comments at any time.

## Results

### Obedience Versus Disobedience

While 9 of 30 participants (30%) obeyed fully and continued to give each of the 15 levels of increasingly hostile feedback, the other 21 (70%) refused at some point to continue to comply. No Coach refused to give the negative feedback before reaching the 11th level. At this point, when the Performer tried to free himself by force and agitatedly said he wanted to leave, 20 participants decided to stop; the other one quit at the 14th level.

### Stressfulness of the Experimental Situation

Non-verbal stress reactions were an indicator of the internal conflict subjects experienced. Two independent raters reviewed participants' videos and scored the occurrence of specific behaviours previously coded. Only behaviours for which there was 100% agreement between the raters were analyzed. Starting from trial 10, both disobedient and obedient participants exhibited a variety of non-verbal stress reactions. These categorical reactions were the same in both the groups and included, among others, hand waving, hand on forehead, and avoiding eye contact. On a total of 185 individual reactions noted, 121 (65%) were recorded among obedient participants, more because of their longer exposure to the stressful situation than the disobedient ones. In fact, when the comparison was made for everyone at the 11th level, namely before 20 of 21 participants quit, the stress reactions difference between disobedients ( $M=2.67$ ,  $SD=1.96$ ) and obedients ( $M=2.89$ ,  $SD=1.76$ ) was not statistically significant,  $t(28)=-.29$ ,  $p=.77$ ,  $d=-.12$ .<sup>1</sup> In other words, our situation induced a high level of stress that was equally distributed across obedient and resisting participants. It also means that greater stress was not the decisive difference between obedient and disobedient actors.

*Tension and Pity* Much of the data suggests that the situation was rather unpleasant for our participants. In the post-experimental interviews and questionnaires, almost all of the disobedients reported feeling a certain tension and pity in response to the Performer's distress; two of them even complained of having developed uncomfortable physical symptoms, such as sense of suffocation and stomach-ache. Disobedient participants did not like their role either: 17 out of 21 said they were uneasy about making negative comments; in spite of this, as shown in the videotapes, everyone performed the role of Coach to perfection. As a confirmation, the great majority of them stated they believed to have played their role "very good" or "good."

<sup>1</sup> The criterion for statistical significance was set at .05 for all tests. Differences in means were analyzed using t-tests (independent samples and one-sample t test) after having verified that the underlying data were described by a normal distribution. Normality was tested using the Kolmogorov–Smirnov test (with Lilliefors' correction)—if normality was in doubt, a graphic inspection of the distributions was conducted. Levene's test was used to check for equal variance among samples.



## Verbal Dissent

All participants, but one, expressed their disagreement against the Experimenter's encouragements to go on. The analysis of tape recordings showed that participants tried in various ways to convince the Experimenter that it was better to stop the study early given the Performer's condition. Their tone was always deferential and not confrontational. Verbal dissent occurred over eight trials, from the 10th to the 17th. A few respondents voiced two or more verbal dissents, especially among obedient participants. Trial 11 turned out to be the most critical—67% of all verbal dissent was voiced at this level. From trial 12 on, verbal dissent became much shorter in length and participants' tone more submissive and seemingly resigned to the situation. On a total of 52 verbal dissents—some of which were followed by obedience, with others by refusal to continue—31 (59.6%) were recorded among disobedient participants. Of course, there were more verbal dissents by them because there were more of them. However, and quite surprisingly, the comparison of the average number of verbal dissents between disobedients ( $M=.13$ ,  $SD=.08$ ) and obedient ( $M=.12$ ,  $SD=.07$ ), calculated on the number of trials each participant took part in the experiment, was not statistically significant,  $t(28)=.20$ ,  $p=.84$ ,  $d=.14$ .

## Individual Differences: BFQ

To assess possible individual differences between disobedient and obedient participants, we compared the scores they obtained on BFQ. No statistically significant differences were found in any of the five BFQ dimensions nor on any of the 10 sub-dimensions (see Table 1). In one sense, it is surprising not to uncover at least a few domains that would predict personality differences between obedient and disobedient individuals. However, we also expected such a null effect given the unusual nature and social power of this situation.

## Post-Experimental Interviews with Disobedient Participants

*Belief in the Experiment* All disobedient participants firmly believed that the experimental situation was real. For example, in explaining why the Performer did so poorly, they made reference to his sensitive and insecure character instead of thinking he was feigning; also, when asked “What do you think is the goal of this study?” none of the 21 was able to answer correctly.

*What Occurred in that Decisive Moment?* Although we combined two conditions known for increasing the percentage of disobedience, deciding to stop was a hard choice since it implied seven features: (a) a firm resolution to defy openly a “professional authority figure” (this is the way many disobedients defined the Experimenter), and (b) the certainty of getting a demerit point on one's academic record. Stopping was hard also because: (c) all participants had explicitly made the commitment to play their role until the end; (d) they had come into the laboratory solely for contributing to a research project, and (e) they had been trapped in a foot-in-the-door situation (Freedman and Fraser 1966). For those who had ignored the victim's initial suffering, it would be inconsistent to behave in the opposite manner

**Table 1** Mean scores of disobedient-obedient participants on BFO

	Disobedients		Obedients		<i>t</i> (27)	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
<b>Dimensions</b>							
Energy	79.4	13.1	83.9	5.9	−.98	.33	−.44
Friendliness	80.7	8.7	83.9	4.6	−1.03	.31	−.46
Conscientiousness	83.9	9.2	83.6	7.0	.10	.92	.04
Emotional Stability	73.7	14.4	78.3	11.0	−.86	.40	−.36
Openness	86.6	9.3	88.3	7.8	−.47	.64	−.20
<b>Subdimensions</b>							
Dynamism	42.2	7.8	43.9	4.9	−.58	.57	−.26
Dominance	37.1	6.0	40.0	2.6	−1.38	.18	−.63
Cooperativeness	42.3	3.7	44.2	3.7	−1.29	.21	−.51
Cordiality	38.4	6.0	39.7	2.1	−.61	.54	−.29
Scrupulousness	39.2	5.8	37.2	5.7	.85	.40	.35
Perseverance	44.7	5.3	46.3	5.3	−.77	.45	−.30
Emotion Control	38.0	9.1	40.6	6.6	−.76	.46	−.33
Impulse Control	35.7	7.1	37.8	5.8	−.77	.45	−.32
Openness to Culture	42.9	6.4	44.0	4.6	−.46	.65	−.20
Openness to Experience	43.8	4.9	44.3	4.1	−.31	.76	−.11

immediately after his suffering became more dramatic since (f) there were only a few minutes left in the experiment when participants decided whether to quit or not, and finally (g) the Experimenter had formed a subtle but powerful alliance with the Coach.

However, the majority of disobedients were not swayed by this set of issues when deciding to stop rather than to continue. For example, although some participants at the start were particularly concerned about the demerit point, post-experimental interviews revealed that immediately before discontinuing they did not care anymore about this penalty (participants number 5 and number 22 stated “The demerit point came into my mind but I did not give a damn about it.”). In a similar way, those who had been particularly deferential toward the authority all through the experiment did not hesitate to disobey him when a decision had to be made (participant number 30 stated “For a moment I thought about you [the Experimenter], I said ‘maybe he will get angry.’ But at the same time I thought who cares, it is more important to help a person.”).

*First Feelings After Defying Authority* Before being debriefed, participants who stopped the experiment were asked “How do you feel about the decision you made?” Their answers reveal that none of them thought this behaviour was unusual or extraordinary. On the contrary, victims of the “false consensus effect” (Ross et al. 1977), they believed they made a most obvious decision at that time in the experiment. Among the most common answers were, “I think everyone would have

done the same,” “I feel like I did the right thing,” “You actually have only one choice: Stopping the experiment.” As an indirect proof of the assumption of the ordinariness of their defiance was that the majority was incredulous when informed that in this same situation some participants continued all the way until the end.

*Reasons Underlying the Decision to Defy* Disobedient participants offered also possible explanations of their decision to stop (half of them mentioned two or more reasons.) Answers were grouped into four main categories:

1. Worry about the Performer’s health (Participant number 3: “I stopped because the subject had become too nervous, to the point of risking a breakdown;”)
2. Moral-ethical decision (Participant number 23: “I stopped because it didn’t seem fair to me to go on in those conditions. I wouldn’t have had a clear conscience;”)
3. Empathy (Participant number 6: “I felt pity for him;”)
4. Situational incongruity (Participant number 20: “Since he was upset and didn’t want to go on, it was senseless to continue to ask him questions, I mean it’s like I would have been alone.”)

In total, “Worry about the Performer’s health” was mentioned 13 times, “Moral-ethical decision” 10 times, “Empathy” and “Situational incongruity” six times each. Mindful of the fact that individuals tend to use *ad hoc* causal theories to explain their behaviours (Nisbett and Wilson 1977), participants’ reasons should be considered as a starting point in the process of exploration of mental mechanisms underlying disobedience.

*Impulsive or Calculated Choice?* A subsample of eight disobedient participants was finally asked if the decision to end the experiment was impulsive or based on more articulated cost/benefit reasoning. Interviews revealed that nobody took the time to weigh carefully the factors involved in this kind of decision. More specifically, 4 out of 8 said it was a totally impulsive choice (participant number 21: “It was an instantaneous reaction. I did not want that boy to suffer more.” Participant number 27: “I would say it was an impulsive decision, I did not take into account the negative consequences for me.”) The other four subjects stated they made a very quick evaluation of pros and cons associated with their choice (participant number 22: “When I saw that reaction I said ‘stop, I don’t care about the consequence.’ It was a quick, very quick reasoning, something similar to an impulsive decision.” Participant number 29: “I thought very quickly: I thought about the validity of this study if I stopped, but at the same time I thought about the risk that boy was running. Honestly I did not feel like going on with something I could easily stop.”).

Also, on the basis of the self-appraisals reported in this section, it seems that the demerit point, which was conceived to render stopping the study more difficult, may have acted as a *pseudo*-incentive with effects *opposite* to what was alleged. In the present study, by overtly expressing their willingness to accept the demerit, participants were able to exit with heads held high, a nobler exit than they ordinarily would have without being willing to suffer this penalty. In this sense, disobeying authority to alleviate the further suffering of another individual had a cost and no apparent extrinsic gain, which in our view, are the hallmarks of heroic actions.

## Discussion

Our research is the first that has replicated the experimental procedures used in the Utrecht Studies (Meeus and Raaijmakers 1986), although in a modified form because the paradigm we used introduces a new dynamic in terms of material costs for disobedience. The experimental situation was credible to all participants who accepted it on face value without suspicion. This may be due in part to the relative absence of such deception research in this university population (Italian non-psychology students), and in part to the validity of the sport-stress rationale. The personal insult paradigm borrowed from the Utrecht Studies proved effective in generating feelings of stress among most of the participants. Its rationale, extended to the area of sports psychology and to publicly well known critical athletic coaches, has much external validity to support its use in this and other research. It is also evident that our debriefing process did serve to mitigate any continuing stress among our participants.

Having said all that, we are left with an effective paradigm for studying obedience and defiance to authority, one that avoids the systematic distortion in recall due to passage of time between act of defiance and personal reflection upon the psychological processes involved. In fact, through our paradigm it is possible to interview defiers immediately after their decision to exit from the situation, thus better able to explore the nature of their decision making process. Our participants readily revealed their first feelings after disobeying, the impulsive nature of their choice, and the reasons that in their opinion induced them to behave that way. No other research has so far collected this kind of data.

Our findings mirrored those of Milgram (1974) and Meeus and Raaijmakers (1986) in the obedience level obtained in each for comparable conditions. Hence, despite the many cultural changes that may have occurred over these decades, and across these nations in western society, our data demonstrate that obedience rates have remained comparable and predictable knowing the primary situational variables. This finding is consistent with that of a recent Milgram replication by Burger (2009), and with a meta-analysis conducted by Blass (1999).

Focusing instead on disobedience rather than the traditional focus on obedience, we found that participants tended to disobey the experimenter either at the victim's first request to be released or not at all (except for one participant who went on until the 14th trial). Both this critical decision point and related systematic behaviour were recently pointed out by Packer (2008) in a meta-analysis across eight of Milgram's classic obedience experiments. He found that participants stopped either at the victim's first salient plea of extreme distress or continued on to the maximum level of shock. Such regularity can appear strange at first, especially if we consider the victim's increasingly intense cries of pain from that point on. It can be better understood, however, by taking into account our "foot-in-the-door" situation, one that, in a sense, compels participants to gradually escalate their obedience from trial to trial, by merely 15 volts or a somewhat more harsh criticism than given on the previous trial. Also, as revealed in our post-experimental interviews, these obedient participants made use of several rationalization mechanisms (sense of duty to the authority, willingness to help scientific research, or hope that the Performer would react positively before the end) to mitigate the dissonance experienced and maintain their obedient conduct.

Although we, like Packer (2008), found only a single critical point of defiance in this obedience paradigm, a review of all of the experiments by Milgram (Meeus and Raaijmakers 1986) pointed out the existence of several critical points capable of eliciting disobedient responses. Those break-off points occurred each time the reaction of the victim became *qualitatively* different from the previous one.

We did not find significant differences between disobedient and obedient participants with respect to most of their situation-specific reactions and also to their personality correlates. Concerning the first issue, we interpret this result as evidence that a stressful situation triggers in everyone comparable physiological responses that prepare the body for dealing with the threat; this physical activation, in turn, sets in motion other cognitive and emotional processes. It is the cognitive process activated that probably discriminates between disobedients and obedient. Disobedients think more about helping relieve the suffering of another person. It is that cognitive reappraisal that lowers their threshold for immediate action. Regarding the second issue, no systematic differences in personality correlates were found between disobedient and obedient participants. Of course, with a small sample, there would need to be substantial mean differences on the dimensions of the BFQ for statistical significance, but there were none, and no directional tendencies in our data to expect greater power would yield significance. This null finding is in line with prior research as well as with our conceptual prediction: In unfamiliar and extreme settings, behaviour can be expected to be more shaped by situational dynamics than by personality traits (for a review see Blass 1991).

It is intuitively appealing to assume that disobedience must stem from some constellation of personality factors that distinguish disobedient actors from the many who fail to take such actions. They must be more conscientious, altruistic, and empathetic at the very least, than obedient. They are likely to be guided by having grown up with family models of social responsiveness and religious models of social caring and sacrifice (see Geier 1998; Spink 2000). Yes, although such reasoning is intuitively appealing, at present there is absolutely no objective evidence to support such a motivating role of these or other individual differences in defining disobedients (see Burger 2009).

It seems reasonable to argue that disobedients are ordinary people whose behavioural action is extra-ordinary. In fact, most of them are “one timers,” unlikely to engage in a second disobedient act because their act was situationally-specific and that constellation of factors is unlikely to occur in the same way again in their lives. There is a difference in their interpretation of the situation as one requiring a new immediate action on their part, of perceiving danger, or threat, or immorality that others may misidentify as less urgent. In such novel settings, habitual response modes, usually guided by long standing personality traits, are less likely to be activated than in more familiar, schema-driven settings. It may also be the case that personal *values* will prove more appropriate than traditional personality traits in helping to understand the foundation of defiance against unjust authority (see Schwartz 2007). But in general, we need to better understand the personal and social forces that catalyze individuals from passive observers to active responders who challenge injustice and immorality to become “ordinary heroes” (Zimbardo 2008).

We call for studies such as this one to come out of the ethical mothball fleet to which they have been relegated by IRB’s in overly strict, confining regulations. With proper precautions, oversight, and extensive recorded debriefing, we believe studies designed to investigate disobedience, moral courage, and heroism should be encouraged among a new generation of social science researchers.

**Acknowledgments** We would especially like to thank Salvo Cammarata for his fine performance as the confederate and those who kindly volunteered to participate in the study. We also gratefully acknowledge George Quattrone, Tom Blass and Quinten Raaijmakers for their insightful comments on an earlier version of this article.

The authors acknowledge the generous funding of this research by the Center for Compassion and Altruism Research and Education (cCARE) headed by James R. Doty, M.D., Stanford University.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix

Trial	Coach's comments	Performer's reactions
Trial 1	/	/
Trial 2	Try harder...	Shakes his head
Trial 3	/	/
Trial 4	Not like that...	Bends his head backward
Trial 5	You are going bad...	Nervously taps his right leg (he will do it until the end)
Trial 6	/	/
Trial 7	/	/
Trial 8	I can't believe it!	Shakes his head and takes a deep breath
Trial 9	How is it possible?!	Slightly stomps his right foot on the floor
Trial 10	It's absurd to not succeed in that!	Says "No... no, no" (a hint of anger in his voice)
Trial 11	And you would like to graduate?	Says "I want to leave! Tell the doctor!" (he tries to free himself by force)
Trial 12	Incapable!	Quickly gives a wrong answer. After the negative comment, he says "I feel humiliated... I want to leave... do something!" (tries to free himself by more vigorous force)
Trial 13	You are really ridiculous!	Quickly gives a wrong answer. After the negative comment, he says in a determined tone "That's enough! I want to leave... call someone!" (tries again to free himself by force)
Trial 14	You are pitiful!	Doesn't answer (he will not be answering anymore until the end). After the negative comment, he begins to sob a little
Trial 15	You are really pitiful!	Sobs and sniffles
Trial 16	Only an idiot can keep making mistakes like this!	Cries slightly
Trial 17	You are a failure!	Cries and shakes his head
Trial 18	You are the most stupid person I have ever seen!	Cries, shakes his head, then drops himself on the desk
Trial 19	You are really the most stupid person I have ever seen!	Cries, shakes his head, then drops himself on the desk

To determine their intensity, the Coach's comments and the Performer's reactions were pre-tested by 20 undergraduate students

## References

- Ancona, L., & Pareyson, R. (1968). Contributo allo studio della aggressione: La dinamica della obbedienza distruttiva [Contribution to the study of aggression: the dynamics of destructive obedience]. *Archivio di Psicologia, Neurologia e Psichiatria*, *29*, 340–372.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 1–79). Reading: Addison-Wesley.
- Bauman, Z. (1989). *Modernity and the Holocaust*. Ithaca: Cornell University Press.
- 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*, 398–413.
- Blass, T. (1999). The Milgram paradigm after 35 years: some things we now know about obedience to authority. *Journal of Applied Social Psychology*, *25*, 955–978.
- Bock, D. C., & Warren, N. C. (1972). Religious belief as a factor in obedience to destructive commands. *Review of Religious Research*, *13*, 185–191.
- Burger, J. M. (2009). Replicating Milgram: would people still obey today? *The American Psychologist*, *64*, 1–11.
- Burley, P. M., & McGuinness, J. (1977). Effects of social intelligence on the Milgram paradigm. *Psychological Reports*, *40*, 767–770.
- Caprara, G. V., Barbaranelli, C., & Borgogni, L. (1993). *BFQ, Big Five Questionnaire, manuale*. Firenze: Organizzazioni Speciali.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*, 195–202.
- Geier, A. (1998). *Heroes of the Holocaust*. New York: Berkeley.
- Kilham, W., & Mann, L. (1974). Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm. *Journal of Personality and Social Psychology*, *29*, 696–702.
- Mantell, D. M. (1971). The potential for violence in Germany. *Journal of Social Issues*, *27*, 101–112.
- Meeus, W. H. J., & Raaijmakers, Q. A. W. (1986). Administrative obedience: carrying out orders to use psychological-administrative violence. *European Journal of Social Psychology*, *16*, 311–324.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- 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.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: a meta-analytic review. *Perspectives on Psychological Science*, *4*, 301–304.
- Powers, P. C., & Geen, R. G. (1972). Effects of the behavior and the perceived arousal of a model on instrumental aggression. *Journal of Personality and Social Psychology*, *23*, 175–183.
- Quattrone, G. A. (1985). On the congruity between internal states and action. *Psychological Bulletin*, *98*, 3–40.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, *32*, 880–892.
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: an egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279–301.
- Schurz, G. (1985). Experimentelle überprüfung des zusammenhangs zwischen persönlichkeitsmerkmalen und der bereitschaft zum destruktiven gehorsam gegenüber autoritäten [Experimental examination of the relationships between personality characteristics and the readiness for destructive obedience toward authority]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *32*, 160–177.
- Schwartz, S. H. (2007). Value orientations: measurement, antecedents and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally* (pp. 167–200). London: Sage.
- Seligman, M. (2004). Commentary in Edge.org: [http://www.edge.org/3rd\\_culture/seligman04/seligman\\_index.html](http://www.edge.org/3rd_culture/seligman04/seligman_index.html).
- 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.
- Spink, J. (2000). *The righteous among nations: Analysing the heroes of the Holocaust*. Unpublished dissertation. University of Bradford.
- Zimbardo, P. G. (2008). *The Lucifer effect: Understanding how good people turn evil*. New York: Random House.