

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## The resistance experiments: Morality, authority and obedience in Stanley Milgram's account

### Journal Item

How to cite:

Kaposi, Dávid (2017). The resistance experiments: Morality, authority and obedience in Stanley Milgram's account. *Journal for the Theory of Social Behaviour*, 47(4) pp. 382–401.

For guidance on citations see [FAQs](#).

© 2017 John Wiley Sons Ltd.

Version: Accepted Manuscript

Link(s) to article on publisher's website:  
<http://dx.doi.org/doi:10.1111/jtsb.12137>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

***The resistance experiments: Morality, authority and obedience in Stanley Milgram's account***

Dr Dávid Kaposi

The Open University

Department of Social Sciences

Milton Keynes

MK7 6AA

Email: [David.Kaposi@open.ac.uk](mailto:David.Kaposi@open.ac.uk)

## ***The resistance experiments: Morality, authority and obedience in Stanley Milgram's account***

### ABSTRACT

The paper seeks to re-conceptualize Stanley Milgram's (in)famous experiments on willing obedience by drawing solely on Milgram's own contemporary account. It identifies a substantial incongruence between the findings Milgram presented (i.e., his description of the experiments) and the meaning he imputed to them (i.e., his interpretation of the experiments). It argues that instead of operationalizing the concepts he claimed to operationalize – legitimate authority, embodied morality and willing obedience –, Milgram's description suggests that the operative forces in the experiments were an illegitimate authority and acts which in effect collude with that authority. As a result, the paper concludes that what the experimental findings represented was not so much obedience out of choice, but out of coercion. Thus, the paper seeks to redirect the conceptual-moral focus of the findings from the participants who “shockingly” obeyed to those who managed to resist the *coercive* force of the total experimental situation.

Keywords: Milgram experiments, morality, authority, obedience, freedom, disobedience

“Now the lack of punishment is, to be sure, a major difference between Milgram’s lab and the jungle war or concentration camp setting. But what happened? An astonishing two thirds obeyed anyway.” (Mook, 1983, p. 386)

“Did they intervene, go to his aid, denounce the researcher, protest to higher authorities, etc? No, even their disobedience was within the framework of ‘acceptability’; they stayed in their seats, ‘in their assigned place’, politely, psychologically demurred, and they waited to be dismissed by the authority.” (Zimbardo, 1974, p. 567)

## 1. Introduction

It is exceptional in psychology to encounter references to “the glaring celebrity status” of an experiment (Miller, 2009, p. 20; cf., Blass, 2000, 2009; Milgram, Collins & Brief, 1995) or to an experiment whose “visibility [is] without precedent in the social sciences” (Miller, 1986, p. v; cf., Burger, Girgis & Manning, 2011, p. 1). In fact, the odds are short that if we ever encounter any such description, it will refer to the series of experiments Stanley Milgram conducted, examining the nature of willing obedience and its relationship to authority. What is more, in its remit this rhetoric is not restricted to psychology or the social sciences. As evidenced by an unprecedented amount of treatises and references across the disciplines, Milgram’s shocking finding that ordinary American citizens would go on and on administering painful and ultimately lethal electric shocks to a completely innocent stranger, doing so in a meaningless learning task and for no other reason than being arbitrarily instructed by an experimenter, is “part of our society’s intellectual legacy” (Ross & Nisbett, 1991, p. 55; cf., Reicher, Haslam & Miller, 2014, pp. 394-397).

This is of no surprise. It is, after all, rare that a scientist would have the bravery to explore the apparently inexplicable event which is the Holocaust, the intellectual creativity to distil the essential aspect of it (i.e., willing obedience), and the scientific rigour to translate this to the transparent confines of the laboratory. Hunches there, in the lab, became facts. Impossibilities became truths. Thus, Milgram’s experiments transcended science and science’s categorical separation from the realm of values. They became powerful demonstrations of the darkest spheres of the human condition.

This, at any rate, is the legacy of the experiments: a legacy which has been canonized in textbooks and, by and large, present both in the mind of the public and that of the discipline of psychology (cf., Griggs & Whitehead, 2015). However, there have been serious challenges in the past, and there is a new wave of contemporary critical commentary, disputing both conceptual and methodological issues around

the experiments. What remains of Milgram's own account in the whirlwind of past and present criticism? Indeed, is it still worthwhile, as this paper suggests, listening to *Milgram's own account* when the archive is now open to scrutinize firsthand what actually took place in his lab at Yale?<sup>1</sup>

## 2. Two waves of criticism

Instead of taking at face value Milgram's shocking findings of the extent of willing obedience, the first wave of criticism remained suspicious as to whether Milgram did in fact find what he claimed to have found (Mixon, 1989; Orne & Holland, 1968; Penner et al, 1973; cf., Miller, 1986, pp. 139-178). The main methodological challenge to the experiments derived from Martin Orne's general concept of experimental "demand characteristics". This claimed, quite simply, that the findings of the experiment are null and void, as participants in laboratory studies tend to follow any instructions given to them by the scientist anyway, and without necessarily believing that these would lead to *real consequences*. They thus not only enter a realm of fiction but treat the laboratory as a terrain of fiction as well: either wholly inconsequential regarding the real world or only of some inscrutable consequence for the experimenter. It is easy to see how such a criticism would render as misguided the crux of Milgram's conclusions:

"Milgram's subjects were obedient, but they were obedient in that they pushed levers. Death-camp executioners were obedient, but they were obedient in that they pushed levers knowing that in doing so they were committing murder. No real relationship may be drawn between these two cases unless it can be established that Milgram's subjects 'knew' that they were shocking the learner as they had been told. This has not been established." (Holland, 1968, p. 71, quoted in Miller, 1986, p. 146)

There have been a number of arguments marshalled from Milgram in response to this claim (cf., Milgram, 1963, p. 374; 1965a, p. 68; cf., de Vos, 2009). Most importantly, it has been pointed out that his findings were not simply that 65 or so percentage of people obeyed. Even in his first publication he enumerated a second major, and for him puzzling, finding: participants did not simply carry out the task of shocking the learner, but experienced visible and extraordinary *tension* whilst doing so (Milgram, 1963, pp. 375, 377; Milgram, 1965a, pp. 66-69; 1967, pp. 4, 7; 1974, p. 41; cf., Miller, 1986, pp. 11-12; 2009, p. 24).

Such a state of affairs resulted in two powerful conclusions. It suggested that participants did not simply push levers according to the wishes of the authority, but also believed the experimental fiction they were supposed to buy into: that they are actually hurting someone. Yet it also appeared to suggest

---

<sup>1</sup> In the case of the Milgram experiments, the dilemma of why we should read Milgram's account at all (instead of engaging, with the help of archival material, with what *really* happened) is further complicated by the fact that Milgram has recently been rather convincingly charged with either very tendentiously/idiosyncratically presenting (Gibson, 2013a, 2013b) or downright *misrepresenting* what happened in the experiments (Nicholson, 2011; Perry, 2012, 2013). Nevertheless, the argument proposed here is that the present re-reading is *legitimate* as it does not seem to significantly conflict with accounts based on archival data. And that it is *desirable* as it presents a new theoretical framework for understanding what happened in the experiments. As any further considerations of the relationship between the archival data of the experiments and their representations by Milgram in his account of the experiments would fall beyond the remit of this paper, they will have to be pursued in footnotes.

that the kind of obedience they displayed was not the “blind” obedience of automatons; they were far from the “cog in the machine” image of a thought-/heartless bureaucrat “only following orders” (Taylor, Peplau & Sears, 2003, p. 225; cf., Baumeister & Bushman, 2011, p. 4; Novick, 2000, pp. 136-137). Their tension, rather, indicated a conflict that they *embodied*, and in which they made an excruciating *decision* to follow one path or another (Miller, 1986, p. 259).

In the face of the tension *embodied* by the participants, the argument based on “demand characteristics” may indeed seem rather weak. For if participants really did not believe in the material consequences of their actions, then there really is no explanation for sweating, stuttering, biting one’s fingernails, laughing nervously and occasionally succumbing to “full-blown, uncontrollable seizures” (Milgram, 1963, p. 375; 1965a, p. 68). It is difficult to believe that participants would have purely feigned the kind of agonized responses they exhibited. As Milgram argued, the participants’ visibly disturbed behaviour seems indeed to have testified to the fact that in the experiments they underwent some genuine conflict and were viscerally torn between competing eventualities (cf., Milgram, 1965a, pp. 62, 69; 1972). They were not just automatically pushing levers devoid of any concern of what consequences their doing so might lead to.

As the introduction to this paper indicated, with no conclusive way to decide what Milgram’s participants actually believed, the impasse that the historically prominent form of methodological debate over the experiment reached meant in effect that Milgram’s claims became accepted.<sup>2</sup> Of course, the reason underlying his participants’ behaviour has still not been fully ascertained and Milgram’s *retroactive* explanation and reference to a supposed “agentic state” have always been considered the weakest points of his project.<sup>3</sup> Of course, Milgram himself was often depicted as a rather unsavoury figure, unethically abusing his participants’ trust. But his interpretation of having found disturbingly high levels of compliance with the requests of authority to carry out what in effect were criminal acts became canonical both in the literature of the discipline and the consciousness of the public. The Milgram experiments had become “shocking facts” – if in demand of proper explanation.

One might think that with the contemporary “second wave” of critical perspectives on the experiments, this state of affairs has of late become ripe for change. And, indeed, new theoretical frameworks, methodological innovation, and the more and more widely utilized access to the wealth of archival data that had been generated by Milgram has made transformative contribution to our understanding of Milgram’s studies.

Occasionally, this leads to valuable insight on the ethical issue of how Milgram misrepresented both the extent of his debriefing and the extent to which his participants remained enthusiastic about participating in the proceedings (Nicholson, 2011; Perry, 2013). Yet, of more theoretical importance, contemporary academics forcefully argue against half a century of social scientific consensus that Milgram’s shocking findings represent the concept of “obedience” in response to “orders”. This conventional (yet by now virtually untenable) explanation of Milgram’s findings in terms of

---

<sup>2</sup> The reason that subsequently neither Milgram nor his proponents overly emphasized the charged atmosphere in the lab probably derived from the fact that what was a good argument in warding off “demand characteristics” challenges was actually an embarrassment in the other historically significant debate about Milgram’s ethical conduct (cf., Baumrind, 1964, 1985; Kelman, 1967, 1973; cf., Miller, 1986, pp. 88-138).

<sup>3</sup> This is a rare point of convergence between both those sympathetic and critical to Milgram’s account in general (cf., Miller, 1986; Reicher & Haslam, 2011; Reicher, Haslam & Miller, 2014; Wrightsman, 1974).

“obedience” is either replaced by a discursive perspective that views the experiment as a rhetorical encounter and that argues that the outcome of the experiments was actually determined by a rhetorical battle of wits where both the participant and the experimenter seek to mobilize discursive resources (Gibson, 2013a, 2013b, 2014; Hollander, 2015). Or the conventional explanation is countered by the social identity theory based idea that compliance in the experiments actually arose from a type of “engaged followership” based on participants’ identification with the experimenter (Haslam, Reicher & Birney, 2014; Reicher & Haslam, 2011; Reicher, Haslam & Smith, 2012).

Thus, contrary to the “first wave”, the “second wave” of critical engagement with the obedience experiments has resulted not only in radically new ways of understanding what went on in Stanley Milgram’s lab but ways that comprehensively seem to replace the conventional interpretation. Yet, as will be argued shortly, it still seems valid to assert that in some essential aspect the reality of Milgram’s experiments and thus the legacy depicted in the introduction remain untouched. Contemporary scholarship thoroughly contests the lasting theoretical framework Milgram constructed for the experiments; yet it still appears to simply take or leave what Milgram presented as the *shocking facts he found*: a highly disturbing number of people willingly complying with experimental instructions to hurt another human being.

Whilst shedding light on the moral character of Milgram himself, contemporary moral criticisms offer no new arguments regarding his findings, rehearsing merely a mixture of established (and contradicting) points about demand characteristics (Nicholson, 2011, pp. 748-9) and the abuse of participants’ trust by “manipulating, “tricking” and “deceiving” them into obedience (Nicholson, 2011, pp. 748, 750, 755, 756; cf., Baumrind, 1964; Perry, 2012, 2013). As such, just as in the past, contemporary moral criticism of the experiments amounts not so much to a novel perspective on them. As authors uncover newer and newer layers of Milgram’s moral shortcomings, we are not offered with a reinterpretation of Milgram’s findings or a systematic programme of inquiry building on Milgram’s research, but a call for their wholesale dismissal.

What of the reconceptualization of the experiments in rhetorical terms? As mentioned above, the proposal, pioneered by Steve Gibson, to understand Milgram’s findings not in terms of the concept of obedience but of a rhetorical battle where “the experimenter [is] essentially trying to argue the case for the continuation of the experiment, and the participant [is] seeking to provide counterarguments in favour of the cessation of the experimental session” (Gibson, 2013a, p. 15) does constitute a highly productive and systematic engagement with the experiments. Yet this is done at the expense of bracketing (if not at least dismissing) essential aspects of Milgram’s findings (Gibson, 2013a, pp. 15-18; 2013b, pp. 188-192; 2017). For it was not any matter that was at stake in these particular rhetorical encounters, but a human being’s life and well-being. And it was not any outcome these rhetorical encounters resulted in, but a shockingly disturbing majority’s systematically losing this battle of wits against an experimenter’s not apparently convincing arguments – and thus criminally hurting that other human being. At the time of writing, the discursive or rhetorical accounts of the experiments do not engage with these essential elements of Milgram’s findings. As long as this remains the case, for all their value, they leave these findings uncontested with regard to these essential aspects.

As for Steve Reicher’s and Alex Haslam’s social identity theory-based reconceptualization of the experimental findings, there can certainly be no charge of either a dismissal or bracketing of such essential aspects of Milgram’s findings. Indeed, taking as their starting point the highly counter-

intuitive fact that what sounds like the only actual *order* of the authority's prods (i.e., "You must continue. You have no other choice.") *never* seems to have led to participants continuing<sup>4</sup>, their replacing the explanatory concept of "obedience" with "engaged followership" actually makes Milgram's findings *even more shocking than they were for Milgram*. Here we have no passive automatons but active followers; not lethargic or reluctant obeyers of orders but "happy" disciples to murderous ideology (Haslam, Reicher & Birney, 2014). Yet, this means that as one (i.e., the explanatory) aspect of Milgram's legacy becomes completely transformed, another one (i.e., his findings with all their inevitable moral purport) remains untouched or enhanced even. Hence Reicher and Haslam's repeated claim that Milgram's experiments constitute a "powerful phenomenon in search of a compelling explanation" (Reicher, Haslam & Miller, 2014, p. 399; cf., Jetten & Mols, 2014, pp. 590-591; Reicher & Haslam, 2011, p. 168), once again positing a gap between the "findings" (and the legacy) and the "explanation". And as the "what happened" and the "why it happened" become separated, the findings themselves continue to hold more or less the same canonical image and meaning as attributed to them by Milgram: representing the extraordinary "willingness of people to inflict serious physical harm on others", in conforming to the mere requests of an authority figure (Reicher, Haslam & Miller, 2014, p. 395; cf., Jetten & Mols, 2014, p. 587).

The present paper does not wish to dismiss or deligitimize, bracket or ignore Milgram's findings. But it does not propose either that the only way to engage with them would be to find them even more shocking. As it does not support the assumption of a gap between findings and moral character, findings and experimental interaction, findings and explanation, the present paper does not wish to treat Milgram's findings, in this sense, authoritative: to take it or leave it.

Instead, it claims that a close reading reveals a gap already *within* Milgram's very account of his findings. That is to say, the main claims advanced by this paper are, first, the existence of a mismatch between Milgram's description of his experimental procedures and the conceptual-interpretative framework he used to interpret these procedures; and second, the potential for a radical re-conceptualization of the experiments on the basis of this mismatch, where *its main finding* being that of the extraordinary "willingness of people to inflict serious physical harm on others" in the conditions Milgram created (Reicher, Haslam & Miller, 2014, p. 395) will be thoroughly contested.

Thus, in what follows, this paper will re-examine the validity of *the conceptual framework* of Milgram's experiments on the basis of *his description* of the experiments. It will, in other words, re-examine how Milgram operationalized the three crucial concepts that have since then been taken for granted to govern the meaning of the experiments: morality, authority, obedience. Were they operationalized in a valid way? Or does perhaps Milgram's *own description* suggest that he may have unwittingly operationalized some other concepts – undermining thereby the long-lasting consensus as to what findings his experiments actually brought to light?

### 3. The concept/force of embodied morality

---

<sup>4</sup> This observation is also made by Gibson (2013a) and Burger (2009).

As mentioned above, the tension experienced by Milgram's participants is of major importance as far as the reality and meaning of his findings are concerned. First, this tension authenticated the participants' conduct: whatever they believed of the reality of the situation, it was powerful enough to trigger extraordinary emotions. Second, it ruled out the possibility that participants would obey in a "blind", automatic manner. Instead, in Milgram's interpretation, they made a choice.

What was the cause of this tension? The story Milgram presents in his papers is that "he" originally conceived of it in purely abstract terms, as an outcome of the conflict between morality and authority (Milgram, 1965a, pp. 60-61; 1974, p. 22; cf., Miller, 1986, p. 171). "He" expected the "ingrained tendency" not to hurt people to be in an ultimately victorious conflict with that of obeying authority. Yet the pilot study we read about did not seem to have led to this presupposed conflict (let alone a righteous victory) as no tension whatsoever was present. In the absence of audible feedback from the learner, virtually all participants "blithely" went on to administer the most powerful shocks (Milgram, 1965a, p. 61). This alerted "Milgram" to the fact that it is not simply the concept of authority that needed to be empirically operationalized, but that of morality as well: "a *force* had to be introduced that would strengthen the subject's resistance to the experimenter's commands, and reveal individual differences in terms of a distribution of break-off points" (Milgram, 1965a, p. 67 – emphasis mine).<sup>5</sup>

The subsequent experiments were therefore taken to be not simply the battleground of presumed abstract powers, but of very concrete (that is, empirically operationalized) *demands* emitted in the actual social situation. It was the immediate conflict between the authority's instructions and the learner's protests which the participants encountered, which they needed to resolve by their conduct, and which, according to Milgram's conceptual framework, they could have easily resolved both in moral and procedural terms.

Thus, from no audible feedback Milgram went on to introduce non-verbal feedback; distant vocal feedback; distant vocal feedback with reference to the learner's heart condition and the fact the proceedings are "bothering" his heart; and, finally, proximal vocal, visual and sensual feedback (Milgram, 1963, 1965a, 1974, pp. 32-57). What needs to be seen here is not only that Milgram finally operationalized the abstract concept of morality instead of just taking its operation for granted. For, in fact, the protests of the learner not only implied the demands of morality. What they made increasingly more immediate to the participants was the viscerally touching expression of human suffering. The pull to break off the experiment did therefore not only derive from abstract moral precepts. These precepts became *embodied* in the learner's reactions, creating thereby, according to Milgram's interpretation, an immediate and progressively more and more powerful "force" (Milgram, 1965a, p. 67).

---

<sup>5</sup> I use these somewhat clumsy quotation marks for "Milgram" in this paragraph to alert the reader to the fact that we have two Milgrams to contend with: the historical Milgram who did the experiments and wrote up the papers, and the "Milgram" who is depicted as the main character in these writings. As has been touched on in footnote 1, the difference is worthwhile noting as it appears to have a methodological and even moral significance in the case of the obedience experiments (i.e., Milgram seems to have tendentiously represented or even misrepresented "Milgram"). The present juncture is a case in point. As opposed to Milgram's own story of a rather naïve "Milgram" amazed at the phenomenon of obedience (that I reconstruct in the main body of the text), Nestor Russel has repeatedly demonstrated how strategic Milgram actually was in designing the experiments to *get* the shocking level of obedience he wanted (Russel, 2011).

So what happened when the authority's demand was immediately opposed not only from an abstract moral vantage point, but by morality embodied in immediate cries of suffering?

"With numbing regularity good people were seen to knuckle under the demands of authority and perform actions that were callous and severe. [...] What is the limit of such obedience? At many points we attempted to establish a boundary. Cries from the victim were inserted; not good enough. The victim pleaded that he be let free [...] subjects continued to shock him. [...] At the outset we had not conceived that such drastic procedures would be needed to generate disobedience, and each step was added only as the ineffectiveness of the earlier techniques became clear." (Milgram, 1965a, p. 74)

"To our consternation, even the strongest protests from the victim did not prevent all subjects from administering the harshest punishment ordered by the experimenter [...]" (Milgram, 1965a, p. 61)

What happened? *Nothing* of major significance changed: a shockingly high level of obedience remained. It made virtually no difference whether the learner pounded on the wall, cried out from behind the wall, or referred to his heart being "bothered". Though obedience in absolute terms decreased in the conditions where participant and learner sat in the same room, given the highly drastic departure from the original procedures in this condition (i.e., with participants either *witnessing* the learner's agony, or directly and actively *forcing* it to happen), the percentage of obedient people can only be perceived in relative terms as still shockingly high (cf., Milgram, 1974, p. 35). The forces were both present and they did now both exert themselves: participants began to feel mental agony. Yet far too many of them still went on and on, rather than, as Milgram depicts himself to have expected, simply "breaking off" and putting an end to both the learner's and their own misery.

"Somehow, the subject [became] implicated in a situation from which he [could] not disengage himself" (Milgram, 1965a, p. 72). "Somehow", as Milgram further theorized, s/he did not realize that subjective feelings are no substitute for real action. "Somehow", s/he still subscribed to the authority's definition of the situation. "Somehow", the powerful counterbalance did still not prove powerful enough to counter the force embodied by the present "experimenter" (Milgram, 1963, pp. 376-377; 1974).

This state of affairs never ceased to unsettle or even outrage the "Milgram" we encounter in Stanley Milgram's writings<sup>6</sup> (cf., 1965a, p. 61, 1974, p. 22; Erickson, 1968), just as it never ceased to unsettle generations who did not simply wish to cleverly explain it away or cynically shrug their shoulders. This is not surprising: in Milgram's description, it is just as morally shocking as it is intellectually inexplicable why the force of moral principles embodied in visceral suffering would not in the participant's mind defeat easily that of the experimenter's requests in a mundane learning experiment (Milgram, 1974, pp. 27-31). Why would the experimenter's simple say-so prove to be more powerful than a combination of the principle of "Thou shall not kill" and the anguished cry of "Experimenter, get me out of here!?" Indeed, the very fact that no explanation beyond Milgram's "somehow" (a non-explanation, that is) ever appeared remotely convincing to match the happenings in the laboratory only contributes to the *power of the findings*. For what happened in Milgram's lab, according to

---

<sup>6</sup> For the clumsy phrasing, see footnote 5.

Milgram's conceptual framework, is simply morally and intellectually inexplicable. Or, rather, it is only conceivable as an extraordinarily enormous indictment against humanity – *precisely in the name of humanity*.

Yet, this is exactly why the question needs to be asked whether Milgram's conceptual framework actually became validly operationalized in the procedures he deployed. Indeed, the counter-intuitive fact was not simply that participants pushed levers, but that they did so as interpreted by Milgram's conceptual framework. As the *meaning* of the experiment derives from this conceptual framework, it is a crucial question to ask how Milgram put into practice the principles of authority and morality/suffering in the form of the experimenter's and the learner's conduct.

The force of embodied morality became manifest in “the form of protest from the victim” (Milgram, 1965a, p. 61). As Milgram explains, “[i]nitially, mild protests were used, but proved inadequate. Subsequently, more vehement protests were inserted into the experimental procedure” (Milgram, 1965a, p. 61).

What exact form did these protests take? As we read,

“[...] at 150 volts the victim cries out, ‘Experimenter get me out of here! I won’t be in the experiment any more! I refuse to go on!’ Cries of this type continue with generally rising intensity, so that at 180 volts the victim cries out, ‘I can’t stand the pain’, and by 270 volts his response to the shock is definitely an agonized scream. Throughout, he insists that he be let out of the experiment. At 300 volts the victim shouts in desperation that he will no longer provide answers to the memory test; and at 315 volts, after a violent scream, he reaffirms with vehemence that he is no longer a participant. From this point on, he provides no answers, but shrieks in agony whenever a shock is administered; this continues through 450 volts. ” (Milgram, 1965a, p. 62)

In both this and other accounts of the victim's reactions, the grammatical format is without exception that of *iterative present*. The repeated use of plural nouns (“cries”, “screams”, “protests”<sup>7</sup>), temporal adverbs (“continuous”<sup>8</sup>, “throughout”<sup>9</sup>), modal adverbs (“insistent”, “vehement”<sup>10</sup>) and iterative verbs (“insists”<sup>11</sup>) all imply that the protests were intense and ongoing. In this way, Milgram's account is certainly in line with his interpretation that it was indeed morality embodied in the learner's suffering that was operationalized by his design, and that this *force* was theoretically capable of compelling the participants' conscience and helping them to disobey.

But let us now compare this account with an extract of the actual transcript of the proceedings that *Milgram himself* had provided his readers with. It is, to all present intents and purposes, typical of the experiments (cf., Milgram, 1973, 1974, pp. 44-54, 73-88).<sup>12</sup> In what follows,

---

<sup>7</sup> cf., Milgram, 1965a, p. 61, 62, 74; Milgram, 1965b, p. 128; cf., Burger, 2009, p. 1.

<sup>8</sup> cf., Blass, 2009, p. 42.

<sup>9</sup> cf., Milgram, 1965a, p. 62

<sup>10</sup> cf., Milgram, 1965a, p. 60; 1965b, p. 128; 1967, p. 5.

<sup>11</sup> cf., Milgram, 1965a, pp. 60, 62; 1965b, p. 131.

<sup>12</sup> As has been demonstrated, in general Milgram's transcripts tended *not* to be accurate, if measured by any reasonable (academic or vernacular) standard (Gibson, 2013b; 2017). However, the elements of the transcript

participant Fred Prozi<sup>13</sup> engages the experimenter after administering the 180-volt shock followed by the learner shouting “Ugh! I can’t stand the pain. Let me out of here!” (Milgram, 1974, p. 56)

Prozi: He can’t stand it. I’m not going to kill that man in there. You hear him hollering?  
 Experimenter: As I told you before, the shocks may be painful but –

that this particular analysis focuses on are replicated by all academics who transcribed the experimental sessions (e.g., Modigliani & Rochat, 1995, p. 119). For comparison, see the Jeffersonian transcript of the whole passage:

01 P: [(['Shocks' L])]  
 02 L: [ ↑AH! ] A:W:.  
 03 (2.3)  
 04 L: I CAN'T STAND THE PAIN.=LET ME=[OUT OF HERE!.  
 05 P: [He can't stand=it now=  
 06 [ ((turns to E))  
 07 =I'm not gonna kill that man there;  
 08 (0.3)  
 09 P: [Y'here him hollerin' ( [ )]  
 10 [((Hand gestures to L))  
 11 E: [As=I said before the shocks  
 12 may be <painful b[ut they're not dangerous.  
 13 P: [ Yeah but =[he's in there=  
 14 [((Glance to L,[hand gesture to L))  
 15 =hollerin' he can't ↑stand it what if some'n  
 16 ↑happens to=him.  
 17 (1.1)  
 18 E: The experiment requires that you continue teacher.=  
 19 P: =#Y:eah# [but ah:: £<Huh£ (0.3) I'm not >gonna get<=  
 20 [((turns away, shaking head))  
 21 =[that man sick in £the(h)re?£  
 22 [((Glance to L, then back to E))  
 23 (0.9)  
 24 P: [#>I=mean<=he's ↑hollerin'# there.  
 25 [((Hand gesture to L))  
 26 (1.2)  
 27 P: Y'know >whadda< mean? ah: I mean huh huh ( [ )  
 28 E: [Whether  
 29 th'=learner likes it or not we m:ust go on until  
 30 [he's learnt all the word pairs.  
 31 P: [ #I don't I mean)# ah:=-I'd refuse to take the=  
 32 [((Shakes head, gaze from E))  
 33 =responsibility[>of=him< getting' hurt in there.  
 34 [((Gaze to E))  
 35 (1.0)  
 36 P: [(Y'know) I'm not[gon- I mean [he's in there hollerin'.  
 37 [ ((Shakes head [glance to L [gesture to L))  
 38 (1.1)

Prozi 'P', Learner 'L', Experimenter 'E'. Note that “glance to L” or “gesture to L” transcription comments refer to L’s (non-visible) location as oriented to by Prozi.

<sup>13</sup> By virtue of featuring in Milgram’s film of the experiments, Prozi was a highly important participant in the experiments (Millard, 2014).

Prozi: But he's hollering. He can't stand it. What's going to happen to him?  
Experimenter: (*his voice is patient, matter-of-fact*): The experiment requires that you continue, Teacher.  
Prozi: Aah, but, unh, I'm not going to get that man sick in there ... know what I mean?  
Experimenter: Whether the learner likes it or not, we must go on, through all the word pairs.  
Prozi: I refuse to take responsibility. He's in there hollering!  
[After an exchange regarding responsibility, Prozi continues.]

(Milgram, 1974, pp. 73-74; cf., Milgram, 1973, p. 63)

We see the famous "prods" in operation here. The authority is undeniably present. We also sense the manifest and intense discomfort of the participant, worrying about "kill[ing] that man in there". What we *do not* "see", however, is the victim. There are no "cries", no "insistence", certainly no "vehement protests throughout" (cf., Milgram, 1965a, p. 62). (Milgram did represent these in his transcripts when they occurred.) There is only silence from the learner, following his one, standardized, utterance.<sup>14</sup>

As such, we can detect a significant gap between Milgram's verbatim account of the experimental procedures and his interpretation of them. This means, of course, that if we accept Milgram's summary of the victim's behaviour above as a valid description of how morality/suffering should be operationalized (cf., Milgram, 1965a, p. 62), then we also have to conclude that, on the evidence of Milgram's own rendering of the experimental procedures (cf., Milgram, 1974, pp. 75-76), morality/suffering was never operationalized in his design in a valid way.

What does this ambiguous operationalization mean? For one thing, and *pace* "demand characteristics" argument, it certainly does not mean that the force which the victim embodied was meaningless and that, consequently, participants could categorically have not believed in the experimental fiction they were supposed to believe. Milgram's verbatim transcript certainly depicts a participant who is struggling heavily to come to terms with the experimental fiction. Fred Prozi confesses his distress over the prospect of "kill[ing] that man in there". For another thing, though, ambiguous operationalization of morality also appears to have a bearing on the experiment which Milgram's interpretation never countenanced. To understand this, we have to look at the exchange above as a piece of interaction; a dispute, as it were, as to whether to continue the experiment or not (cf., Billig, 1996; Edwards, 1998; Gibson, 2013a, 2014).

It is the victim's "cry" that initiates or triggers this dispute. Without this, no argument would ensue as no moral dilemma would be present. It is the victim that occasions the perspective of his own well-being and juxtaposes it with the progression of the experiment. Fred Prozi duly joins the debate on his (i.e., the victim's) side, and he seeks to advance this perspective by

---

<sup>14</sup> One might claim that the victim was simply blocked by the wall to do any of this. However, as Milgram made it clear, "his [i.e., the victim's] complaints could be heard clearly through a door left slightly ajar, and through the walls of the laboratory" (Milgram, 1965a, p. 62).

making frequent references to the victim's protests and well-being. Using the present continuous tense, he refers to the victim's state on three occasions: "hear him hollering?", "he's hollering", "he's in there hollering".

Except, of course, that, as his responses are standardized by experimental protocols, by this point the victim is audibly *not* hollering.

This is not to say that this standardized silence is without meaning in the emerging interactional context jointly created by the victim's utterance ("Ugh! I can't stand the pain. Let me out of here!") and Prozi's uptake of it. For Prozi's utterances are not simply statements about the world, but seek to perform two things in the world (Potter, 2001). Explicitly, they form a moral argument; and implicitly, they form a request that the victim back up or enjoin this moral argument – an argument *on behalf of and initiated by the victim*. The victim's standardized silence is, therefore, a *response* to these acts of Prozi. For this reason, this silence unequivocally signals, first, the victim's active *rejection* to enjoin Prozi's argument; and second, the victim's active *refutation* of this argument of Prozi's about his "hollering". Despite his three clear attempts to engage the learner/victim and demonstrate the learner/victim's suffering, therefore, Prozi's moral argument is three times rejected and refuted. Given that this very moral argument was initiated by and is pursued on behalf of the person who is now rejecting and refuting it by being silent, not only has Prozi lost the dispute and not only has he been let down but he might also be said to have become utterly confused and humiliated.

Milgram's verbatim description therefore appears to firmly contradict Milgram's interpretative framework regarding the concept/force of morality. Far from being an unequivocal operationalization of morality supposedly clashing with authority, the victim's conduct embodied something very different. It was not simply the authority's conduct that kept the participants obeying. Counter-intuitively and bewilderingly, the victim too in part rejected and refuted their arguments to disobey; in fact, the victim implicitly *colluded with the authority* just when he appeared to explicitly collide with him.

#### 4. The concept/force of the legitimate authority of science

Having examined one aspect of Milgram's design, one concept and its operationalization that enabled the participants within Milgram's design to make a choice, we now have to turn to the other one: without any ambiguity, participants were understood to obey a *legitimate authority* whose legitimacy derived from the enterprise of *science*. It was this that was taken to constitute the second of the "forces" in the situation (Milgram, 1965a, p. 67; 1974, pp. 6, 143; cf., Fromm, 1973, p. 74).

So, was the legitimate authority of science operationalized in a valid way? There is an old argument, of course, to the contrary. The "demand characteristics" stance did not simply oppose the validity of Milgram's experiment on the grounds of principle. Proponents typically marshalled another argument of their counter-thesis as they examined the behaviour of the experimenter within the settings.

"When the pounding on the wall occurred, followed by silence, it seemed clear at the very least that something had gone wrong with the experiment. Yet the 'experimenter' gave no indication, either verbally or nonverbally, that he was

surprised or in the least bit concerned about what happened in the other room. [...] the experimenter might at least be a bit expected to show some concern for himself. [...] The only motivation that would explain the ‘experimenter’s’ composure and lack of interest in what had happened in the other room is certain knowledge on his part that the subject is not harmed.” (Mixon, 1989, p. 28; cf., Mixon, 1972, 1977)

Such a state of affairs was originally termed “incongruous” by Orne and Holland, and it highlighted a fact that was supposed to empirically support their principled incredulity concerning the participants’ credulity (Holland, 1968; Orne & Holland, 1968). The argument ran that it was the very conduct of the experimenter that invalidated the reality of events within the laboratory. No real scientist (i.e., the legitimate authority that the participants allegedly obeyed) would behave in the way Milgram’s actual experimenter did. Therefore, no participant should be considered to have believed what they were asked to believe: that they were actually hurting anyone.

Leaving momentarily aside the issue that the participants’ tension and visible emotional discomfort are left unaccounted for by this argument, it is difficult to see any fault with it – *provided*, that is, we accept Milgram’s claim that what he operationalized and exclusively operationalized was the concept of legitimate scientific authority. Indeed, scientists do not arbitrarily and meaninglessly kill and torture people; and if they accidentally start doing so, the expectation is reasonable that they would behave differently from the way Milgram’s experimenter did.

Have we then reached an empirical impasse concerning the credibility of Milgram’s study, where participants just cannot possibly take the situation seriously – but where they still somehow feel highly stressed? Where either their gullibility or their genuine tension is left unaccounted for?

As indicated above, this is only the case *if* we take for granted Milgram’s claim that what his experiments exclusively operationalized was the impact of the legitimate and benevolent authority of science. Importantly, the “incongruity” argument suggests that the lack of *disapproval* on the part of the experimenter regarding the terrible things going on in the other room means a lack of *recognition and validation* of those terrible things. This, and only this, is why the hypothetical conclusion ensues that nothing really was going on in there.

However, it must be clear that what characterized Milgram’s authority was by no means a “lack of interest in what had happened” (Mixon, 1989, p. 24). Though he most certainly did not communicate his “surprise” or “concern”, he *did communicate his interest*; though, of course, his interest did not pertain to the discontinuation of the experiment but to its very continuation (cf., Gibson, 2013a). Thus, *both* Mixon’s and Milgram’s interpretative arguments ignore Milgram’s actual description of the experiment where the absence of disapproval and the presence of the demand to continue actually communicate *approval*: that there is really something terrible going on in the other room, and that this is exactly what should be going on in there.

Yet if so, this state of affairs does not just refute the “incongruity” argument but, unexpectedly perhaps and against the intentions of Mixon, also raises the dilemma of the valid operationalization of the second crucial concept – the legitimate authority of science – of Milgram’s interpretative framework. For if the experimenter is not simply a part but actually the *originator* of the destructive proceedings,

then by definition he turns from being the legitimate authority of science into a criminally destructive (and, arguably, an illegitimate) authority.<sup>15</sup>

Thus, just as there seemed to be a gap between Milgram's interpretative framework and his descriptive account of the experiment in terms of his first concept/"force" (i.e., embodied morality), there seems to be a similar gap between his claim to have operationalized the legitimate authority of science and his description of what he actually operationalized. But, equally, just as with the case of the first "force", it also seems possible that Milgram's account of the experiments did not describe some meaningless procedure. In other words, just as his account suggests to have operationalized something meaningful in addition to or in the place of embodied morality, so it appears that, at the crucial points of disagreement, the authority figure of the experiment did not represent or only represent the legitimate authority of science, but rather something else: an authority that seemed to have both the wherewithal (i.e., the shock machine) and the clear intention (i.e., prods) to deliver painful and possibly lethal shocks. Therefore, what in the experiments Milgram ultimately operationalized with some validity might have been an *arbitrarily destructive and criminal (and, hence, illegitimate) authority*.

If this is acceptable, however, it becomes our task now to understand what, according to his own description of the experimental proceedings, Milgram's *very findings* were. For if neither of the "forces" manipulated by Milgram stood for the concepts that they were supposed to represent, and if the experiments actually seem to have operationalized alternative and meaningful concepts that Milgram's interpretative framework did not account for, then this state of affairs leads us to ask the question: what, then, does Milgram's description suggest as regards his "dependent" concept – in other words, his very findings?<sup>16</sup> Did he, as he claimed in his interpretation, measure the concept of willing obedience in his lab? Or, as was the case with the "forces" of morality and authority, did he possibly measure something else?

## 5. The concept of willing obedience

In understanding the concept of obedience and rendering the empirical findings meaningful, it was a crucial element of Milgram's conceptual framework from the very beginning that his participants were conceived of being active and responsible agents in the experiment. Implying this, he variously described them using an active language: as people who did not "break off" (Milgram, 1965a, p. 66), who "go [very far] in complying", who "gives himself over" (Milgram, 1974, p. xx), who "continue" and, of course, who do "willingly" (Milgram, 1967, p. 5) whatever they do. What Milgram's rhetoric repeatedly suggests is that the behaviour of the participants derives from their *choice*. To be sure, that particular choice occurs in a peculiar social situation, responding to peculiar experimental

---

<sup>15</sup> Indeed, Milgram regularly referred to the gradual evolution of the situation in terms of destruction. Mostly, he talked of "destructive obedience" (Milgram, 1964, p. 848; cf., Miller, 1986, pp. 138-140; Miller, 2009, p. 21) or "destructive process" (Milgram, 1967, p. 5; 1974, p. 6). On occasion, even the "commands" were depicted as "destructive" (Milgram, 1965a, p. 68) and the experimenter as displaying "extraordinary harshness [...] toward a victim" (Milgram, 1964, p. 851). Milgram's account, therefore, attributed a quality of arbitrary and criminal destructivity to every element of the evolving experimental procedures – bar, of course, the "experimenter" himself.

<sup>16</sup> Milgram did not use the terms "independent variable" or "dependent variable", using instead figurative language such as "force". He identified the levels of the concept of "obedience" by what he called "dependent measures" (Milgram, 1964, 1965a).

characteristics; and as such, it might not be all that easy to make such a choice. But it is in essence both a morally necessary and an empirically free choice nonetheless, given that, first, the moral alternative to disobedience was hurting and possibly killing another human being, and second, according to Milgram's framework, disobedience was procedurally easy to accomplish by simply discontinuing the experiment. Yet, in Milgram's interpretation, *despite the presence* of clearly uncomfortable emotions, a disturbing proportion of participants chose not to exercise it and just obeyed instead (Milgram, 1963, p. 376, 1965a, p. 67, 1965b, p. 128, 1974, p. 4; Miller, 1986, p. 12.; cf., Milgram, 1963, p. 377, 1965a, p. 72; 1965b, p. 130, 1967, p. 7, 1974, p. 41).

What had been implied in the early publications first became more explicitly expressed in Milgram's reply to Diana Baumrind, and then elaborated on in his book on the experiments.

Foreshadowing subsequent moral critiques, Baumrind famously disparaged Milgram for treating his participants with disrespect, and exposing them to a situation which may have resulted in long-term psychological damage (cf., Nicholson, 2011; Perry, 2013). What is more, Baumrind contended that such a state of affairs was even more problematic as it was the experimental situation created by Milgram that "*made*" (Milgram, 1964, p. 851) people behave in the way they did:

"The dependent, obedient attitude assumed by most subjects in the experimental setting is appropriate to that situation. The 'game' is defined by the experimenter and he makes the rules. By volunteering, the subject agrees implicitly to assume a posture of trust and obedience." (Baumrind, 1964, p. 421)

Baumrind's criticism was clearly different from the "demand characteristics" position. Whilst pointing out the peculiar characteristics of the experimental situation, her point was not that the laboratory allows people to behave in an inconsequential manner. Rather, it was that the situation *as experimental situation* may in itself be accountable for the acts the participants found themselves carrying out and the subsequent psychological torture they had to endure whilst doing so. In a state of suggestibility and prone to obey anyway, they were presented with a situation they could not extricate themselves from: mesmerized, as it were, by science. In Baumrind's interpretation, it is this natural tendency for experimental obedience and trust in the experimenter that Milgram abused.

To this, Milgram's riposte was that

"[a] concern with human dignity is based on a respect for a man's potential to act morally. Baumrind feels that the experimenter *made* the subject shock the victim. This conception is alien to my view. The experimenter tells the subject to do something. But between the command and the outcome there is a paramount force, the acting person who may obey or disobey. I started with the belief that every person who came to the laboratory was free to accept or reject the dictates of authority. This view sustains a conception of human dignity insofar as it sees in each man a capacity for *choosing* his own behaviour" (Milgram, 1964, p. 851)

It was at this point, then, that the third crucial feature of the conceptual framework of the experiment became explicit. Participants in Milgram's interpretation did not merely act but *chose freely* to act. They could always have chosen to disengage. They could always have chosen to defy the legitimate authority of the experimenter.

Indeed, by the time Milgram published his book summarizing, interpreting and explaining his main findings, the participants' capacity to make free choices was repeatedly highlighted:

“While people will comply with a source of social control under coercion (as when a gun is aimed at them), the nature of obedience under such circumstances is limited to direct surveillance. When the gunman leaves, or when his capacity for sanctions is eliminated, obedience stops. In the case of voluntary obedience to a legitimate authority, the principal sanctions for disobedience come from within the person. They are not dependent upon coercion, but stem from the individual's sense of commitment to his role. In this sense, *there is an internalized basis for his obedience, not merely an external one*” (Milgram, 1974, p. 141)

“To the degree that an attitude of willingness and the absence of compulsion is present, obedience is colored by a cooperative mood; to the degree that the threat or punishment against the person is intimated, obedience is compelled by fear. Our studies deal only with obedience that is willingly assumed in the absence of threat of any sort, obedience that is maintained through the simple assertion by authority that it has the right to exercise control over the person” (Milgram, 1974, p. xx-xi).

In his interpretation, then, Milgram's interest lay in obedience stemming from free choice and not from coercion or threat. It is easy to see why the capacity to act freely and out of one's own volition is a conceptual cornerstone of the findings of the experiments. For insofar as human action is carried out in the shadow of (physical) threat, and insofar as obedience stems from coercion rather than a choice, responsibility may rest partly or wholly with the agents of coercion rather than their instrument. Our interpretations of these acts of obedience may therefore shift as we will be inclined to see the scene not as an experiment but akin to torture.

So what was the foundation of Milgram's “belief” (1964, p. 851) that his participants had the empirical (i.e., not the metaphysical) capacity to *choose* freely their behaviour? In other words, how did he include the concept of free choice in his research design? How did he make sure that, as his participants executed the harshest of acts against an innocent fellow human being, they simply followed “the instructions of an authority who had no special powers to enforce his commands” and that they realized that “[t]o disobey would bring no material loss to the subject; no punishment would ensue” (Milgram 1963, p. 376; cf., Milgram, 1974, p. 41; 1983; Morelli, 1983)?

These may sound rather nebulous questions. After all, in reality the participants freely volunteered for the experiment and could have left it just as freely. In reality, the experimenter *really* did not have any physical “power to enforce his imperatives” (Milgram, 1967, p. 4). Yet, by the same token, in this experiment there *really* was no reason to leave either, as no one *really* was hurt at all. And if Milgram's interpretative framework readily continued to assume that his participants subscribed to the fiction proposed to them (and thus discounted the reality), there is no *a priori* reason to assume that they just as readily assumed certain other aspects of reality (i.e., the lack of punishment and their right to leave just as and when they wished to). Indeed, inasmuch as participants are supposed to have subscribed to no other fiction than that of lethal and uncompromising destruction in the lab, there is *absolutely no reason* to suppose that they still “somehow” knew that it was them rather than the experimenter who had the power to change the scene of destruction.

Fiction did not merely act in parallel with reality in the experiment. Milgram's description of his design and procedures suggests that the experimental fiction of arbitrary destruction erased the reality of free choice.

What participants voluntarily agreed to in the beginning was, of course, an innocuous "learning experiment". However, on the basis of Milgram's description of the experimental procedures, as Milgram's experimental fiction started to saturate the social situation with an arbitrarily destructive and downright criminal authority starting to emerge in the place of (or in addition to) a scientific one, we have to ask the question that Milgram never asked: to what extent did this shift from constructive benevolence to meaningless and criminal destructive malevolence impact on the very findings of the experiment, and to what extent did it amount to not only credible if very uncomfortable commands, but also to *conceivably very real threats*? In other words, can we still simply assume that participants continued to be aware that they were still voluntary participants and that they had the free choice to discontinue at their disposal? Do we have any basis for assuming from Milgram's description of the experiments that participants may have been in a situation where they really had *no choice but to continue*; where freedom of choice and responsibility were not so much cognitively or rhetorically disclaimed later but, to all intents and purposes, *could have been legitimately perceived as not being present*?

Needless to say, we cannot give an unequivocal answer to this most crucial of questions.<sup>17</sup> But, as mentioned in the previous section, Milgram's design depicts an actual "gun" in the laboratory in the form of the shock generator. It depicts manifest physical suffering and the tangible spectre of death. It also depicts the potential inter-changeability between the roles of the learner and the teacher as these roles were merely decided by an arbitrary draw. Thus, it depicts with some validity an increasingly malevolent, meaningless and arbitrarily destructive authority, suggesting and then specifying the categorical nature of his commands and never at this point alluding to any lack of consequences. For this reason, just as in the realm of reality it was unambiguous that there was no punishment involved in the proceedings, it may have been equally or even more plausible in the realm of experimental fiction that there actually was. And if this is so, Milgram's interpretation/assumption that participants acted freely and willingly becomes problematic; and correspondingly, the counter-assumption that their obedient response was not an outcome of choice but one of coercion in the face of torture becomes distinctly possible, on no other basis than Milgram's own description of the procedures.

To state the alternatives clearly. Milgram's interpretation suggests that his participants may have fully believed the fiction of some terrible or even lethal things going on in the laboratory. And they may still have also believed the reality that it is only their dispositions that make these terrible things happen. They may have believed the experimenter telling them that they have "no choice" yet may have paradoxically still known that they indeed had a choice. As a result, they may have not been able to

---

<sup>17</sup> As there are claims being made here about what Milgram's participants' might have felt or known or realized (that is to say, what they might have *experienced*), it is tempting to want to settle these crucial issues by going back either to the archive (and Milgram's post-experimental surveys and interviews) or directly to the participants (e.g., Haslam et al, 2014; Nicholson, 2011; Perry, 2012). However, the present inquiry would caution against in settling political-moral dilemmas with recourse to a simple, direct and exclusive reliance on participants' capacity to report of their state of mind. What participants report(ed) to have experienced (*horribile dictum*, what participants consciously experienced) cannot simply settle issues of such conceptual, epistemological and moral complexity that we are dealing with.

extricate themselves from the situation *despite* the fact that they felt it torturing, because a legitimate authority commanded them to do so. And they may have entered the “agentic” state to effectively deny the responsibility that was ultimately solely theirs. (Or they may just have claimed to have done this.)<sup>18</sup>

Yet the present re-analysis of Milgram’s description of his design and procedures suggests some other possibility. It suggests that participants may have believed (or feared) enough of the fiction of some terrible or even lethal things going on in the laboratory. And they may have also believed, *accordingly*, that the reality of the context of the experimenter, laboratory, science and free participation they volunteered to take part in turned out to be an illusion. They may have believed that the experimenter telling them that they have “no choice” was acting truthfully, as he was the authority who would also torture or even kill some innocent in the other room. As a result, they may have not been able to extricate themselves from the situation *because* of the fact that they felt it torturing and never knew who would be tortured next if they refused to go on. And they may therefore *really* have been in a situation where any responsibility for what happened did not lie with them.

Due to the fact that not only did Milgram only ever assume and never actively incorporated the concept of free choice into his research design, but made ample efforts to suggest that it was not present, we will never know for sure which of these scenarios is closer to the truth. Yet, based on the present close reading of *Milgram’s description*, we can at the very least punctuate over half a century of consensus regarding the findings of the Milgram experiments. Namely, we can see now obedient acts not simply stemming from free choice (or at any rate the illusion of free choice) but also, to some considerable extent, from coercion. And, as a corollary, we can finally see the first act of disobedience in the experiment not as the naturally and normatively expected reaction, but something truly extraordinary – both in terms of “the starry sky above [...] and the moral law within” (Kant, 1997, p. 133).

## 5. Conclusion

The present paper’s aims were twofold. Re-analysing Milgram’s description of his design and procedures, it located a mismatch between Milgram’s description of his findings and the conceptual-interpretative framework he constructed to understand them; and it proposed a fundamental reconceptualization of the experiments on the basis of this mismatch. It suggested that, *by his own description*, Milgram did not (or, at any rate, did not exclusively) operationalize in a valid way the concepts he claimed to have underpinned his experimental findings. Instead (or, at any rate, in addition) he operationalized concepts he took no account of when reporting his findings of high levels of “willing obedience”, but which are crucial for the interpretation of his experimental findings.

First, the concept or “force” of morality, embodied in the victim’s suffering, was not (or, at any rate, not exclusively) what Milgram operationalized by the experimental procedure. Despite what is

---

<sup>18</sup> Of course, most of the attributions mentioned in this paragraph sound untenable in light of the contemporary re-evaluations of the experiments introduced here earlier. Yet the point here is that they are just as untenable on the basis of the re-analysis of Milgram’s own description of his experiments – and that the alternative attributions presented in the next paragraph amount to a fundamental re-interpretation of the experiments.

explicitly stated by Milgram's interpretative framework, his description of the proceedings clearly suggests that the learner did by no reasonable standard protest continuously and despite the occasional outbursts never engaged in argumentations about his own existential or moral status. What this meant, however, as the incongruous nature of the learner's behaviour may actually have undermined the very validity of his outbursts and protests, was not simply that the "force" of morality was far less powerful or persuasive as Milgram claimed. Rather, his design also undermined the participants' attempts to argue their way out of the conflict by engaging these outbursts and protests. As such, it not only represented a less powerful "force" to help the participant disobey than Milgram claimed, but, arguably, one that actually *colluded* with the authority to ensure the participant's obedience.

Second, the concept or "force" of the legitimate authority of science was not (or, at any rate, not exclusively) what Milgram operationalized at the last and crucial phase of the experiment. Such an authority would of course not countenance the scene of arbitrary destruction that had eventually emerged in the lab with the victim crying out for help. Milgram's authority, however, not only tolerated the ghastly scene but brought about its existence with his requests/orders, and with having both the means (i.e., shock generator) and the apparent intention (i.e., the arbitrarily hurting of an innocent human being) to enforce these requests/orders. As such, he not only ceased to be the legitimate authority of science but, arguably, became the illegitimate and criminal authority of arbitrary destruction.

Third, although the concept of obedience was at all points based on the capacity of participants to make free choices (i.e., to willingly obey as opposed to be coerced), Milgram's description indicates no attempt whatsoever to incorporate it in the design of his experiment. This is not to say that he should not have assumed or "believed" in human beings' capacity to choose freely (Milgram, 1964, p. 851). Rather, it is to point out that Milgram's belief here was absolutist in that his stated assumption concerning the existence of free choice seems to have equalled his unstated assumption of the absolute power of free choice over the situation. As such, he did not (could not) explore the conditions within which free choice would have been exercised in his experiments, free choice's interaction with those conditions, and, ultimately, the extent to which free choice was possible within those conditions. What is more, these do not seem to be mere academic exercises of thought as the experimental presence of an arbitrary and criminally destructive authority figure alongside an incongruous "victim" implicitly colluding with this figure may reasonably be seen to have radically undermined Milgram's implicit assumption regarding the absolute power of free will. Inasmuch as participants subscribed to this fiction of the authority and the "victim", they had every reason to also believe that the possibility of making free decisions was not any more at their disposal. In other words (and as evidenced both by visible signs of physical discomfort during and explicit disclaimers of responsibility after the experiments), that any decision to disengage may result in some destructive consequences regarding their own wellbeing.

As Milgram's interpretation of his findings was at all points based on these three concepts of embodied morality, legitimate authority of science and willing obedience, this paper has to conclude that *Milgram's interpretative claims* that he found a shocking number of people to willingly follow the instructions of an experimenter in causing lethal damage to a fellow human being are, *according to Milgram's own description of his experimental proceedings*, unsubstantiated. At the same time, this paper also concludes that Milgram's own description does suggest the distinct possibility that the

experiments actually resulted in some, as yet unreported but very meaningful, alternative finding. This was that despite being psychologically tortured (by the potentiality of physical torture), a strikingly high percentage of Milgram's participants managed to extricate themselves from the situation. *That is to say*, far from being simply an important element to "complete" the picture (Jetten & Mols, 2014; Reicher, Haslam & Miller, 2014, p. 398), *resistance in the face of the perception of a criminally destructive authority* (embodied both by the experimenter and the "victim" colluding with him) *may have been the very conceptual core of the Milgram experiments.*

To conclude here: as regards the actual experiments, as opposed to Milgram's construction of them, this statement above is certainly but a hypothesis. It needs to be further tested, with regard now not any more to Milgram's interpretive words but the archival data registering the events in his lab. Ultimately, it is there that we can see whether the conceptual framework this paper has been arguing for is well suited to help us understand the dynamics of, on the one hand, the workings of extreme domination and, on the other, the realisation of the capacity to resist such coercion. Dynamics, that is, which both Milgram and his participants exhibited whilst clearly being unaware of them.<sup>19</sup>

List of references:

Baumeister, R.F. & Bushman, B.J. (2011). *Social psychology and human nature*. 2<sup>nd</sup> ed. London, UK: Wadsworth Publ.

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

Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited. *American Psychologist*, 40, 165-174.

---

<sup>19</sup> This paper has so far persistently refrained from engaging with the crucial issue of disparity between the account Milgram gave of the experiments and the accounts contemporary scholars on the basis of archival material reconstruct for the experiments. Yet at this point one glaring point of contradiction between the present reconstruction and what archival data seem to suggest must be reflected on. Namely, the very conclusion that Milgram's major finding in the experiments was not the social psychological phenomenon of "willing obedience" but that of "acting or resisting under conditions of coercion/torture" appears to conflict head on with what is arguably the most robust as well as intriguing finding emerging from the archive. As remarked in the Introduction, Milgram's seemingly most categorical, most threatening, and most order-like "prod" (i.e., "You must continue. You have no other choice.") *never* actually worked (Burger et al, 2011; Gibson, 2013a, 2013b; cf., Haslam, Reicher & Birney, 2014; Reicher & Haslam, 2011; Reicher, Haslam & Miller, 2014). How is it that the ultimate expression of forceful destruction would *not* have led to obedience? The answer is that utterances are indexical, i.e., they mean one thing in one context and something else in another (Billig, 1996). Thus, the final "prod" may indeed be an instance of a categorical command or even coercion in certain contexts. Yet, *those participants* who managed to reach the phase when this command was uttered had already created a context of negotiation and argument. As such, the fiction of the experiment as well as the psychological torture exercised by it (in short *the fiction that there is no argumentation to take place in this experiment*) has already unravelled. This was what may have rendered the last "prod" completely ineffective and simply the last grotesque grunt of the falling tyrant.

- Billig, M. (1996). *Arguing and thinking: A rhetorical approach to psychology*. London: Cambridge UP.
- Blass, T. (2004). *The man who shocked the world. The life and legacy of Stanley Milgram*. Cambridge, MA: Basic Books.
- Blass, T. (Ed). (2000). *Obedience to authority: Current perspectives on the Milgram paradigm*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Blass, T. (2009). From New Haven to Santa Clara: A historical perspective on the Milgram obedience experiments. *American Psychologist*, 64, 37-45.
- Burger, J.M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, 64, 1-11.
- Burger, J.M., Girgis, Z.M. & Manning, C.C. (2011). In their own words: Explaining obedience to authority through an examination of participants' comments. *Social Psychology and Personality Science*. doi: 10.1177/1948550610397632
- De Vos, J. (2009). Now that you know, how do you feel? The Milgram experiment and psychologization. *Annual review of Critical Psychology*, 7, 223-246.
- Edwards, D. (1998). *Discourse and cognition*. London: Sage.
- Erickson, M. (1968). The inhumanity of ordinary people. *International Journal of Psychiatry*, 6, 278-279.
- Gibson, S. (2013a). Milgram's obedience experiments: A rhetorical analysis. *British Journal of Social Psychology*, 52, 290-309.
- Gibson, S. (2013b). "The last possible resort": A forgotten prod and the *in situ* standardization of Stanley Milgram's voice-feedback condition. *History of Psychology*, 16, 177-194.
- Gibson, S. (2014). Discourse, defiance and rationality: 'Knowledge work' in the 'obedience' experiments. *Journal of Social Issues*, 70, 424-438.
- Gibson, S. (2017, forthcoming). Re-visiting Milgram: Developing psychology's archival sensibilities. *Qualitative Psychology*.
- Griggs, R. A., & Whitehead, G. L. (2015). Coverage of recent criticisms of Milgram's obedience experiments in introductory social psychology textbooks. *Theory & Psychology*, 25, 564-580.
- Haslam, A., Reicher, S. & Birney, M. (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.
- Holland, C.H. (1968). Sources of variance in the experimental investigation of behavioural obedience. Unpublished PhD dissertation, University of Connecticut.

- Hollander, M.M. (2015). The repertoire of resistance: Non-compliance with directives in Milgram's "obedience" experiments. *British Journal of Social Psychology*. Advance online publication. doi: 10.1111/bjso.12099
- Jetten, J. & Mols, F. (2014). 50:50 hindsight: Appreciating anew the contributions of Milgram's obedience experiments. *Journal of Social Issues, 70(3)*, 587-602.
- Kant, I. (1997). *Critique of practical reason*. London: Cambridge UP.
- Kelman, H.C. (1967). Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin, 67*, 1-11.
- Kelman, H.C. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues, 29*, 25-61.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371-378.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist, 19*, 848-852.
- Milgram, S. (1965a). Some conditions of obedience and disobedience to authority. *Human Relations, 18*, 57-76.
- Milgram, S. (1965b) Liberating effects of group pressure. *Journal of Personality and Social Psychology, 69*, 127-134.
- Milgram, S. (1967). Obedience to criminal orders: The compulsion to do evil. *Patterns of Prejudice, 1*, 3-7.
- Milgram, S. (1972). Interpreting obedience: Error and evidence; A reply to Orne and Holland. In A.G. Miller (Ed.), *The social psychology of psychological research* (pp. 138-154). New York, NY: The Free Press.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York, NY: Harper and Row.
- Milgram, S. (1983). Reflections on Morelli's "Dilemma of obedience". *Metaphilosophy, 14*, 190-194.
- Millard, K. (2014). Revisioning obedience: Exploring the role of Milgram's skills as a filmmaker in bringing his shocking narrative to life. *Journal of Social Issues, 70*, 439-455. doi: 10.1111/josi.12070
- Miller, A.G. (1986). *The obedience experiments: A case study of controversy in social science*. New York, NY: Praeger.
- Miller, A.G. (2009). Reflections on "Replicating Milgram" (Burger, 2009). *American Psychologist, 64*, 20-27.
- Miller, A.G., Collins, B.E. & Brief, D.E. (1995) Perspectives on obedience to authority: The legacy of the Milgram experiments. *Journal of Social Issues, 51*, 1-19.

- Mixon, D. (1989). *Obedience and civilization: Authorized crime and the normality of evil*. London, UK: Pluto Press.
- Mixon, D. (1972). Instead of deception. *Journal for the Theory of Social Behaviour*, 2, 145-174.
- Mixon, D. (1977). Why pretend to deceive? *Personality and Social Psychology Bulletin*, 3, 647-653.
- Modigliani, A. & 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.
- Mook, D.G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379-387.
- Morelli, M. (1983). Milgram's dilemma of obedience. *Metaphilosophy*, 14, 183-189.
- Nicholson, I. (2011). "Torture at Yale": Experimental subjects, laboratory torment and the "rehabilitation" of Milgram's "Obedience to Authority". *Theory & Psychology*, 21, 737-761.
- Novick, P. (2000). *The Holocaust and collective memory*. London, UK: Bloomsbury.
- Orne, M.T. & Holland, C.H. (1968). On the ecological validity of laboratory deceptions. *International Journal of Psychiatry*, 6, 282-293.
- Penner, L. Hawkins, H.L., Dertke, M.C., Spector, P., & Stone, A. (1973). Obedience as a function of experimental competence. *Memory and Cognition*, 1, 241-245.
- Perry, G. (2013). Deception and illusion in Milgram's accounts of the obedience experiments. *Theoretical & Applied Ethics*, 2, 79-92.
- Perry, G. (2012). *Behind the shock machine: The untold story of the notorious Milgram psychology experiments*. London: Scribe.
- Potter, J. (2001). Wittgenstein and Austin. In Wetherell, M., Taylor, S. and Yates, S.J. (eds.) *Discourse Theory and Practice: A Reader* (pp. 39-46). London: Sage.
- Reicher, S. & Haslam, S.A. (2011). After shock? Towards a social identity explanation of the Milgram "obedience" studies. *British Journal of Social Psychology*, 50, 163-169.
- Reicher, S. Haslam, A. & Miller, A. (2014). What makes a person a perpetrator? The intellectual, moral, and methodological arguments for revisiting Milgram's research on the influence of authority. *Journal of Social Issues*, 70(3), 393-408.
- Reicher, S. Haslam, A. & Smith, J. (2012). Working towards the experimenter: Reconceptualizing obedience within the Milgram paradigm as identification based followership. *Perspectives on Psychological Science*, 7, 315-324.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: McGraw-Hill.
- Russell, N.J.C. (2011). Milgram's obedience to authority experiments: Origins and early evolution. *British Journal of Social Psychology*, 50, 140-162.
- Taylor, S.E., Peplau, L.A. & Sears, D.O. (2003). *Social Psychology*. (11<sup>th</sup> ed). London, UK: Pearson.

Wrightsman, L.S. (1974). The most important social psychological research in this generation?  
*Contemporary Psychology, 19*, 803-805.

Zimbardo, P.G. (1974). On "obedience to authority". *American Psychologist, 29*, 566-567.