WILEY

The second wave of critical engagement with Stanley Milgram's 'obedience to authority' experiments: What did we learn?

David Kaposi [ID]

Open Psychology Research Centre, School of Psychology and Counselling, The Open University, Milton Keynes, UK

Correspondence
David Kaposi, Open Psychology Research Centre, School of Psychology and Counselling, The Open University, Milton Keynes, MK7 6AA, UK.
Email: David.Kaposi@open.ac.uk

Abstract

The study critically examines contemporary academic engagement with Stanley Milgram's classic 'obedience to authority' experiments. It argues that, following what will be termed the 'first wave of criticism' (1964 to mid-1980s) and 'consensus and canonization' (mid-1980s to mid-2000s), this present-day phase has constituted a veritable 'second wave of criticism'. The 'second wave' is reviewed in terms of, first, a return to fundamental dilemmas around the experiments (i.e., replication, ethics, methodology, and theory); and second, a recourse to novel data sources as well as theoretical and epistemological-methodological developments. Such integration of traditional concerns and new perspectives has resulted in focus being redirected to a forgotten aspect of the experiments: the *interactional production of affective dynamics in the lab*. Although the 'second wave' will be critiqued for selectively focusing on either affective atmosphere or interactional process, the paper will conclude with the suggestion that bringing these two aspects together continues to promise a theoretical contribution to the understanding of the experiments not seen before.

1 | INTRODUCTION: THREE PHASES OF ACADEMIC RESPONSE

Despite being the standard feature of psychology textbooks throughout the decades, academic engagement with Stanley Milgram's 'Obedience to authority' experimental series has in neither magnitude nor character been homogenous (Milgram, 1963, 1974). As a simple year-by-year citation count of Milgram's classic 1963 paper demonstrates (Figure 1, updated from Reicher et al., 2014, p. 395), an early spike in interest up to the mid-1980s was followed by a somewhat less productive phase lasting up to the mid-2000s, and by a resurgence that is ongoing.[1] What is more, there is yet more to these phases than the ebb and flow of academic interest; the nature of this academic interest has been periodically changing too. Introducing here briefly the first two, the present paper will examine in depth the most recent phase, and argue for it to be considered a veritable *second wave of criticism*: a period with a substantial contribution not seen since the mid-1980s.

The *first wave of critical engagement* started in the immediate aftermath of Milgram's 1963 publication, and witnessed an unprecedented flurry of arguments concerning the very fundamentals of the experiments. Accepting the validity of the experimental procedures, numerous researchers either sought to experimentally *replicate* Milgram's findings (e.g., Mantell, 1971; Rosenhan, 1969), or interrogated his *theory* to account for those findings (e.g., Helm & Morelli, 1979; Mantell & Panzarella, 1976). Others took a strongly critical stance regarding the status of the experiments' findings themselves, and belaboured the *methodological* point that participants may have seen through the experimental deception (e.g., Holland, 1967; Orne & Holland, 1968). And yet others were less interested in the findings, and focused rather on the *ethics* of the experiments (e.g., Baumrind, 1964; Kaufmann, 1967).

This turbulent phase lasted for around two decades. In 1985, the last directly comparable replication was published[2]; and in 1986, an authoritative monograph by social psychologist Arthur Miller declared Milgram to be either a largely unequivocal winner on critical points (i.e., Miller, 1986, pp. 137 [ethics], 161 [replicability] and 176 [internal validity]), or at least the only one with a plausible argument to offer (p. 253 [theory]). In the years to come, there was still considerable insight offered on the experiments, with a special regard to theoretical matters that have never quite been settled (e.g., Blass, 1991; Modigliani & Rochat, 1995). Yet the hallmark of the period was a discipline-wide agreement concerning the experiments' status and significance. Fittingly, the conclusion of this *period of consensus and canonisation* may be found in a biography of Milgram by Thomas Blass (2004). Published nearly
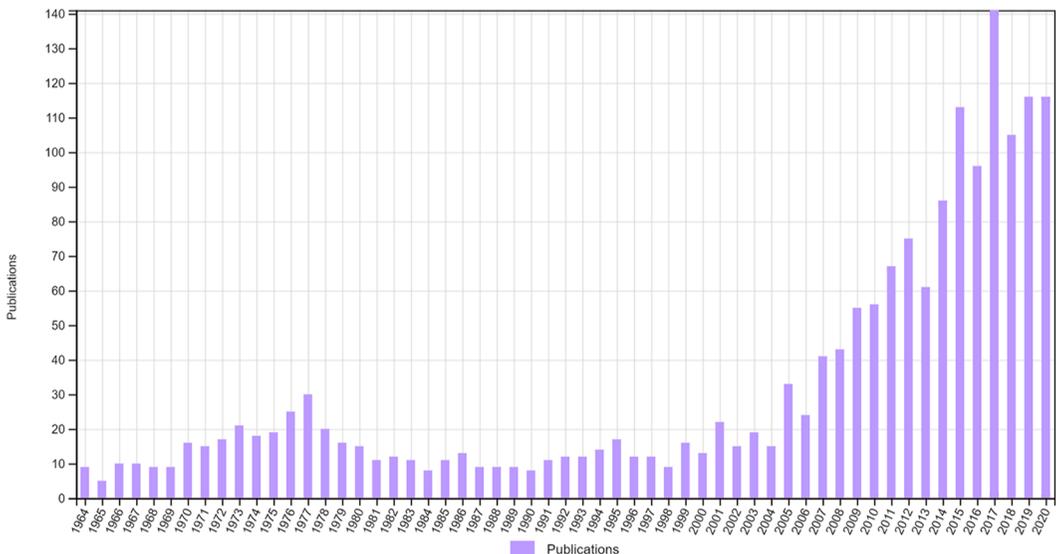

FIGURE 1 Number of publications citing Milgram (1963) from 1964 to 2020 (Updated version of Reicher et al. (2014) p. 395; *Source*: Web of Science)

two decades later, the book's psychological perspective remained synoptic with that of Miller's. Where novelty lay concerned a turn to the man and his legacy, as well as the ever-present cultural phenomenon that had become of the experiments.

All of this was to radically change from the mid-2000s as a *second wave of criticism* has started to take shape. As this paper will demonstrate, the title is warranted not just on account of a surge in interest, but of a novel way of engagement with Milgram's work. For one thing, with a vehemence not seen for decades, contemporary researchers are re-visiting the most fundamental aspects of the experiments—and they are re-visiting them in highly innovative ways. The traditional matter of replication was taken up by Jerry Burger (Section 2.1); ethics and method/validity were critiqued by Ian Nicholson, Gina Perry, and Nestar Russell (Section 2.2); and theoretical/conceptual matters were re-interpreted by Steve Reicher and Alex Haslam, Matthew Hollander and Jason Turowetz, and Stephen Gibson (Section 2.3)—*all* with recourse to theoretical, methodological, analytical innovation, and by drawing on novel sources of data.

For another, this integration of old (concerns) and new (approaches) has redirected focus to two aspects of the original experiments that, despite referred to in Milgram (1963) already, for six decades eluded proper scrutiny. These are the affective dynamics that prevailed in Milgram's lab (in short: that those who obeyed looked tormented by doing so) and the interactions that shaped them (in short: how participants' stress and reluctance was transformed into obedience). The paper will conclude that a fuller understanding of these *affective dynamics as interactional phenomenon* may yet prove to be the second wave's most enduring legacy.

2 | THE SECOND WAVE OF CRITICISM

How may we explain the recent resurgence of interest in the fundamentals of psychology's arguably most well-known experiment?[3] Whilst it would be tempting to attribute it to the seemingly sudden emergence of novel data sources (i.e., experimental and archival), there is nothing in principle in either of these sources which would have prevented academics of the *period of consensus* to systematically draw on them. The Stanley Milgram Papers, in particular, were established at Yale University shortly after Milgram's death in *1984*. What therefore 'suddenly emerged' was not the data themselves but an *academic perspective* that facilitated the systematic and authoritative use of such data. And to account for *this*, we need to look at the intellectual context around and within psychology.

Regarding the experimental replication to be shortly examined, we have to remain speculative in the extreme as social psychologist Jerry Burger's work largely continues to stand on its own.[4] As such, we may link it to the incipient 'replication crisis' in psychology (Maxwell et al., 2015)—or we may simply regard it as a stroke of genius that could have happened at virtually any time. Regarding the use of archival data, however, it has not just become a widespread trend among scholars of the experiments, but one that fully dominates present discourse. Whilst it would be well beyond the remit of this study to provide a full theoretical account for its emergence, a number of trends may tentatively be listed.

The first is a transdisciplinary shift regarding interest in subjective experiences with a special regard to traditionally marginalised voices (e.g., Herman, 2015; Kohler Riessman, 2007); work examined in Section 2.2 concerning ethical aspects of the experiments may not have come about without such developments. The second is a somewhat corresponding transformation in political-historical sensitivity, and the emergence of a genre not simply applying a politically oriented 'hermeneutics of suspicion' to well-established pieces of a disciplinary canon as attempting to 'debunk' them wholesale (e.g., Manning et al., 2007; Le Texier, 2019; Perry, 2018); this trend may again be detected in some of the work in Section 2.2. In addition, a third contributing factor can be the development of certain theoretical or epistemological-methodological perspectives within psychology itself (e.g., Burr, 2015; Postmes & Branscombe, 2010); these were essential to contributions reviewed in Section 2.3.

WILEY

2.1 | Replication

With the 'lull' since the mid-1980s, it was not surprising that in 2009 the *American Psychologist* devoted a special issue to a 'partial replication' of Milgram's *baseline* variation (American Psychologist, 2009; cf., Milgram, 1974, pp. 55–57). Social psychologist Jerry Burger's solution to the *ethics* versus *knowledge* dilemma was to terminate proceedings at the earliest instance where a meaningful comparison could be made with the original experiment. The proposed cut-off point was participants' reading the next question to the learner, having just administered the 150-V shock and heard the learner's subsequent noisy withdrawal of consent from participation. This was convincingly argued to be a watershed moment: 79% of those who continued here went on to be fully obedient (Burger, 2009; cf., Milgram, 1974, p. 59).

Burger (2009, p. 8) found that there was no statistically significant difference in obedience at this point between the original *baseline* condition (80%) and his replication (72.5%). Being told to do so by an experimenter, he concluded, there was no real change from the 1960s in ordinary American citizens' readiness to administer a shock to an innocent stranger, despite his explicit protest to the contrary.

The result was duly taken up by the professional community as well as the media (Cohen, 2008). But in an important sense is was not conclusive, just as, arguably, *none* of the first phase replications had been. What was invariably overlooked was that from his very first publication on Milgram mentioned not one but *two findings.* Other than the simple percentages of obedience rates, he reported that 'the other unanticipated effect was the extraordinary tension generated by the procedures' (Milgram, 1963, p. 377; cf., pp. 375, 376–377; 1964, p. 848; 1965, pp. 66–69; 1974, pp. 41–43).[5]

The significance of this affective atmosphere of 'extreme stress' is conceptual (Milgram, 1963, p. 376). If people do not simply obey but do so visibly against their inclinations, then there is a moment of reluctance that even technically fully obedient participants need to overcome. What we also learn from Milgram, graphically conveyed in his monograph by two chapters devoted in full to participants' often agonizing conduct in-between authority and victim, is that this moment of reluctance often manifested in, and had to be overcome by, interactions with the experimenter (Milgram, 1974, pp. 45–54, 73–88; cf.,; Hollander, 2015). Moreover, this prompts the question that is at the very heart of obedience: what attribute of the situation made unwilling participants to overcome their burgeoning hesitation/reluctance?

The answer, of course, may be assumed to lie with the famous 'prods' Milgram developed for such eventualities (Milgram, 1974, p. 21; cf., Table 1). Whilst Miller (1986, p. 8) called these 'the most central methodological elements in the paradigm', constituting 'the operationalization of authority', the astonishing fact is that as of 2009 literally nothing was known about how they worked either in the original experiments or in replications.[6] From Milgram (1963, 1965, 1974) to Burger (2009, 2014, 2015), explanations as to what features of the social context may have influenced participants' behaviour invariably listed unchanging attributes of the situation, with no regard to the 'prods'. Such unchanging attributes cannot address the question posited above, however, as they are by definition unresponsive to the moments of reluctance/resistance that, according to Milgram, characterised even fully obedient participants.

TABLE 1 The main prods used in the "Obedience to authority" experimental series (Milgram, 1974, pp. 21–22; for further, ad hoc, prods, see Gibson, 2019a, pp. 106–116)

Prod 1: Please continue, and variants.

Prod 2: The experiment requires that you continue.

Prod 3: It is absolutely essential that you continue.

Prod 4: You have no other choice, you *must* go on.

In addition, when relevant, the following two special prods were also used:

Although the shocks may be painful, there is no permanent tissue damage.

Whether the learner likes it or not, you must go on until he has learned all the word pairs.

Miller very reasonably hypothesised that it was the stern and authoritative Prod 3 and 4 that 'strongly influenced subjects to shock' (Miller, 1986, p. 8; cf., 2009, p. 25), and it is for this reason that Burger's lasting contribution may not simply be his replication of Milgram's obedience percentages but a follow-up article *demonstrating that in his replication Prod 4 never worked* (Burger et al., 2011).

Indeed, it is when he notes that the finding 'challenges interpretations of the obedience studies based on the notion that participants were following orders' that Burger exceeds the horizons of both the 'first wave' and the 'consensus' (Burger et al., 2011, p. 466). Granted, subsequent commentaries overstate the case somewhat by equating the powerlessness of Prod 4-in-use with Milgram's conception of obedience lacking in validity.[7] Still, the observation definitely demonstrates how little we know about what went on Milgram's lab to transform stress and unwillingness into the administration of shocks.

Despite making the observation and duly concluding that it is challenging the traditional interpretation, Burger himself left the matters unexplored. His subsequent publications did not even make reference to the finding concerning Prod 4 (Burger, 2014, 2015). Others, however, soon followed the trail he blazed.

2.2 | Ethics and internal validity

As mentioned in the introduction, Burger's replication aside, the dominant source for second wave inquiries has become primary data from the original experimental series: research funding applications, notebooks, audio tapes of actual experimental sessions, post-experimental questionnaires and correspondence with participants, etc. Milgram had dutifully kept these and following his death in 1984 they were donated to Yale University (Kaplan, 1996). Although the Stanley Milgram Papers were sporadically consulted during the period of canonisation (cf. Blass, 2004; Modigliani & Rochat, 1995), from around the mid-2000s they have become the dominant source for understanding the experiments.

The first findings drawing on archival material immediately troubled the canonical image of the experiments. First, evidence was reported of considerably more post-experimental distress from participants than Milgram had admitted (Nicholson, 2011; Perry, 2012). Second, the experiment's debriefing/dehoaxing procedures were revealed to have continued keeping participants in the dark as regards the fictitious nature of the experiments (Brannigan et al., 2015; Nicholson, 2011, 2015; Perry, 2012, 2013). Third, a number of significant divergences from the published record were detected in the experimental procedures (Gibson, 2013a; Perry, 2012; Russell, 2011, 2018).[8]

To start with ethical matters, it may come as surprise but the archival revelations do not quite bear out claims to an interpretative revolution. Concerning after-effects on participants, Milgram, as mentioned above, himself talked about 'extreme tension' in the very first paragraph published on the experiments (Milgram, 1963, p. 375), and reported in 1964 that 0.5% of his participants responded to a follow-up questionnaire as 'very sorry' to have participated (Milgram, 1964, p. 849). Whilst archival findings rendered this 'very sorry' well and truly experience-near, they have not to date questioned the original figure. Concerning Milgram's debriefing practice, the picture is more complex. What most participants were immediately told was that the hoax of the experiment comprised of the shock machine only generating dangerous shocks for 'small animals, mice and small rats' but not for humans, and it is clear that this strange (and untrue) story seems to have contributed to confusion and upset. Yet this misinformation also contained the most important element of the truth (i.e., that the shocks the learner received were not dangerous); and Milgram moreover *never claimed* either to have conveyed more of the truth during the immediate dehoax: 'at the very least all subjects were told that the victim had not received dangerous electric shocks' (Milgram, 1964, p. 849).

Ethical criticisms of the second wave may therefore fail in matching sensitivity to participants' concerns with equal sensitivity to Milgram's actual claims—and indeed to the context within which those claims were made.[9] But, from the perspective of this study, their major shortcoming is of a different order. It lies with their exclusive focus on why the experiments should never have happened in the first place, and the corresponding neglect of attempting to understand what had actually happened in the experiments. This is most unfortunate, as it would be of obvious

ethical and theoretical importance to understand how the affective dynamics of the experimental procedures may have contributed to the experiments' results.

Moving now on to concerns about methodology/internal validity, the paper will turn to what appears to be a truly significant discrepancy: the experimenter's handling of participants' misgivings. Gina Perry (2012, pp. 132–134) claimed that the experimenter pressed the exclusively female participants of Experiment 8 far more forcefully than others; Nestar Russell detected the experimenter generally to have 'invented his own vigorously demanding prods' (Russel, 2018, p. 175); and Stephen Gibson documented various previously unknown practices, most striking of which is the experimenter three times leaving the room in the early Experiment 2 to 'check' whether the learner is fine (Gibson, 2013b, 2015, 2019a).

The significance of such findings is twofold. First, they suggest that, according to standards widely accepted today, the 'obedience to authority' experiments may in the strict sense not have been *experiments*: the experimenter's conduct constituted no standardised environment across, or possibly even within, the various conditions. Yet, second, these observations are substantially more significant than even this. They promise to shed light onto the interactional production of affective dynamics that was highlighted but left unexplored by both Burger and the ethical criticism. The findings should therefore be understood not simply as dismissing the experimental status of Milgram's findings, but as re-interpreting them.

Gibson's framework will be examined in the next section. For their part, Russell and Perry both assert that the experimenter's use of prods amounted to such pressure that participants acted under 'coercion' (Perry, 2012, pp. 53, 111, 132–134; Russell, 2018, pp. 175, 211, 221). And given that the term 'coercion' implies the use of force, to whatever degree these authors understand Milgram's experimenter to apply such force (i.e., bullying, threatening, extorting, etc.), it has momentous implications regarding participants' responsibility. Namely, the significance of Milgram's findings always depended on the assumption that, putting their purely internal struggles and fantasies aside, participants were completely free to obey or disobey (Milgram, 1964, p. 851; 1974, pp. xx-xi, 32, 41, 141–145, 175; cf.,; Blass, 1991, p. 398; 2004, p. 93; Miller, 1986, p. 8); and it was as a consequence of this assumption that they were understood to be fully responsible for the 'torture' in the lab. If, however, the social action transforming their stress and reluctance into obedience was an act of coercion, this freedom partly or even fully disappears—and so does their responsibility on which the canonical meaning of the experiments has always depended.[10]

It is little short of astonishing therefore that, promising such truly transformative insight, *both Perry and Russell ultimately settle on completely different arguments in accounting for participants' widespread obedience.* Perry, following her repeated assertions that the experimenter put 'enormous pressure' on participants (Perry, 2012, pp. 54, 11, 132), rather counter-intuitively rekindles Martin Orne's 'demand characteristics' thesis, arguing that it was simply participants sceptical of the reality of the experimental set-up that tended to obey (Perry et al., 2020; cf., Orne & Holland, 1968; Perry, 2012, pp. 159–177). And Russell, equally surprisingly, ends by invoking Milgram in all but name as he argues that participants consciously chose to jeopardize the learner's wellbeing in order to evade an 'awkward' or 'impolite' confrontation with the experimenter (Russell, 2018, p. 227; Russell & Gregory, 2011, p. 502; cf., Milgram, 1974, pp. 149–152).

How is it possible that neither Perry nor Russell follow up their claim regarding the experimenter's coercive action? The reason may be that it appears very difficult to find actual data (in vivo or recollective alike) to back it up. Importantly, it may be noted that Perry and Russell's interpretative position has actually a lot in common with that of Milgram's. To reiterate once more, 'extreme tension' in the lab has been public knowledge from the very inception of the experiments. Milgram clearly and repeatedly conveyed that a substantial number of participants obeyed *as though* they were coerced to do so. Yet it also seemed clear to him that all the experimenter did was telling them in a polite if determined way to continue, backing up this request with some simple arguments. This is why Milgram felt justified in attributing participants' 'extreme tension' not to anything the experimenter did but to participants' psychological characteristics (Milgram, 1974, pp. 152, 162).

As regards Perry and Russell, the data they bring to support a different conclusion is extremely scarce—and highly problematic even at that. Perry (2012, p. 134) offers but two pieces/types of evidence. She claims that the

experimenter 'insisted that one woman continue 26 times. He argued with two others 14 times [etc.].' And she notes that at one point 'Subject 2029 [...] switched off the machine in defiance', whereupon the experimenter 'switched it back on and insisted that she continue' (Perry, 2012, p. 134). The first claim appears to be based on a simple misunderstanding. The experimenter was not limited in 'arguing' with the participants as many times (i.e., after as many shocks) as such arguments were initiated by the participants themselves. The only *rule* concerned the use of Prods one to four *within* one voltage interaction. Regarding the second claim, 'Subject 2029's' said interaction unfolds from 7 min 28 seconds on the tape, following the learner's forceful outcry at the 270V-shock. It is indeed a highly revealing exchange but seems to have been misheard by Perry. 'We have to be careful', says the participant first, and then, when the experimenter replies with a 'Pardon me?', whispers (!): 'We have to be careful. I turned it off.' 'Yeah, turn it back on again', says the experimenter 0.6 seconds thereafter, followed by 'Subject 2029' (presumably) herself switching back on the machine immediately—and continuing without further ado. Similarly, the few and unsystematic examples that Russel (2018, p. 175) offers include the experimenter telling a participant the technical clarification that they cannot go over and talk to the learner, as well as the experimenter formally reassuring a participant that they can keep the $4.50 cheque. Later on, Russell repeatedly mentions the experimenter taking full responsibility for the proceedings; that is, once again, reassuring participants. Certainly, all of these examples by Russell may be classed as devious or manipulative. But it is not obvious at all why/how they would constitute an atmosphere of coercion.

To sum up, then, just as Burger's contribution, archival work on both ethics and internal validity feels for now somewhat incomplete. There, we find the highly intriguing interactional observation regarding the connection between disobedience and the experimenter's the use of Prod 4; yet the affective or conceptual implications of such observations are neglected. Here, we see the twin intuitions that Milgram's participants may have had traumatic experiences and that the experimenter may have coerced them to do something they did not want to; yet the empirical grounding of these claims in actual interactions in the lab is neglected.

It is with this sense of incompleteness that the paper now turns to contemporary researchers who proposed a more systematic theoretical and empirical inquiry into the experimenter's conduct.

2.3 | Theory and concepts

The perspectives reviewed in this last section offer intriguing possibilities to illuminate the experiments. They are distinguished from the approaches previously examined in three important respects. First, they are grounded in theoretical and epistemological-methodological developments in social psychology that have emerged since the time Milgram conducted his experiments. Second, they do not simply make observations about the experiments, but proceed with an aim to systematically re-theorise/-conceptualise Milgram's findings. Third, they offer a radically different picture of the experimenter-participant dynamic: in the place of a potentially traumatising coercive atmosphere, we will see first the dynamic of motivational leader and enthusiastic follower, and second that of two debaters.

The hypothesis that the operative dynamics in the experiments was not obedience to an authority but following a leader was put forward in a vigorous succession of papers by social psychologists Stephen Reicher and Alex Haslam (e.g., Haslam & Reicher, 2018; Reicher et al., 2014; Reicher & Haslam, 2011). They stated that whilst Milgram produced a robust fact, not only did his 'agentic shift' theory (positing a cognitive shift whereby an actual agent feels no responsibility for their action) remain empirically unproven, it is downright inadequate to the experiments' multivocal nature. In the laboratory situation participants had constantly to choose between the competing voices of the experimenter and the learner. The theoretical explanation, however, depicted them as robotic, 'inattentive bureaucrat[s]' who are simply deaf to one of these voices (Haslam et al., 2015, pp. 58–60).

Not so, asserted Reicher and Haslam. The competing explanation forms part of the authors' wider ambition to re-theorise the phenomenon of leadership from a social identity theory-base (Haslam et al., 2010; Reicher & Haslam, 2019). The theory holds that human action is intertwined with agents' sense of who they are and what groups they belong to (Haslam et al., 2010). Accordingly, it is claimed that participants in the experiments actively

identified with the experimenter, who, in turn, made efforts to cultivate participants' engagement with his scientific project. In this reinterpretation, we do not see obedient participants passively following orders or turning a blind eye to an immoral course of action; we have the willing assumption of what they understand to be a *virtuous* course of action (Haslam et al., 2015, pp. 60–62). Little surprise that in Reicher and Haslam's 'engaged follower' model a formerly unknown affective side of the participants' conduct is revealed: a positive and symmetrical relationship with the experimenter that involves 'loyalty, trust, helpfulness, and the fulfilment of obligation' (Haslam et al., 2016, p. 67).

This image of evil enthusiastically done in the name of what is understood to be the greater good is certainly an evocative (not to mention: frightening) one. Yet, contradicting as it does all previous accounts of the experiments,[11] it is highly intriguing that supporting it Haslam and Reicher only ever offered a single piece of empirical evidence that draws on primary data from the archive (for evidential data generated by the authors themselves, see Haslam et al., 2014; Reicher et al., 2012).

It also appears to be data that do not at face value allow Reicher and Haslam to make the claims they wish to put forward. The evidence is comprised of pre-structured items and spontaneous responses to the follow-up questionnaires Milgram sent out to participants in the aftermath of the series. It is not just that Reicher and Haslam's conclusion that participants were 'happy to have been of service' (Haslam et al., 2015, p. 80) is based on retrospective accounts, and with clearly transformative events occurring in-between (i.e., the debriefing, or the reception of a research report from Milgram). As sociologists Matthew Hollander and Jason Turowetz argue in a respectfully phrased but devastating paper, the fundamental issue is a confusion of reference. Quite simply, in the data Haslam and Reicher analyse, Milgram's respondents write about their feelings/views regarding what they now know to have been 'the obedience to authority experiments'. They make absolutely no reference to their cognitive or affective processes at the time when they had been under the impression of administering electric shocks in a 'memory and learning experiment' that subsequently proved to be fictitious (cf., Hollander & Turowetz, 2017, p. 671).[12]

What is more, in the main body of their paper Hollander and Turowetz analyse the accounts participants did offer of/for their conduct in this 'memory and learning experiment', and offered it moreover precisely before its fictitious nature had been revealed. In these pre-debriefing interviews, they report very little in the way of cognitive themes that could be interpreted as 'identification', and absolutely nothing that can possibly be understood as 'happiness' or 'loyalty'. What they find instead is participants' trying to make sense of their 'often stressful experience' (Hollander & Turowetz, 2017, p. 665) in distinctly *Milgramesque* ways: that is, that they merely 'followed instructions' and accepted the experimenter's claims regarding the learner 'not being harmed', or that they just would not have disputed the claim that the experiment required them to continue (Hollander & Turowetz, 2017).[13]

If the 'engaged follower' model was found categorically wanting in its lack of engagement with the original experimental sessions, no such accusation can be made concerning the last piece of second wave research to be examined in this paper. Little after Haslam and Reicher introduced their social identity framework, social psychologist Stephen Gibson too proposed a novel account of the experiments (Gibson, 2013a, 2013b, 2014, 2015). Gibson's position was developed by drawing on rhetorical/discursive psychological analysis of audio recordings of disobedient sessions from Milgram's Experiments 2 and 8 (Gibson, 2013a, p. 292; 2014, p. 425; cf., 2013b, 2015). A psychological approach with origins dating back to the second half of the 1980s, discursive psychology understands participants' talk-in-interaction as performative of emerging factual, cognitive, affective realities. Gibson put forward the highly interesting thesis that such approach reveals the experiments to be a site of 'rational debate […], a contest of persuasion and rhetorical skill' (Gibson, 2014, pp. 434–435; cf., 2013a, p. 304; 2019a, p. 103).

What did participation in a 'rational debate' entail? Something perhaps less surprising regarding disobedient participants: they were demonstrated to have engaged in argumentation as they approached the point where they were considered technically disobedient. Something far more counter-intuitive regarding the experimenter: he was found not to have robotically reacting with standardised prods, but flexibly engaging with participants' arguments. But the most radical element of Gibson's rhetorical contribution was challenging the fundamentals of the traditional account of *obedience*. Gibson did not just once again demonstrate that the ultimate order of Prod 4 never worked, but also claimed that the reason for this may have been that its deployment also implied a counterargument (i.e.,

freedom of choice) which could be routinely used by participants to bat it back (Gibson, 2013a, pp. 300–304). The emphatic conclusion followed that whatever prompted Milgram's participants to continue, it was not explicit orders; and that inasmuch as the cornerstone of the concept of obedience has always been the passive following of orders, this needed to accommodate or even be replaced by the concept/practice of argumentation (Gibson, 2013a, 2014, 2015).[14]

Gibson's re-conceptualisation of the experiments was quickly taken up within the discipline, and rightly so: his claims concerned precisely that area of interactional processes already highlighted by Milgram to be of crucial importance, yet never systematically addressed by either him or any of his critics. At the same time, there was in Gibson's early contribution an absolutely crucial issue not attended to, and hence potentially compromising the validity of his conclusions. He was, of course, upfront about the fact that the data extracts he analysed were exclusively from disobedient sessions.[15] Yet this acknowledgment of the nature of his data did not extend to the appreciation of just how restricting such an empirical base is regarding the remit of possible conceptual conclusions. It is not just that any inference to 'obedience', or to the experiments *as such*, must clearly be treated as tentative in the extreme if based on an analytical sample that is in fact restricted not simply to disobedient sessions but to exact moments of immediate disobedience even within those sessions.[16] It is also that, arguably, even claims regarding 'disobedience' become questionable if based on so partial a sample.[17]

This troubling matter came to the fore when, to his absolute credit, years later Gibson suddenly introduced 'experimental sessions in which participants proceeded with the experiment with little or no attempt at disobedience' (Gibson, 2019a, p. 12; 2019b). It was here that he finally revealed that '*many* obedient participants' proceeded in the 'absence of argumentation (Gibson, 2019a, p. 168). Indeed, we now learned that it was not just that the prods whose flexible use was described at such length in Gibson's earlier papers/chapters were actually unnecessary to make participants obey, but that '*they were not generally very effective*' (Gibson, 2019a, p. 197—emphasis mine).

In other words, it would appear that what Gibson's complete work on the experiments reveals is not a reality where people verbally argue with the experimenter: some (i.e., disobedient ones) with success whereas others (i.e., obedient ones) with no success. It is rather something altogether more intriguing: a reality where some argue (and disobey) whilst some simply do not argue at all (and obey). Likewise, it is not a reality where the experimenter's success lies in flexibly engaging with these arguments. It is one rather where his success lies in keeping participants from initiating arguments in the first place.

What do we make of this somewhat odd state of affairs? What kept obedient participants from arguing when there simply is no example in Gibson's work for an argument that would have proved unsuccessful?[18]

To address this question, it is noteworthy that whilst Perry or Russell, or Reicher and Haslam failed to anchor their intuitions/hypotheses regarding affective dynamics in the actual interactions in the lab, Gibson appears to have done the exact opposite. In his work, the interactional process of what he understands to be a 'rational debate' unfolds with no reference whatsoever to the presence, or indeed absence, of any corresponding affective dynamics. Debate, just as the conspicuously complete lack of debate, is analysed here as though taking place in an atmosphere of equanimity, where participants can freely deliberate and put forward arguments. Yet if participants really chose freely to resist from arguments that would have interrogated the experimenter's willingness to torture someone, then the only explanation of some psychological plausibility for this situation is not, as Gibson continues to propose, that they must have been convinced by the 'walls' of why torture is an acceptable thing in the name of science—but that the presence of such 'walls' as Yale University's fully convinced them that this supposed 'torture' was in fact a *hoax*.

To be sure, no one has ever put forward the thesis that the lab's affective atmosphere was that of equanimity, or that interactions between participants and the experimenter represented something of a free exchange of ideas. For this reason, it is hard not to conclude at this point that Gibson's work may have for theoretical or other reasons overlooked a feature of the proceedings which Milgram, Perry, and Russell all sensed. And if the observation that there was 'extreme tension' in the lab is of some validity, then the dichotomous image where a participant, by and large, either argues-and-disobeys or sits simply in silence (at any rate: sits without arguing), invokes not so much the agora in Athens, fifth century BC, but Kossuth Square in Budapest 1956.

3 | CONCLUSIONS

This paper sought to examine the contribution of current critical engagement to our understanding of the Milgram experiments. Its conclusions are twofold.

First, the integration of fundamental concerns of old with more contemporary ethical sensitivities and theoretical-methodological developments warrants a designation of the present period as a distinctive *second wave of criticism.* Most importantly, second wave researchers from various angles tapped into an aspect of the experiments that had already been intimated by Milgram yet eventually left unaccounted for both by him and previous periods of academic engagement: the techniques by which participants' manifest reluctance and distress were transformed. In the estimation of the present paper, these techniques appear to be an absolutely crucial ingredient in understanding how obedience was produced in Milgram's lab.

Second, though, the paper also concludes that the second wave has some way to go in accounting for these interactional processes and affective dynamics. Namely, a curious disjuncture characterised the approaches examined here in that they either focused on affective dynamics whilst largely ignoring interactional processes (e.g., Perry, Russell, Reicher and Haslam); or focused on interactional processes as though they were divorced from affective realities (e.g., Burger, Gibson). Integration eluded researchers just as, curiously, it seemed to have eluded Milgram himself 60 years ago.

Yet the present paper will end on a note of optimism, and not solely for the reason that the focus now on interactional process, now on affective dynamics, clearly holds the promise of integration. Another reason is the growing sophistication in the use of data sources. Part of this concerns Burger's creative (if ultimately incomplete) use of his own experiments not simply as an arbiter in the question of replication but as a basis from which to explore the phenomenon of obedience. And another part concerns the use of the archive. Again, early use of the archive may have conveyed a somewhat naïve belief in primary data simply being the unambiguous arbiter on Milgram's experiments (or even character). In effect, though, researchers have consequently demonstrated a considerably more productive use of the archive, for the characterization of which we may conclude with the evocative formulation of French philosopher, Jacques Derrida: '[t]he interpretation of the archive […] can only illuminate, read, interpret, establish its object, namely a given inheritance, by inscribing itself into it, that is to say by opening it and by enriching it enough to have a rightful place in it.' (Derrida, 1998, p. 67).

If largely refraining from direct recourse to archival data, the present paper wished to further the *second wave*'s most valuable contribution of such 'enriching' the archive Stanley Milgram had brought to being.

ORCID

David Kaposi https://orcid.org/0000-0001-9910-0172

ENDNOTES

[1] Controlling for effects of changing citational practices and general trends within social psychology does not significantly alter this graph, save perhaps for flattening somewhat the spike of the last phase.

[2] For a full list of all 17 first wave replications, see Blass (2004, p. 302).

[3] As for the immediate quantitative change around the mid-2000s, reference can also be made to external events at the American prison camp of Abu Ghraib (Zimbardo, 2008).

⁴ Doliński, Grzyb and colleagues' widely known partial replication in Poland is certainly an exception here. The reason why it is not mentioned in the main body is that the ethics versus knowledge dilemma was solved in that experiment in such a way so as to make any comparison with Milgram or Burger impossible. Namely, as Gibson (2019a) notes, in the Polish replication the experiment was terminated before 150 V, and thus before the first verbal protest of the 'learner' (Doliński et al., 2017).

⁵ An exception is Rosenhan (1968) who in an unpublished manuscript reflects on the fact that his replication failed to reproduce this 'extreme tension'.

⁶ What is more, and somewhat shocking, despite Milgram (1963) clearly describing his prods, replications diverged from these with liberty and without providing arguments for doing so (e.g., Kilham & Mann, 1974, p. 699; Shanab & Yahya, 1977, p. 268). The learner's grunts or protests were also deployed in a haphazard manner, although Milgram only published the exact protocol for these in his book (Milgram, 1974, p. 56; cf., Milgram, 1965, p. 62).

⁷ See fn 14, below.

⁸ Corresponding to such claims, a number of critical comments concerned not so much Milgram's work but his character. It would be beyond the remit (and perhaps propriety) of this paper to attend to these.

⁹ What may also be added here is, first, the claims that the idea of the partial debrief may actually have originated not from Milgram himself, but from "the National Science Foundation [that] had warned him that an honest debriefing would increase the risk of 'subject contamination…'" (Russell, 2018, pp. 113–114). And second, that in the 1960s not just the scientific but even the *ethical* merits of full debriefing were still hotly debated (cf., Ring et al., 1970; Rosenhan, 1968).

¹⁰ It is for this reason that Milgram has repeatedly displayed an active lack of interest in conceptual issues surrounding obedience—save for the equally repeated and forceful statement that his participants always and fully acted out of free will (Milgram, 1965, p. 58; 1983, p. 190).

¹¹ Intriguingly, the only account of some resemblance to the 'engaged follower' model is actually Milgram's own 'agentic shift' theory. For instance: '[…] the idea of science and its acceptance as a legitimate social enterprise provide the overarching ideological justification for the experiment. […] Obedience could be secured [that is, without such justification], but it would not be the form of willing obedience, in which the person complies with a strong sense of doing the right thing' (Milgram, 1974, p. 142). And 'Morality does not disappear [that is, in the "agentic state"], but acquires a radically different focus: the subordinate person feels shame or pride depending on how adequately he performed the actions called for by authority. Language provides numerous terms to pinpoint this type of morality: *loyalty, duty, discipline* […]' (Milgram, 1974, p. 144—emphasis in the original).

¹² Interestingly, parts of the overall dataset Haslam et al. (2015) consult do contain references to participants' thoughts and feelings at the time of the experiments. Yet the authors do not analyse these, nor do they explain why they choose to exclude data of such obvious apparent relevance.

¹³ The authors themselves propose that their third (and numerically rather infrequent) theme would lend support to the 'engagement followership' model. This study has to disagree. The ideological construct of science featured prominently already in Milgram's account; any authority figure was argued to have derived its status from an ideological context (Milgram, 1965, p. 71; 1974, pp. 70, 142). Where the novelty of Reicher and Haslam's contribution would come is not simply the relevance of such context or acceptance of it, but *enthusiastic identification* with it on the participant's side and *active cultivation* of this identification on the experimenter's side. It is for the same reason that, as Hollander and Turowetz correctly note (2018, pp. 5–6), Reicher and Haslam's interpretative effort in their rejoinder to find evidence for their theory in the "learner was safe" theme is invalid: acceptance of or even trust in an authority's reassurance is something altogether different from identification with his overall project (Reicher & Haslam, 2019).

¹⁴ Such argument appears now to routinely follow the finding regarding Prod 4 (e.g., Gibson, 2015, p. 650; Reicher & Haslam, 2011, p. 167). However, it is one thing to say, validly, that the finding troubles the traditional account of the experiment given in terms of obedience to orders, and quite another thing to claim that it conclusively invalidates it. Any soldier who at some point disobeys an order would have by that point obeyed a number of previous orders; otherwise they would not be a soldier in that moment (cf., Kaposi, 2017, last fn).

¹⁵ It should be mentioned though that Gibson's very first abstract published on the experiments does not mention that his evidence base consists of exclusively disobedient sessions/moments. Yet it concludes: "These observations are discussed in terms of their implications for theory and research on dis/obedience, with a specific focus on the concepts of choice and agency and the nature and meaning of dis/obedience" (Gibson, 2013a, p. 290).

¹⁶ The 18 extracts analysed by Gibson (2013a, 2013b, 2014, 2015) are followed by a mean number of 0.76 shocks administered, with both mode and median being nil.

[17] In other words, we do not know what had happened before those moments of argumentation that immediately preceded the experimenter's termination of the procedure. Did these participants argue throughout their sessions? Or were they silent? Or did they do something else?

[18] Gibson continues at this point to assert the relevance of argumentative processes, positing that "rhetoric is not only to be found in the words used in the obedience experiments, but in the walls of the laboratory" (Gibson, 2019a, pp. 10, 194–198; 2019b, p. 255). Yet, unless we are capable of imagining rational arguments (of the walls) that were extremely effective only on the condition of not being made explicit (by the experimenter); and unless we can conceive of participants who literally could not think up an argument of some plausibility to offer in a debate on whether or not to torture someone, we will conclude that the late addition of obedient sessions does not complete Gibson's early account but subverts his entire argumentation-focused perspective. And this subversion, indeed, includes the perspective on disobedience too, for once we see that it was not 'creative' (Gibson, 2013a, pp. 292–293) arguments that led to disobedience but *any argument*, then we also see that what demonstrably differentiated disobedient participants from obedient ones was not their rhetorical skill but their simple willingness to initiate an (i.e., any) argument.

REFERENCES

American Psychologist. (2009). *Obedience: Now and Then*, 64(1), 1–45.

Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's "Behavioural study of obedience". *American Psychologist*, 19, 421–423. https://doi.org/10.1037/h0040128

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. https://doi.org/10.1037/0022-3514.60.3.398

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

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

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

Burger, J. M. (2014). Situational features in Milgram's experiment that kept his participants shocking. *Journal of Social Issues*, 70(3), 489–500. https://doi.org/10.1111/josi.12073

Burger, J. M. (2015). Obedience. In S. Harkins, K. D. Williams, & J. Burger (Eds.), *The Oxford handbook of social influence*. Advance Online Publication. https://doi.org/10.1093/oxfordhb/9780199859870.013.5

Burger, J. M., Girgis, Z. M., & Manning, C. C. (2011). In their own words: Explaining obedience to authority through an analysis of participants' comments. *Social Psychological and Personality Science*, 2, 460–466. https://doi.org/10.1177/1948550610397632

Burr, V. (2015). *Social constructionism* (3rd ed.). Routledge.

Cohen, A. (2008). Four decades after Milgram, we are still willing to inflict pain. *New York Times*, 28(Dec), 32.

Derrida, J. (1998). *Archive fever: A Freudian impression*. University of Chicago Press.

Doliński, D., Grzyb, T., Folwarczny, M., Grzybala, P., Krzyszycha, K., Martynowska, K., & Trojanowski, J. (2017). Would you deliver an electric shock in 2015? Obedience in the experimental paradigm developed by Stanley Milgram in the 50 years following the original studies. *Social Psychological and Personality Science*, 8, 927–933. https://doi.org/10.1177/1948550617693060

Gibson, S. (2013a). Milgram's obedience experiments: A rhetorical analysis. *British Journal of Social Psychology*, 52, 290–309. https://doi.org/10.1111/j.2044-8309.2011.02070.x

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. https://doi.org/10.1037/a0032430

Gibson, S. (2014). Discourse, defiance, and rationality: "Knowledge work" in the "obedience" experiments. *Journal of Social Issues*, 70, 424–438. https://doi.org/10.1111/josi.12069

Gibson, S. (2015). Rhetoric and resistance. *The Psychologist*, 28, 648–651.

Gibson, S. (2019a). *Arguing, obeying and defying: A rhetorical perspective on Stanley Milgram's obedience experiments*. Cambridge University Press.

Gibson, S. (2019b). Obedience without orders: Expanding social psychology's conception of 'obedience'. *British Journal of Social Psychology*, 58(1), 241–259.

Haslam, S. A., Ellemers, N., Reicher, S. D., Reynolds, K. J., & Schmitt, M. T. (2010). The social identity perspective today: An overview of its defining ideas. In T. Postmes & N. R. Branscombe (Eds.). *Rediscovering Social identity: Key readings* (pp. 341–356). Psychology Press.

Haslam, S. A., & Reicher, S. D. (2018). A truth that does not always speak its name: How Hollander and Turowetz's findings confirm and extend the engaged followership analysis of harm-doing in the Milgram paradigm. *British Journal of Social Psychology*, 57, 292–300. https://doi.org/10.1111/bjso.12247

Haslam, S. A., Reicher, S. D., & Birney, M. E. (2014). Nothing by Mere authority: Evidence that in an experimental analogue of the Milgram paradigm participants are motivated not by orders but by appeals to science. *Journal of Social Issues*, *70*(3), 473–488. https://doi.org/10.1111/josi.12072

Haslam, S. A., Reicher, S. D., & Birney, M. E. (2016). Questioning authority: New perspectives on Milgram's 'obedience' research and its implications for intergroup relations. *Current Opinion in Psychology*, *11*, 6–9. https://doi.org/10.1016/j.copsyc.2016.03.007

Haslam, S. A., Reicher, S. D., Birney, M. E., Millard, K., & McDonald, R. (2015). 'Happy to have been of service': The Yale archive as a window into the engaged followership of participants in Milgram's 'obedience' experiment. *British Journal of Social Psychology*, *54*, 55–83. https://doi.org/10.1111/bjso.12074

Helm, C., & Morelli, M. (1979). Stanley Milgram and the obedience experiment: Authority, legitimacy, and human action. *Political Theory*, *7*, 321–345.

Herman, J. (2015). *Trauma and recovery: From domestic abuse to political terror*. Basic Books.

Holland, C. H. (1967). *Sources of variance in the experimental investigation of behavioral obedience [Unpublished PhD thesis]*. 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*, *54*, 425–444. https://doi.org/10.1111/bjso.12099

Hollander, M. M., & Turowetz, J. (2017). Normalizing trust: Participants' immediately post-hoc explanations of behaviour in Milgram's 'obedience' experiments. *British Journal of Social Psychology*, *56*, 655–674.

Hollander, M. M., & Turowetz, J. (2018). Multiple compliant processes: A reply to Haslam and Reicher on the engaged followership explanation of "obedience" in Milgram's experiments. *British Journal of Social Psychology*, *57*(2), 301–309.

Kaplan, D. E. (1996). The Stanley Milgram papers: A case study on appraisal of and access to confidential data files. *American Archivist*, *59*, 288–297.

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

Kaufmann, H. (1967). The price of obedience and the price of knowledge. *American Psychologist*, *22*, 321–322. https://doi.org/10.1037/h0037664

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. https://doi.org/10.1037/h0036636

Kohler Riessman, C. (2007). *Narrative methods for the human sciences*. Sage.

Le Texier, T. (2019). Debunking the Stanford prison experiments. *American Psychologist*, *74*, 823–839. https://doi.org/10.1037/amp0000401

Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses. *American Psychologist*, *62*(6), 555–562. https://doi.org/10.1037/0003-066X.62.6.555

Mantell, D. M. (1971). The potential for violence in Germany. *Journal of Social Issues*, *24*(4), 101–112.

Mantell, D. M., & Panzarella, R. (1976). Obedience and responsibility. *British Journal of Social & Clinical Psychology*, *15*, 239–245. https://doi.org/10.1111/j.2044-8260.1976.tb00030.x

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*, 487–498. https://doi.org/10.1037/a0039400

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378. https://doi.org/10.1037/h0040525

Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, *19*, 848–852. https://doi.org/10.1037/h0044954

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, *18*, 57–76. https://doi.org/10.1177/001872676501800105

Milgram, S. (1974). *Obedience to authority: An experimental view*. Harper & Row.

Milgram, S. (1983). Reflections on Morelli's "dilemma of obedience". *Metaphilosophy*, *14*, 190–194.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. Praeger.

Miller, A. G. (2009). Reflections on "replicating Milgram" (burger, 2009). *American Psychologist*, *64*, 20–27. https://doi.org/10.1037/a0014407

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*(3), 107–123. https://doi.org/10.1111/j.1540-4560.1995.tb01337.x

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

Nicholson, I. (2015). The normalization of torment: Producing and managing anguish in Milgram's "obedience" laboratory. *Theory & Psychology*, *25*, 639–656. https://doi.org/10.1177/0959354315605393

Orne, M. T., & Holland, C. C. (1968). On the ecological validity of laboratory deceptions. *International Journal of Psychiatry*, *6*, 282–293.

Perry, G. (2012). *Behind the shock machine: The untold story of the notorious Milgram psychology experiments*. Scribe.

Perry, G. (2013). Deception and illusion in Milgram's accounts of the obedience experiments. *Theoretical & Applied Ethics*, *2*(2), 79–92. https://doi.org/10.1353/tha.2013.0018

Perry, G. (2018). *The lost boys. Inside Muzafer Sherif's Robber Cave experiment*. Scribe.

Perry, G., Brannigan, A., Wanner, R. A., & Stam, H. (2020). Credibility and Incredulity in Milgram's obedience experiments: A Reanalysis of an unpublished test. *Social Psychology Quarterly*, *83*(1), 88–106. https://doi.org/10.1177/0190272519861952

Postmes, T., & Branscombe, N. R. (Eds.). (2010). *Rediscovering social identity: Key readings*. Psychology Press.

Reicher, S. D., & Haslam, A. (2019). Engaged followership and engaged fellowship. In L. Newman (Ed.), *Confronting humanity at its worst: Social psychological perspectives on genocide*. Oxford University Press.

Reicher, S. D., 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. https://doi.org/10.1111/josi.12067

Reicher, S. D., & Haslam, S. A. (2011). After shock? Towards a social identity explanation of the Milgram 'obedience' studies. *British Journal of Social Psychology*, *50*, 163–169. https://doi.org/10.1111/j.2044-8309.2010.02015.x

Reicher, S. D., Haslam, S. A., & Smith, J. R. (2012). Working towards the experimenter: Reconceptualizing obedience within the Milgram paradigm as identification-based followership. *Perspectives on Psychological Science*, *7*, 315–324. https://doi.org/10.1177/1745691612448482

Ring, K., Wallston, K., & Corey, M. (1970). Mode of debriefing as a factor affecting subjective reaction to a Milgram-type obedience experiment: An ethical inquiry. *Representative Research in Social Psychology*, *1*, 67–88.

Rosenhan, D. (1968). *Obedience and rebellion. Unpublished manuscript*. Stanford University Library.

Rosenhan, D. (1969). Some origins of concern for others. In P. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology* (pp. 134–153). Holt, Rinehart & Winston.

Russel, N. J. C. (2018). *Understanding willing participants: Milgram's obedience experiments and the Holocaust. Volt I*. Palgrave Macmillan.

Russell, N. J. C. (2011). Milgram's obedience to authority experiments: Origins and early evolution. *British Journal of Social Psychology*, *50*, 140–162.

Russell, N. J. C., & Gregory, R. J. (2011). Spinning an organizational "web of obligation?" Moral choice in Stanley Milgram's "obedience" experiments. *The American Review of Public Administration*, *41*, 495–518. https://doi.org/10.1177/0275074010384129

Shanab, M. E., & Yahya, K. A. (1977). A behavioural study of obedience in children. *Journal of Personality and Social Psychology*, *35*, 530–536. https://doi.org/10.1037/0022-3514.35.7.530

Zimbardo, P. (2008). *The Lucifer effect: How good people turn evil*. Rider.