Rethinking the Nature of Cruelty: The Role of Identity Leadership in the Stanford Prison Experiment

S. Alexander Haslam
The University of Queensland

Stephen D. Reicher
University of St Andrews

Jay J. Van Bavel
New York University

The Stanford Prison Experiment (SPE) is one of the most famous studies in the history of psychology. For nearly a half century it has been understood to show that assigning people to a toxic role will, on its own, unlock the human capacity to treat others with cruelty. In contrast, principles of identity leadership argue that roles are unlikely to elicit cruelty unless leaders encourage potential perpetrators to identify with what is presented as a noble ingroup cause and to believe their actions are necessary for the advancement of that cause. Although identity leadership has been implicated in behavior ranging from electoral success to obedience to authority, researchers have hitherto had limited capacity to establish whether role conformity or identity leadership provides a better account of the cruelty observed in the SPE. Through examination of material in the SPE archive, we present comprehensive evidence that, rather than guards conforming to role of their own accord, experimenters directly encouraged them to adopt roles and act tough in a manner consistent with tenets of identity leadership. Implications for the analysis of conformity and cruelty as well as for interpretation of the SPE are discussed.

Keywords: leadership, followership, social identification, identity entrepreneurship, Stanford Prison Experiment

Supplemental materials: http://dx.doi.org/10.1037/amp0000443.supp

In August, 1971, Professor Philip Zimbardo and his research team recruited 24 young men to participate in a study on prison life. The volunteer participants were randomly assigned to be either prisoners or guards in a mock prison that had been constructed in the basement of the Stanford psychology department. After just a few days, guards began to repress the prisoners and their cruelty escalated, to the point where, after 6 days, the study was prematurely terminated. What happened during that week—and what it means for our understanding of human behavior—has been the focus of scientific and public debate for the past half century.

The Stanford Prison Experiment (SPE; Haney, Banks, & Zimbardo, 1973a, 1973b) is one of the most famous psychology studies of all time. It is covered in most introductory psychology and social psychology textbooks and courses (Bartels, Milovich, & Moussier, 2016; Griggs, 2014; Griggs & Whitehead, 2014) and is a standard point of reference for media stories on tyranny and repression. The study has also been presented to government officials and in court cases to help understand events ranging from prison riots to the abuse of detainees during the Iraq War. On top of this, it has provided the material for an extremely influential website (www.prisonexp.org), a best-selling book on

S. Alexander Haslam

evil (Zimbardo, 2007), and several feature films that have grossed over $12 million at the Box Office (Bratman & Alvarez, 2015; Scheuring et al., 2010; see also Conrad, Preuss, Wildfeuer, & Hirschbiegel, 2001). As a result, the study has had a major role to play in shaping how millions of people think both about the nature of human cruelty and about the power of the situation to encourage toxic behavior.

The SPE became famous for the striking evidence it appeared to provide that normal, well-adjusted young men could come to harm innocent civilians simply as a consequence of having been randomly assigned the role of guard (rather than prisoner). Zimbardo and his colleagues have expressed this provocative idea in a number of slightly different ways to different audiences. In their first publication, they wrote that guards made their own decisions about how to run the prison and that their aggression was "emitted simply as a 'natural' consequence of being in the uniform of a 'Guard' and asserting the power inherent in that role" (Haney et al., 1973b, p. 12). Later on, in a summary chapter, Zimbardo (2004) argued that "each subject's prior societal learning of the meaning of prisons and the behavioral scripts associated with the oppositional roles of Prisoner and Guard was the sole source of guidance" (p. 39). More recently still, in their textbook, Zimbardo, Johnson, and McCann (2012) claimed that "the mere fact of putting on uniforms was sufficient to transform [participants] into passive prisoners and aggressive guards." This point has now become canonical in the field in seeming to show that toxic situations are sufficient to stimulate toxic behavior.

Whatever the precise wording, the core of Zimbardo, Banks, Haney, and Jaffe's (1973) argument is that what the

guards did, they did "without training from [the Experimenters] in how to be Guards" (p. 40; see also pp. 65–66). In the researchers' words, "our results go one step further [than Milgram's] in removing the immediate presence of the dominant experimenter-authority figure, [and] giving the subjects-as-guards a freer range of behavioral alternatives" (Haney et al., 1973a, p. 90). Indeed, it is the fact that the study points to the power of social roles on their own to engender cruelty that has helped make the SPE so influential in the scientific literature. The implication is that good people will generally turn bad if they happen to be put in a powerful position in a toxic place.

This is the lesson that Zimbardo himself has explicitly drawn from the SPE, naming it *The Lucifer Effect*—the title of his best-selling 2007 book. It is a lesson that has been widely propagated, not only in psychology but also to students and researchers in a wide array of other disciplines (e.g., history and criminology; Browning, 1992; Jacoby, Severance, & Bruce, 2004). Indeed, alongside Milgram's (1963, 1974) classic research on obedience to authority (OtA), the SPE may have had more impact on the public consciousness than any other piece of psychological research (Banyard, 2007; Blum, 2018; Konnikova, 2015).

Problems With the Role Account

Over the 5 decades since it was conducted, the SPE has been subjected to increasing scrutiny and criticism (Griggs, 2014; Turner, 2006). Although textbook and popular accounts of the study generally make little or no mention of limitations in either the study's design or Zimbardo and colleagues' interpretation of its findings (Bartels et al., 2016; Carnahan & McFarland, 2007; Griggs, 2014; Griggs & Whitehead, 2014), there are two principal lines of criticism in the scientific literature.

First, from the sources that are available—notably, Zimbardo's (1992) film of the study, *Quiet Rage*, and his 2007 book (though, as Griggs, 2014, pointed out, these fail to include many important details of the study)—it is clear that many participants did not conform to role. Many prisoners continued to resist authority until the end of the study (Le Texier, 2018; Reicher & Haslam, 2006). Likewise, many guards refused to assert their authority. While a few were brutal and cruel, others were not, and some even sided with the prisoners (Zimbardo, 1992). Such variability requires a more nuanced interpretation of the SPE than is typically provided in media and textbook reports (Haslam & Reicher, 2007). For many participants, then, it appears that being given a role was not sufficient to elicit the type of behavior for which the study has become well known.

Second, to the extent that some guards and some prisoners did conform, there are suggestions that, rather than doing so of their own accord, the actions of the experimenters may have played an important part in producing their

Stephen D. Reicher

behavior. Early on, Banuazizi and Movahedi (1975; see also Gray, 2013) noted that the behavior of participants—especially the guards—could be explained in terms a range of salient demand characteristics that encouraged them to behave in particular ways. The importance of the experimenters' actions was subsequently demonstrated in a study by Lovibond, Mithiran, and Adams (1979) in which guards were instructed to adopt either authoritarian, democratic, or participatory approaches to their role. Here, only the authoritarian instructions produced toxic behavior similar to that seen in the SPE. Moreover, in a more recent prison study by Reicher and Haslam (2006), guards were given no direct instructions about how to behave (other than prohibiting violence, as in the SPE) and, subsequently, showed no inclination to treat prisoners cruelly.

This body of research led Haslam and Reicher (2007a, 2012a) to propose that the experimenters' leadership may have been critical to the emergence of guard cruelty in the SPE. There is some evidence to support this in the materials Zimbardo has previously made available, notably the briefing he gave to the guards during the "orientation day" before the prisoners arrived. This is included in Zimbardo's (1992) film of the SPE. During the briefing, Zimbardo announced,

> You can create in the prisoners feelings of boredom, a sense of fear to some degree, you can create a notion of arbitrariness that their life is totally controlled by us, by the system, you, me—and they'll have no privacy. They'll have no freedom of action, they can do nothing, say nothing that we do not permit. We're going to take away their individuality in various ways. In general what all this leads to is a sense of powerlessness. (Haslam & Reicher, 2007a, p. 618)

Two features of this passage are notable. The first is Zimbardo's use of the terms *we*, *us*, *they*, and *their*. In effect, he positions himself as part of the guard's ingroup, standing with them against the prisoners (for evidence of the power of such rhetoric, see Donnellon, 1996; Steffens & Haslam, 2013). As Zimbardo said in an interview in 1972, "I trained the guards and said 'Look, this is a serious situation, we're in this together and it's you and I and us and the people of California versus the prisoners'" (cited in Le Texier, 2018). In short, he created a sense of multiple overlapping identities that defined the experimenters and guards together (as people interested in science, as reformers, and as good Americans) as a united force against the prisoners. The second feature is the explicit guidance that guards are given about the general manner in which they should act—being instructed to create a sense of fear and boredom, to remove the prisoners' privacy and freedom, and so on.

In short, then, this briefing can be seen as an act of leadership on the part of Zimbardo. And although this point has never been acknowledged either by Zimbardo and his colleagues (e.g., Zimbardo, 2004, 2006, and see www.prisonexp.org) or in textbooks that describe the SPE (see Griggs, 2014), it raises the question of whether there were further acts of leadership during the study, something suggested by a number of the people who were directly involved in the study (e.g., Guard 4[1], see Mark, 2007; Ronson, 2015) and also by other scholars (Konnikova, 2015; Le Texier, 2018). It also raises the question of how important the leadership of the experimenters might have been in producing the cruelty of the guards.

Yet, interesting as these questions are, neither could ever be answered as long as the only clear evidence of leadership was the single passage cited above. Certainly, this provides a thin basis on which to argue that the experimenters' interventions were crucial to producing guard cruelty, let alone to develop an alternative theoretical account of both parties' behavior. We may have suspected that leadership was important in producing toxic behavior, but for substantive data to support this analysis, we have previously had to look elsewhere for evidence.

An Alternative Account: Identity Leadership and Engaged Followership

Elaborating on the foregoing observations, Haslam and Reicher (2007b, 2012a) suggested that what may have been going on in the SPE is a specific process referred to as *identity leadership*. This analysis is grounded in hypotheses derived from social identity theorizing (after Tajfel &

---

[1] To preserve their anonymity, guards are referred to by number. The number refers to the position of a given guard's last name in alphabetical sequence.

Jay J. Van Bavel

Turner, 1979; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987).

This analysis emerges from a tradition of social identity research (Tajfel & Turner, 1979) that proposes that human beings have the capacity to define themselves not only as individuals (e.g., "Abigail," "Barbara," "Colin") but as group members (e.g., "Americans," "Buddhists," "Cubs fans") and that the way we define ourselves is bound up with social context. In particular, following Bruner (1957), we adopt a given group membership to the extent that it allows us to make sense of our role in the situation that confronts us (i.e., so that it is fitting; Oakes, Haslam, & Turner, 2006). For this reason, Reicher and Haslam (2006) argued that people do not automatically take on roles (in ways that Haney et al., 1973a, suggested), but do so only to the extent that these roles make sense—or, rather, have been made to make sense—in the context of a salient social identity.

Once we do self-categorize ourselves and others in collective terms (Turner et al., 1987; Turner, Oakes, Haslam, & McGarty, 1994), we perceive and evaluate ourselves, our world, and our goals in terms of the social group. Thus, it is the esteem in which the group is held that defines our self-esteem (Tajfel & Turner, 1979) and indeed when the group is defined positively and is a source of pride, we are more likely to identify with it (Tyler & Blader, 2003). Equally, it is the interest of the group which defines our self-interest such that we may sacrifice our personal gain or even our personal existence for the collective cause (Swann et al., 2014).

Perhaps most importantly, it is the understanding of the group (its collective beliefs, values and norms) that guides us in pursuing this interest (Reicher, Haslam, Spears, & Reynolds, 2012; Turner & Reynolds, 2011). The significance of this is that it provides the basis for a model of social influence termed *referent informational influence* (Turner, 1982, 1991). According to this, when group identity becomes salient, individuals seek to ascertain and to conform to those understandings which define what it means to be a member of the relevant group. This then provides distinctive answers to the three key questions for any model of influence. Who is influential? Those in a position to know the group beliefs, values and norms by virtue of their being representative of the group (formally, after Rosch, 1978, those who are prototypical of the group). Who is influenced? Those who share a common salient social identity. What achieves influence? Messages seen to be consonant with group understandings.

This model of influence is also an implicit theory of leadership. For effective leaders can be understood as those who are able to (a) represent themselves as prototypical of the group, (b) make potential followers represent themselves as members of a common social group, and (c) represent their proposals as the actualization of group understandings. Put slightly differently, leaders are able to influence followers largely as a function of their capacity to represent, create and advance a sense of social identity that they share with those followers (a sense of "us-ness"). As we have put it elsewhere, effective leadership is therefore a process of social identity management (Haslam, Reicher, & Platow, 2011).

Over the last 20 years, this model of leadership has been made explicit, elaborated, and tested by a range of authors (Ellemers, de Gilder, & Haslam, 2004; Hogg, 2001; Reicher, Haslam, & Hopkins, 2005; Steffens, Haslam, & Reicher, 2014; van Dick et al., 2018; Turner & Haslam, 2001). In our own work, we provide an analysis that centers on four core insights (Haslam et al., 2011; Reicher et al., 2005). The first is that effective leaders (those who influence and harness the energies of followers) need to be seen to be representative of a shared ingroup. However, it is not enough just to be prototypical of the group. Many leaders who are ingroup members and who understand group norms and values may nonetheless act for their own ends rather than the group's. Indeed, in contemporary politics this belief is central to the rise of populist "antipolitics" (Eatwell & Goodwin, 2018). Accordingly, second, effective leaders need to be seen to be advancing ingroup interests. But what the group and its interests are is never just given. For this reason, third, leaders, need to be entrepreneurs of identity (Reicher & Hopkins, 2001) who actively construe followers as part of a common group (Reicher et al., 2005) and actively construe themselves as both prototypical and working for the group (Haslam et al., 2011). Additionally, they need to present their own messages and projects as an actualization of group beliefs values and norms. This brings

us to the fourth and final point: that because leadership is not just about how leaders act but also about their capacity to shape the actions of followers, effective leaders need to be impresarios of identity who promote policies and practices that help translate group values into material lived reality.

Support for these four propositions is provided by a large body of research in social, organizational, and political psychology (Haslam et al., 2011; Steffens, Haslam, Reicher, Platow, et al., 2014). For example, recent meta-analysis has found that leaders garner more support, exert more influence, and are seen as more charismatic the more prototypical they are of the group they lead (Barreto & Hogg, 2017, 35 studies, $r = .49$; Steffens, Munt, van Knippenberg, Platow, & Haslam, 2018, 70 studies, $r = .39$). Moreover, a recent study conducted across 21 countries shows that the above four aspects of identity leadership predict leader success better than many other theoretical constructs (notably transformational leadership and authentic leadership; van Dick et al., 2018). Principles of identity leadership thus appear to characterize and underpin effective leadership across a wide range of leaders, groups and social contexts.

But can principles of identity leadership help us better understand the culture of cruelty that emerged in the SPE? Haslam and Reicher (2007, 2012) argued that they might but, as noted above, the lack of relevant detail made this question hard to answer definitively. To test their ideas Haslam and Reicher therefore turned to other classic studies in which ordinary people were led to be cruel to others, notably Milgram's (1963, 1974) classic studies of OtA. On inspection, there are a number of aspects of these studies that were consistent with an identity leadership account. In the first instance, archival material provided clear evidence of the considerable lengths that Milgram took to encourage participants to identify with his research enterprise and to see its goals—ostensibly to improve scientific understanding of the effects of punishment on learning—as positive, worthy and a source of pride (Gibson, 2013; Haslam, Reicher, Millard, & McDonald, 2015; Russell, 2011). This identity entrepreneurship, in turn, appears to have underpinned participants' willingness to harm the learner within the OtA paradigm. In particular, correlational and experimental research shows that in both the original Milgram studies and in more recent conceptual replications, participants prove willing to follow destructive instructions only to the extent that they identify with the (male) experimenter and the scientific community that he represents (Haslam, Reicher, & Birney, 2014; Reicher, Haslam, & Smith, 2012). Indeed, Haslam and Reicher argued—and have shown experimentally (e.g., Haslam et al., 2014, 2015)—that the power of situational cues within the paradigm derives precisely from their capacity to encourage identification with the experimenter and his scientific agenda (e.g., so that participants identify more with the experimenter when he is

in the same room or from a prestigious institution, or when he states that the experiment requires that they continue, rather than ordering to them to go on).

As well as validating an identity leadership analysis of toxic behavior, these studies also point to some distinctive aspects of leadership as it applies to the cultivation of cruelty. The key issue is that, as we now know from the SPE and the BBC Prison Study (Reicher & Haslam, 2006), few people spontaneously identify with roles that require them to be cruel to others. Hence, as Haslam and Reicher (2007a, 2012a, 2017) have argued, those who advocate harming others need to work hard to do the identity entrepreneurship necessary to (a) get others to identify with them and their group cause; (b) construe the group's goals as noble, virtuous, and a source of pride; and (c) show how harmful acts are absolutely necessary for the achievement of those goals.

In the case of the OtA studies, archival research makes it clear that Milgram indeed worked hard to do precisely this (e.g., Haslam et al., 2015; Russell, 2011). Yet in the case of the SPE, Zimbardo and colleagues' public statements argued directly against the identity leadership analysis, and, as we have seen, in the absence of strong evidence to the contrary their role account has continued to serve as the dominant explanation in textbooks, movies, and the media.

New Evidence From the SPE

In 2018, the situation described above regarding the lack of evidence from SPE changed. The Zimbardo archive at Stanford University became available online and was subjected to forensic examination by Le Texier (2018). This has allowed many of the previous gaps in knowledge about what happened in the study to be filled in. Of particular relevance to the present argument, the archive presents a much richer picture of how the experimenters directly engaged with participants in the study. As a consequence, we are now in a position rigorously to examine whether an identity leadership analysis can explain what happened in the SPE.

There are at least four new pieces of evidence provided by Le Texier (2018) that point to the importance of the experimenters' leadership in the SPE in producing the cruelty of the guards as well as other related outcomes.

First, following Zimbardo's orientation day briefing, the guards had a much longer session with the "chief warden." As Le Texier (2018, p. 74) noted, this lasted 5–6 hr, but only a portion was recorded. In a key passage, the warden told them,

> I guess when you come into prison you have to sit and wait and wait and wait and so we'll have that, the waiting with the uniforms, stocking cap and all this business. The idea is to order them around, you do not request them to do anything, you tell them, and I think that all of you could probably put on some sort of a firm policemanlike voice. I

do not know but we'll probably try to role play that a little bit this afternoon. . . . If we do not do anything, then we're just playing a game, and nothing happens—we do not learn anything. So your inventiveness and whatever else—stick-to-it-ness as far as devising things, working out a schedule, and then rewards and punishments and all of this kind of thing are very important part of making the thing run. To a large extent we will either create this prison environment or not on the basis of what you do during the shift. (cited in Le Texier, 2018, pp. 80–81)

The important thing to note here is that this briefing does not merely reiterate Zimbardo's general guidelines to the guards, but provides very detailed instructions about to how to act. When one of the guards expressed concerns that "some of these sadisms" might create "some kind of bad problems" the warden gives further practical advice on what should be done if the damage to the prisoners requires medical attention: "We have access to the Cowell Health Center Facilities . . . and should we have need for that sort of thing, we can get them over [t]here very fast." This does two things: It implies that the level of abuse might rise to a level that requires medical attention, while simultaneously signaling that this should not be seen as a problem. Moreover, in line with principles of identity leadership, the warden stresses that these actions are necessary for the research team to achieve its important goals, and with his repeated references to "we," he makes it clear that he sees the guards as part of that team.

Second, it is clear that such detailed instructions were not limited to orientation sessions but continued to the end of the study. Thus, Guard 10 noted in his poststudy evaluation that "the Warden or Prof. Zimbardo specifically directed me (us) to act a certain way (ex. hard attitude Wednesday following Tuesday leniency)" (cited in Le Texier, 2019). In a similar vein, when writing to Zimbardo after the study had ended, another guard described how the warden told them to do such things as stand outside the cells in the middle of the night and blow their whistles. He concluded, "I thought that the Warden was very creative, not just then but through the experiment, he gave us very good sado-creative ideas" (Guard 3, cited in Le Texier, 2019). And another guard also reported that he gleaned the experimenters' expectations not only from what he was told to do, but also from what he was not discouraged from doing: "We would ramp up the general harassment, just sort of crank it up a bit. Nobody was telling me I shouldn't be doing this. The professor is the authority here, you know. He's the prison warden. He's not stopping me" (Guard 4, cited in Le Texier, 2018, p. 81). In this way, multiple guards independently reported direct, concrete instruction from the experimenters about how to torment the prisoners.

Third, the guards were aware of what they were expected, or indeed required, to do to ensure the goals of the experimenters were achieved. One, Guard 2, wrote,

> I consciously felt that for the experiment to be at all useful "guards" had to act something like guards. [. . .] I felt that the experiment was important and my being "guard-like" was part of finding out how people react to real oppression. (cited in Le Texier, 2019)

Again, this speaks to the ways in which the experimenters' identity leadership ensured that guards saw their cruelty as necessary for the advancement of a worthy ingroup cause.

Echoing Guard 4's statement that "it's the Professor who decides here," another (unidentified) guard said to Zimbardo and his colleagues during the debriefing that "You're the Experimenter and in a sense we're kind of like the employees" (cited in Le Texier, 2019). Yet, by the same token, another guard who was unconvinced of the study's worth, dropped out because of the requirement to act harshly: "I object to the way prisoners will be treated. As the orientation meeting went on, the way things were stressed, there will be too much harassment" (cited in Le Texier, 2018, p. 105). All guards thus inferred that the experimenters were keen to encourage harassment and cruelty. Moreover, the fact that this assumption extended to a guard who did not behave this way, suggests this is not simply a post hoc rationalization or a strategy for shifting blame.

Fourth, there is evidence that the experimenters shaped not only the guards' understandings but also the way they structured the prison regime. While the traditional account of the SPE suggests that the guards drew up the prison rules and regulations themselves, the new information suggests that they were effectively dictated by the warden's orientation session. Of the 17 rules in the SPE, 11 were copies of those that the warden had drawn up for an earlier prison study of his own and the other 6 were largely adaptations to the specific circumstances of the SPE (e.g., a rule forbidding playing with the light switches because there were no such switches in the earlier study; see Le Texier, 2018, p. 61; Le Texier, 2019). Likewise, many of the punishments were based on those that had been devised in the earlier study. It is not plausible that these similarities are mere coincidence.

In combination, these various pieces of evidence show clearly that the level of intervention by the experimenters in the SPE was greater than has been reported in the literature or media. More importantly, it is apparent that these explicit and implicit instructions had more of an impact on the guards than has been previously acknowledged. In the light of this, it is very difficult to sustain Zimbardo and colleagues' claim that guards acted cruelly entirely of their own accord or slipped naturally into cruel roles.

We can therefore now say with confidence that leadership is important to an understanding of cruelty in the SPE. Nevertheless, while there are suggestions that this took the

specific form of identity leadership—for instance, in guards' statements that they felt that they were acting in pursuit of a worthwhile cause—the evidence we have presented thus far is hardly definitive. Not least, this is because the material discussed above does not provide an opportunity for in-depth analysis of the experimenters' leadership in action.

There is, however, one further piece of evidence in the archive which is particularly pertinent to this issue. This is the recording of a formal meeting early in the SPE between the warden and Guard 7. Two features of this meeting make it particularly relevant to the key question of how the experimenters sought to produce guard toxicity. The first is that Guard 7 had been reluctant to act repressively (as he confirmed to us in a telephone interview on June 13, 2018) and so the whole meeting centered on the warden trying to get him to adopt the role of a tough guard. The second is that a recording of the entire interchange is available (whereas this is true of less than 10% of the entire study; Le Texier, 2019) and therefore can be analyzed in its entirety. The recording is held in the Department of Special Collections and University Archives in Stanford University Libraries (Source ID: SC0750_s5_b2_21). It can be accessed directly online at http://purl.stanford.edu/wn708sg0050 and a full transcript is provided at https://osf.io/8dpqz. This allows students and researchers to review the transcript themselves and determine the degree to which it supports our identity leadership analysis rather than Zimbardo's role account.[2]

This interview provides a new and unprecedented opportunity to investigate a series of questions concerning not only the degree of leadership in the SPE but also the form taken by such leadership. To start with, did the warden accept the position taken by Guard 7 or did he actively attempt to change his stance? That is, to paraphrase Zimbardo's (2004, p. 39) outline of the role conformity account, did the warden encourage the guard to "rely on his own prior societal learning of the meaning of prisons and the behavioral scripts associated with the oppositional roles of prisoner and Guard as [his] sole source of guidance"? Or did he instead employ the signature characteristics of the identity leadership account?

More formally, we can break the identity leadership account down into the three elements that we outlined above and then gauge the presence of each of them in the text. The first involves seeking to establish a common cause and a common group membership that links the experimenters and guards. The second involves presenting "tough" or cruel behavior as necessary to advance this shared cause. The third involves characterizing the group cause as worthy and noble to justify the toxic behavior that advances it. If these elements are present, this would provide strong novel evidence that identity leadership had a role to play in efforts to encourage the guards to behave cruelly.

The Meeting Between the Warden and Guard 7

On Leadership

The warden's meeting with Guard 7 is replete with evidence that the guard was not conforming (blindly or otherwise) to the role he had been assigned and that, rather than simply accept this, the warden sought to get him to change his mind. Faced with the guard's admission that "I'm not too tough" [line 14] and that "if it was just entirely up to me, I wouldn't do anything. I would just let it cool off" [line 38], the warden made it clear at multiple points in the meeting that the experimenters expect (and require) him both to be more "involved" [lines 12, 80] in proceedings and to be more "tough" [lines 13, 17, 244]. Indeed, the meeting started with (and was presumably motivated by) the observation that "We noticed this morning that you weren't really lending a hand . . . but we really want to get you active and involved because the guards have to know that every Guard is going to be what we call a tough Guard" [lines 1–2, 11–13].

In urging the guard to be more tough, it is also apparent that the warden contextualized this toughness as an aspect of the role that the guard was being asked to play to ensure the experiment is a success. That is, toughness was not valorized as a positive attribute in and of itself. Rather, it was valorized as an important dimension of an identity that the warden wanted the guard to adopt. Importantly too, the warden emphasized that "every Guard" was going to play this same role—underscoring the experimenters' expectations about the norms for the entire group and mirroring what both the warden himself and Zimbardo had said when briefing the guards as a whole.

Again, though, Guard 7 was reluctant to take on this identity and the attributes associated with it. Far from passively or naturally adopting the role of tough guard, he actively and repeatedly resisted the pressure from the warden. This is the exact opposite of what the role account would predict. In the face of this resistance, the warden repeated and intensified his demands for conformity:

> We'd like you to try a little bit more, to get into the action . . . instead of sitting in the background, if you can get in and start doing something yourself, get involved [lines 77–80] . . . It's your *job* to make sure these things [episodes of Prisoner

---

[2] Before reading our own analysis of these issues, we recommend that readers listen to the interview and read the transcript for themselves. Given the necessity of applying open science principles to what has previously been a closed debate, we think it important that readers judge independently how the original materials speak to the respective merits of different conceptual accounts of the toxic behavior that unfolded in the SPE and evaluate for themselves the adequacy of our identity leadership account. For reasons of openness and transparency, we also illustrate our points with verbatim extracts from the meeting accompanied with line numbers related to the meeting transcript in the online supplemental materials. This allows readers to check that our use of these extracts is accurate and appropriate.

revolt] do not happen, to the best of your ability. [lines 171–172]

Finally, toward the end of the meeting when the guard is still showing no willingness to fall into line, the warden urged him to "forget some of the more sophisticated psychology that you might know" [lines 245–246], to forgo his "individual style" [line 253], and instead to embrace the role of "the stereotype Guard" [line 252]. In this way, the warden emphasized the importance of the group and encouraged the guard to self-categorize, and behave, as a member of the group. Importantly, rather than allowing the guard to define the role for himself, the warden defined it for him, and did so in explicitly stereotypic terms. Here again, though, the guard resisted by refusing to embrace the categorical "us–them" identity that the warden invoked to structure his understanding of the situation. Indeed, the guard actively contested the stereotypic self-definition that the warden proposed:

> Well I've met a lot of police that, er, that act a whole lot of different ways. You cannot do it just by the movies or something. 'Cause you know I've met plenty of police. [lines 248–250]

Far from naturally adopting the role or accepting the stereotypic role, the Guard thus made an explicit case for rejecting them. It is thus with more than a hint of exasperation that the Warden concluded the meeting with explicit and concrete instructions about the behavior expected of the Guard:

> When there's a situation . . . [you have] to have to go in there and shout if necessary. To be more *into* the action. [line 257–258]

In sum, two conclusions can be drawn from this evidence. First, the guard was certainly not left to his own devices in deciding what to do within the prison. Indeed, he was assertively pressured by the prison leadership to conform to group-based expectations. Second, any theoretical analysis of guard behavior must include the active interventions of the experimenters in seeking to ensure conformity to a brutal role. In short, there was clearly sustained leadership going on. We now focus on the forms that this took with a view to understanding the strategies through which the warden sought to shape the guard's behavior.

On Identity Leadership

The transcript of the meeting between the warden and Guard 7 provides striking evidence that the warden encouraged the guard to discard his personal identity, to adopt a collective identity, and to embody stereotypic expectations associated with that collective identity. In this respect, as with Zimbardo's briefing of the guards that we discussed earlier, one of the more striking features of the warden's

discourse is his repeated reference to the collective "we." Indeed, the warden used the first-person plural pronouns *we*, *our*, and *us* 57 times in this single meeting, or once every 30 words. Previous research has found that this use of collective pronouns is associated with effective leadership. For example, across the 34 Australian elections that have been held since federation in 1901, winning candidates use these collective pronouns once every 79 words while losing candidates use them once every 136 words (Steffens & Haslam, 2013). Thus while politicians are often seen to exemplify principles of identity leadership (e.g., Augoustinos & De Garis, 2012; Gleibs, Hendricks, & Kurz, 2018; Reicher & Hopkins, 2001), the warden appears to have enacted this aspect of the process of cultivating a sense of shared identity more vigorously than even the most successful political leaders.

In this regard, it is clear that the guard was encouraged to see himself, along with the experimenter, as part of this collective "we" (i.e., "us the prison authorities"). In particular, the warden highlighted their shared disdain for the correctional system, for example, by stressing that "We happen to agree with you that basically it's rotten" [lines 200–201]. He also repeatedly portrayed the guard and the experimenters as having the same interests and goals: "These things [prisons and mental hospitals] are all over the place. And we want to know about them" [lines 83–85]. In this way, the warden appealed to a shared goal outside the experiment to justify cruelty within it.

Throughout the interaction, the warden thus went to great lengths to represent the guard and the experimenters as being "on the same side" while also sharing common enemies—notably those who defend the prevailing criminal justice system and "the pigs" [line 246]. He also encouraged the guard to acknowledge and embrace their shared identity and at one point accomplished this through a four-step rhetorical maneuver where, first, he stated his own values; second, he stated those of the group as a whole; third, he made a statement about what he saw the guard's values to be; and fourth, he asked the guard to confirm their alignment: "I'm very deeply committed to that [rehabilitation]. And I think all of us are. . . . I think you feel the same way. Is that true?" [lines 204–209]. In line with identity leadership, the construction of shared identity based on a common cause can thus be seen to lie at the core of the warden's intervention to shape the guard's behavior.

The warden also tried to convince Guard 7 that what he was asking him to do—that is, be tough—was essential to the success of their mutual cause. More specifically, in the face of the guard's refusal to act tough, the warden repeatedly reminded him that in doing so, he was putting the research enterprise as a whole at risk. Guard toughness, the warden explained, is "really important for the workings of the Experiment" [lines 19–20] because "whether or not we can make this thing seem like a prison, which is the aim of

the thing, depends largely on the guards' behavior" [lines 22–24]. This is a point that the warden underlined toward the end of the meeting when he noted:

> I think you could be a better guard. Er, better in the sense of, you know, the guards have more responsibility at this point for making the thing work. That's, that's the real clincher right here. That, er, that if the guards fall apart, the experiment falls apart. The prisoners do not have that power in this study. [lines 157–158]

In making these observations, the warden attuned Guard 7 to two interrelated considerations. First, that while it may be unpleasant and not what the guard would otherwise want to do, his willingness to be tough was essential for the success of the project. Second, that if he failed to rise to this challenge, the guard was effectively letting the team down. The dual implication, then, is that the guard's reluctance to conform was not only jeopardizing the group's collective mission but also selfish. At the same time, tough behavior was framed as something that was not only essential to the advancement of shared group goals but also expected of any committed group member.

The warden repeatedly pointed out that the experiment would fail if the guards were not sufficiently tough, and he repeatedly underscored—and asked the guard to recognize—the profound and progressive significance of the experiment for both science and society. Thus, he asked, "Do you understand the rationale behind doing something like this? The importance of it?" [lines 186–187]. He then explained in detail how the design of the study would allow the researchers to make important scientific claims, because "There isn't any prison in this country is going to let you set up, you know, observational measures 24 hours a day. So that's why we have to do it here" [lines 56–58]. He further explained how important the study was in terms of societal impact: "hopefully what will come out of this study is some very serious recommendations for reform, at least reform, if not, you know, revolutionary-type reform" [lines 49–53]. In this way, rather than being pernicious, cruelty was represented as an essential driver of social progress.

Critically, this impact was also characterized as a social good with reference to the progressive identity that the warden and guard ostensibly shared. This aligned the guard with the shared long-term goals of the research. At the same time, the warden made it clear that the worthy goal of exposing penal pathology and facilitating penal reform could only be achieved through cruel behavior on the part of the guard:

> What we want to do is be able to study the thing that exists, or as nearly as we can make it to what exists and to be able to go to the world with what we've done and say "Now look, this

is what happens when you have guards who behave this way" . . . But in order to say that we have to have guards who behave that way. [lines 211–215]

In setting out the case for guard cruelty in this way, the warden also explicitly rejected the suggestion that such behavior might reflect the fact that guards are naturally cruel: "We're not trying to do this just because we're sadists" [lines 53–54]. Indeed, the worthiness of the cause was used as an explicit defense against such an accusation:

> If you need an excuse, and I think most of us do really, it is so we can learn what happens in a total institution . . . And we want to know about them. So that we can, we can get on the media and, um, and, and into press with it. And, and, and say "Now look at what, what this really about. [lines 80–86]

Overall, then, the arguments of the warden served to validate cruelty in the study in the name of challenging cruelty in society—only if the guards were toxic could toxicity be exposed and eliminated. What would otherwise be seen as wicked behavior was thereby recast as worthy action performed in the service of a greater good. This helped assign positive and distinct value to the ingroup, and it also facilitated identification with that ingroup. It also made oppression of the prisoners consonant with group norms and served to offset—and inoculate against—any doubts or hesitations that Guard 7 might had about acting brutally. In short, this new evidence displays all the hallmarks of identity leadership.

The Need to Rethink the Nature of Cruelty in the SPE

The meeting between warden and guard provides clear evidence that guards in the SPE were not left to their own devices when it came to making decisions about how to behave and run the prison. On the contrary, they were subjected to active leadership from the experimenters. It follows that any account of the SPE which fails to highlight the leadership of the experimenters and their concerted attempts to make the guards act in role is both partial and misleading.

As noted above, numerous guards independently reported that they believed the experimenters played an active role in guiding their behavior. Beyond this, though, it is apparent from various pieces of evidence—notably, the transcript of the meeting between the warden and Guard 7 and the warden's 5-hr briefing of all the guards—that this leadership took a very specific form. In particular, we see that the warden's efforts to encourage the guard to conform to stereotypic role requirements centered on strategies of identity leadership. That is, he sought to influence the guard through appeals to a sense of shared identity that promoted "toughness" as an ingroup-defining attribute necessary to (a) run the prison properly, (b) advance science, and thereby

(c) achieve the valued goal of exposing the toxicity of the American penal system. Indeed, by aligning "toughness" with the goals of multiple identities, the warden can be seen to have increased the likelihood of it being seen as an appropriate way to behave (in ways suggested by identity complexity theory; Roccas & Brewer, 2002).

It should be stressed that these identity appeals were not just a subset of themes that were mingled in with others in the meeting. Instead, the meeting can be seen as one long series of identity appeals. Time and again, the guard insisted on his unsuitability for his role and his unwillingness to be tough. Time and again, the warden responded by explaining how important it was that he act like a tough guard to help him, Zimbardo and others in their noble enterprise. Remove this interplay between the guard's role rejection and the warden's identity entrepreneurship, and little would be left.

To summarize, then, the new evidence from the SPE archive sustains three theoretical claims. First, the traditional notion that guards became cruel of their own accord is very hard (if not impossible) to sustain. Second, the experimenters' leadership was a central feature of the study. Third, and more specifically, we see that the experimenters engaged in identity leadership in an effort to encourage guard cruelty. Although support for these conclusions emerges from the guards' reports and the transcript of the meeting between the warden and Guard 7, it is also supported by a wealth of other new material (see Le Texier, 2018). In particular, full transcripts of the guard briefings indicate that the experimenters went to considerable trouble to create a sense of shared identity with all guards (not only one), to persuade them that they were coproducers of important scientific knowledge, and to indicate in some detail exactly how they were expected to behave. What happened to Guard 7 was thus not exceptional, but simply reflects the fact that he was failing to behave in ways expected (and demanded) by the experimenters.

At the same time that we make these claims, it is important to be clear about what we are not claiming. First, we do not suggest that identity leadership always produces consent or that it did so in the case of Guard 7. Nevertheless, the efforts of the warden did have some effect on him. In particular, this is because Guard 7 recollected being fully aware that his behavior was at odds with the experimenters' goals and that toughness was presented, and affirmed, by the experimenters as a core norm for the guard group. He therefore recalled being poorly positioned to challenge the cruelty of others—not least because, although he himself was not cruel in his dealings with the prisoners, he was sidelined from the shift in which most of the toxic behaviors reported in the SPE took place. Thus, identity leadership facilitated guard cruelty even if Guard 7 did not display this behavior himself.

Second, even though it is clear that the experimenters were fairly explicit about sanctioning extreme behavior, and at times provided explicit guidance as to what forms this should take, we are not suggesting that the guards had no autonomy and were simply following a script. We agree with Zimbardo and colleagues that the guards clearly improvised and were creative in what they did and how far they went. However, this enriches rather than undermines the identity leadership analysis. For as Haslam and Reicher (2012b) have argued, the guards' creativity can be seen to reflect the fact that effective identity leadership does not produce passive conformity so much as *engaged followership*. That is, when followers identify with a leader and his or her shared cause, they ask themselves what it is that that leader wants them to do and then strive to interpret the instructions they have been given enthusiastically and creatively. This is a process that the World War II historian Ian Kershaw has referred to as "working towards the Führer" (Kershaw, 1993; see also Sofsky, 1993). Here, then, we would suggest that those guards who were inspired by the identity leadership of Zimbardo and his colleagues would have been "working towards the experimenter," in ways that translated an appeal to be tough into a willingness to be cruel and oppressive (Haslam & Reicher, 2007b).

A critical point here is that the identity leadership analysis we have presented—and which is represented schematically in Figure 1—does not simply replace one form of automaticity with another (moving from "people automatically take on the roles they are thrust into" to "people automatically take on the identities that are thrust upon them"). People are well able to resist the categories and identities proposed to them (as shown above in Guard 7's responses to the warden), especially if these are at odds with other identities that are important in their lives (as shown in our previous critiques of Zimbardo's work; see Haslam & Reicher, 2012b; Reicher & Haslam, 2006, 2013). For this and other reasons, identity leadership may elicit very different behavior from different individuals, and invoking cruelty as a group norm will not always lead people to embrace it. This will be especially true if leaders fail—as the warden did with Guard 7—to persuade would-be followers that cruelty is a normative aspect of an identity that they value.

Can we conclude, then, that even if identity leadership failed to "turn" Guard 7, it was responsible for such guard cruelty as did occur? Again, to make such a claim is beyond the scope of the current evidence. The type of evidence we have from the SPE simply does not allow us to make any definitive causal claims of this form. In particular, the small sample and lack of experimental control mean we cannot be sure that these guards would not have been cruel in any case (e.g., for reasons suggested by Carnahan & McFarland, 2007), or that their cruelty was not produced by some other feature of the study besides the leadership of the experimenters. Nevertheless, what we can say with confidence is that (a) the experimenters undoubtedly did try to exercise identity leadership, (b) several guards independently re-
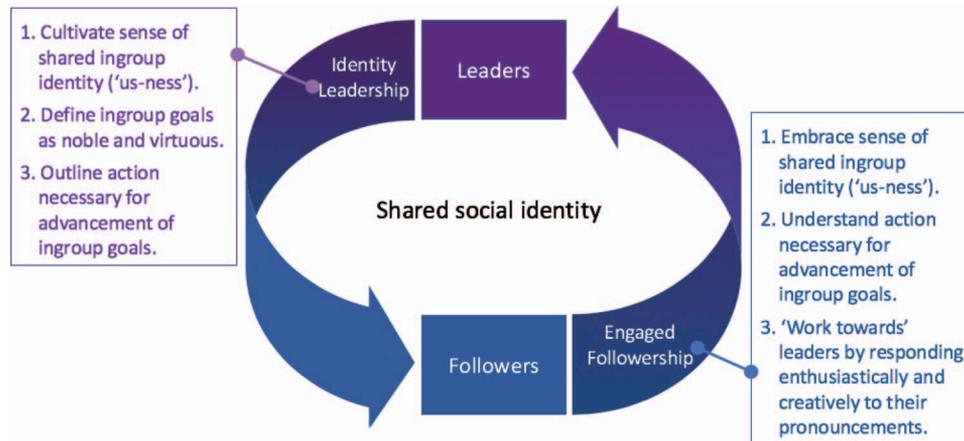
Figure 1.   Schematic representation of the relationship between identity leadership and engaged followership. See the online article for the color version of this figure.

ported being aware of the experimenters' efforts to engage in identity leadership, and (c) identity leadership therefore constitutes a plausible framework for explaining guard cruelty in the SPE. At minimum, then, the results of our analysis are plainly more consistent with an identity leadership account than they are with the standard role account. We would add too that the patterns we have identified above accord with a very large body of research which shows identity leadership to be a critical component of effective leadership (both toxic and benign) in the world at large (e.g., see Steffens et al., 2014; van Dick et al., 2018).

In relation to these various points, the significance of the Stanford archive is how closely it supports theoretical claims that were first made over a decade ago (but which could not be substantiated)—claims that (having now been substantiated) provide an important platform for revising our understanding of the important issues that the SPE addresses. In this regard, it is also clear that there are a great many ways in which the material in the archive might not have substantiated our analysis. Most obviously, this would have been the case if it had revealed no evidence of the experimenters' identity leadership or if the behavior of the participants had mapped closely on to Zimbardo and colleague's role account (e.g., if the experimenters had provided the guards with no guidance about how to interpret their role). Indeed, the falsifiability of our analysis is apparent from the fact that, hitherto, Zimbardo (2006) has dismissed our evidence-based critiques of the SPE as fundamentally wrong-headed and "scientifically irresponsible" (p. 47).

Identity Leadership and Cruelty

The plausibility of our position is enhanced by two further considerations. First, as well as being applied to the study of leadership in general (e.g., Haslam et al., 2011; van Dick et

al., 2018), the identity leadership analysis has successfully been used to explain toxic behavior in other classic studies (notably, Milgram's OtA research). Indeed, once one acknowledges the role of leadership in both Milgram's paradigm and Zimbardo's, the major distinction between the two disappears (Haslam & Reicher, 2012a). Nevertheless, the value of this corroboration lies the fact that the OtA studies were more carefully controlled than the SPE, used larger samples, and have been independently replicated (Blass, 2004; Haslam & Reicher, 2017; Reicher, Haslam, & Miller, 2014). This advances our claim that the identity leadership framework is useful not only for understanding behavior in the SPE, but also for understanding toxic (and nontoxic) behavior in a broad range of experimental contexts.

Second, looking beyond the laboratory to the wider world, it is also clear that identity leadership is a common feature in episodes of human toxicity and brutality, perhaps especially in its most extreme forms (Koonz, 2003; Muller-Hill, 1988; Sofsky, 1993; Vetlesen, 2005). A recurrent observation is that leaders work hard not only to create a common identity with would-be perpetrators, but to convince them that cruelty to outgroups (e.g., Muslims, migrants, dissidents) is necessary for the protection and advancement of the ingroup (e.g., keeping Serbia strong, America great, Turkey safe). However, if we were to select just one example to exemplify these dynamics of human inhumanity, we would choose Himmler's infamous Poznan speech of October 6, 1943, in which he rallied SS officers to persevere with the challenges of exterminating Jews in occupied Poland. The substance of this is exemplified by the following passage:

> It is one of those things that is easily said: "The Jewish people is being exterminated," every party member will tell you . . . [But] none of them has seen it happen, not one has had to go

through with it. Most of you men know what it is like to see 100 corpses lie side by side, or 500 or 1,000. To have stood fast through all this and . . . at the same time to have remained a decent person . . . has made us hard. This is an unwritten and never-to-be-written page of glory in our history. All in all, however, we can say that we have carried out this most difficult of tasks in a spirit of love for our people. (Grobmes & Landen, 1983, pp. 454–455; cited by Haslam et al., 2015, p. 78)

By invoking this example, we are not suggesting that the toxicity observed in Milgram's studies or in the SPE was in any way comparable with that of the Holocaust (Miller, 2004). What we believe is striking, though, is the consonance in the processes through which this was encouraged in these different settings. Framing cruelty as essential for the achievement of noble collective goals thus appears to be a critical strategy for mobilizing people to hate and harm others in theaters of conflict both small and large.

It follows from all this that just as cruelty does not inhere simply in the nature of the perpetrators, neither does it inhere only the demands of the situation. An understanding of how it is produced additionally requires an analysis of leadership, of how leaders persuade, and of how they are able to portray toxic behavior as worthy action in defense of a noble group cause. Central to this endeavor are leaders' efforts to construct a sense of shared identity that encompasses both the source and the target of persuasion. Indeed, accounts which seek to naturalize cruelty and harm-doing as an inevitable outcome of human behavior serve to help leaders avoid accountability for the part they have played and so, not surprisingly, are often invoked by repressive leaders in their own defense (Cesarani, 2004). A case in point involves Radovan Karadzic, who was indicted before the International Criminal Court for the former Yugoslavia for his war crimes. His defense centered on the claim that the hatred between groups and the violence that was perpetrated "arose of its own accord, from the bottom up and cannot be attributed to Franco Trujman" (cited in Elcheroth & Reicher, 2017, p. 29).

Conclusion

The new evidence that we have examined in this article makes it clear that the specific narrative of the SPE and our broader understanding of cruelty in the world both need to be rewritten. The totality of evidence indicates that, far from slipping naturally into their assigned roles, some of Zimbardo's guards actively resisted. They were consequently subjected to intense interventions from the experimenters. These sought to persuade them to conform to group norms for the purpose of achieving a shared and admirable group goal. Where previously we only had inklings that the be-

havior of guards in the SPE could have been produced in response to forces of identity leadership, we now have sufficiently strong and clear evidence of this as to need to place it on the scientific record. This in turn means that we must, of necessity, focus on the role of these forces in spawning cruelty and repression more generally, both in our studies and in the world beyond.

We hope that future presentations of the SPE—whether in the scientific literature, in classrooms, or in boardrooms—will now tell a richer story of what happened in the basement of the Stanford Psychology Department during the summer of 1971. If that happens, then the SPE can continue to help us achieve a richer appreciation of the dark social processes that blight the human condition.

References

Augoustinos, M., & De Garis, S. (2012). "Too black or not black enough": Social identity complexity in the political rhetoric of Barack Obama. *European Journal of Social Psychology, 42,* 564–577. http://dx.doi.org/10.1002/ejsp.1868

Banuazizi, A., & Movahedi, S. (1975). Interpersonal dynamics in a simulated prison: A methodological analysis. *American Psychologist, 30,* 152–160. http://dx.doi.org/10.1037/h0076835

Banyard, P. (2007). Tyranny and the tyrant. *The Psychologist, 20,* 494–495.

Barreto, N. B., & Hogg, M. A. (2017). Evaluation of and support for group prototypical leaders: A meta-analysis of twenty years of empirical research. *Social Influence, 12,* 41–55. http://dx.doi.org/10.1080/15534510.2017.1316771

Bartels, J. M., Milovich, M. M., & Moussier, S. (2016). Coverage of the Stanford prison experiment in introductory psychology courses: A survey of introductory psychology instructors. *Teaching of Psychology, 43,* 136–141. http://dx.doi.org/10.1177/0098628316636290

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

Blum, B. (2018). The lifespan of a lie. *Medium.* Retrieved from https://medium.com/s/trustissues/the-lifespan-of-a-lie-d869212b1f62

Bratman, L. (Producer), & Alvarez, K. P. (Director). (2015). *The Stanford Prison Experiment* [Motion picture]. United States: IFC Films.

Browning, C. (1992). *Ordinary men: Reserve Police Batallion 101 and the final solution in Poland.* London, United Kingdom: Penguin Books.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64,* 123–152. http://dx.doi.org/10.1037/h0043805

Carnahan, T., & McFarland, S. (2007). Revisiting the Stanford prison experiment: Could participant self-selection have led to the cruelty? *Personality and Social Psychology Bulletin, 33,* 603–614. http://dx.doi.org/10.1177/0146167206292689

Cesarani, D. (2004). *Eichmann: His life and crimes.* London, United Kingdom: Heinemann.

Conrad, M., Preuss, N., Wildfeuer, F. (Producers), & Hirschbiegel, O. (Director). (2001). *Das experiment* [Motion picture]. Germany: Senator Film.

Donnellon, A. (1996). *Team-talk: The power of language in team dynamics.* Boston, MA: Harvard University Business School Press.

Eatwell, R., & Goodwin, M. (2018). *National populism: The revolt against liberal democracy.* London, United Kingdom: Penguin.

Elcheroth, G., & Reicher, S. D. (2017). *Identity, violence and power: Mobilising hatred, demobilising dissent.* London, United Kingdom: Palgrave. http://dx.doi.org/10.1057/978-1-137-31728-5

Ellemers, N., De Gilder, D., & Haslam, S. A. (2004). Motivating individuals and groups at work: A social identity perspective on leadership and

group performance. *Academy of Management Review, 29,* 459–478. http://dx.doi.org/10.5465/amr.2004.13670967

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

Gleibs, I. H., Hendricks, K., & Kurz, T. (2018). Identity mediators: Leadership and identity construction in campaign speeches of American Presidential candidates' spouses. *Political Psychology, 39,* 939–956. http://dx.doi.org/10.1111/pops.12448

Gray, P. (2013). Why Zimbardo's prison experiment isn't in my textbook. *Psychology Today.* Retrieved from https://www.psychologytoday.com/au/blog/freedom-learn/201310/why-zimbardo-s-prison-experiment-isn-t-in-my-textbook

Griggs, R. A. (2014). Coverage of the Stanford Prison Experiment in introductory psychology textbooks. *Teaching of Psychology, 41,* 195–203. http://dx.doi.org/10.1177/0098628314537968

Griggs, R. A., & Whitehead, G. I., III. (2014). Coverage of the Stanford prison experiment in introductory social psychology textbooks. *Teaching of Psychology, 41,* 318–324. http://dx.doi.org/10.1177/0098628314549703

Grobmes, A., & Landen, D. (1983). *Genocide: Critical issues of the Holocaust.* West Orange, NJ: Berman House.

Haney, C., Banks, C., & Zimbardo, P. G. (1973a). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology, 1,* 69–97.

Haney, C., Banks, C., & Zimbardo, P. G. (1973b). *Naval research reviews: A study of prisoners and guards in a simulated prison.* Washington, DC: Office of Naval Research.

Haslam, S. A., & Reicher, S. (2007a). Beyond the banality of evil: Three dynamics of an interactionist social psychology of tyranny. *Personality and Social Psychology Bulletin, 33,* 615–622. http://dx.doi.org/10.1177/0146167206298570

Haslam, S. A., & Reicher, S. D. (2007b). Identity entrepreneurship and the consequences of identity failure: The dynamics of leadership in the BBC Prison Study. *Social Psychology Quarterly, 70,* 125–147. http://dx.doi.org/10.1177/019027250707000204

Haslam, S. A., & Reicher, S. D. (2012a). Contesting the "Nature" Of Conformity: What Milgram and Zimbardo's studies really show. *PLoS Biology, 10,* e1001426. http://dx.doi.org/10.1371/journal.pbio.1001426

Haslam, S. A., & Reicher, S. D. (2012b). When prisoners take over the prison: A social psychology of resistance. *Personality and Social Psychology Review, 16,* 154–179. http://dx.doi.org/10.1177/1088868311419864

Haslam, S. A., & Reicher, S. D. (2017). 50 years of "obedience to authority": From blind obedience to engaged followership. *Annual Review of Law and Social Science, 13,* 59–78. http://dx.doi.org/10.1146/annurev-lawsocsci-110316-113710

Haslam, S. A., Reicher, S. D., & 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,* 473–488. http://dx.doi.org/10.1111/josi.12072

Haslam, S. A., Reicher, S. D., 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" experiments. *British Journal of Social Psychology, 54,* 55–83. http://dx.doi.org/10.1111/bjso.12074

Haslam, S. A., Reicher, S. D., & Platow, M. J. (2011). *The new psychology of leadership: Identity, influence and power.* New York, NY: Psychology Press.

Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review, 5,* 184–200. http://dx.doi.org/10.1207/S15327957PSPR0503_1

Jacoby, J. E., Severance, T. A., & Bruce, A. S. (Eds.). (2004). *Classics of criminology.* Long Grove, IL: Waveland Press.

Kershaw, I. (1993). Working towards the Führer: Reflections on the nature of the Hitler dictatorship. *Contemporary European History, 2,* 103–108. http://dx.doi.org/10.1017/S0960777300000382

Konnikova, M. (2015, June 12). The real lesson of the Stanford prison experiment. *The New Yorker.* Retrieved from https://www.newyorker.com/science/maria-konnikova/the-real-lesson-of-the-stanford-prison-experiment

Koonz, K. (2003). *The Nazi conscience.* New Haven, CT: Harvard University Press.

Le Texier, T. (2018). *Histoire d'un mensonge: Enquête sur l'experience de Stanford.* Paris, France: Editions la Découverte.

Le Texier, T. (2019). Debunking the Stanford Prison Experiment. *American Psychologist, 74,* 823–829. http://dx.doi.org/10.1037/amp0000401

Lovibond, S. H., Mithiran, X., & Adams, W. G. (1979). The effects of three experimental prison environments on the behaviour of non-convict volunteer subjects. *Australian Psychologist, 14,* 273–287. http://dx.doi.org/10.1080/00050067908254355

Mark, J. (2007). Letter to the editor. *Stanford Daily (June/July)* Retrieved from https://alumni.stanford.edu/get/page/magazine/article/?article_id=32561

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

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York, NY: Harper and Row.

Miller, A. G. (2004). What can the Milgram obedience experiments tell us about the Holocaust? Generalizing from the social psychology laboratory. In A. Miller (Ed.), *The social psychology of good and evil* (pp. 193–239). New York, NY: Guilford Press.

Muller-Hill, B. (1988). *Murderous science.* Oxford, United Kingdom: Oxford University Press.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality.* Oxford, United Kingdom: Blackwell.

Reicher, S., & Haslam, S. A. (2006). Rethinking the psychology of tyranny: The BBC prison study. *British Journal of Social Psychology, 45,* 1–40. http://dx.doi.org/10.1348/014466605X48998

Reicher, S. D., & Haslam, S. A. (2013). Obedience and tyranny in psychology and history. In A. Golec & A. Cichocka (Eds.), *Social psychology of social problems* (pp. 172–195). Houndmills, United Kingdom: Palgrave Macmillan. http://dx.doi.org/10.1007/978-1-137-27222-5_7

Reicher, S. D., Haslam, S. A., & Hopkins, N. (2005). Social identity and the dynamics of leadership: Leaders and followers as collaborative agents in the transformation of social reality. *The Leadership Quarterly, 16,* 547–568. http://dx.doi.org/10.1016/j.leaqua.2005.06.007

Reicher, S. D., Haslam, S. A., & Miller, A. G. (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,* 393–408. http://dx.doi.org/10.1111/josi.12067

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

Reicher, S. D., Haslam, S. A., Spears, R., & Reynolds, K. J. (2012). A social mind: The context of John Turner's work and its influence. *European Review of Social Psychology, 23,* 344–385. http://dx.doi.org/10.1080/10463283.2012.745672

Reicher, S. D., & Hopkins, N. P. (2001). *Self and nation.* London, United Kingdom: Sage.

Roccas, S., & Brewer, M. B. (2002). Social identity complexity. *Personality and Social Psychology Review, 6,* 88–106. http://dx.doi.org/10.1207/S15327957PSPR0602_01

Ronson, J. (2015). *So you've been publicly shamed*. New York, NY: Riverhead Books.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Russell, N. J. C. (2011). Milgram's obedience to authority experiments: Origins and early evolution. *British Journal of Social Psychology, 50,* 146–162. http://dx.doi.org/10.1348/014466610x492205

Scheuring, P., Adelstein, M., Milio, J. B., Johnson, B., Nemes, S., Parouse, D. (Producers), & Scheuring, P. (Director). (2010). *The experiment* [Motion picture]. United States: Stage 6 Films.

Sofsky, W. (1993). *The order of terror: The concentration camp*. Princeton, NJ: Princeton University Press.

Steffens, N. K., & Haslam, S. A. (2013). Power through "us": Leaders' use of we-referencing language predicts election victory. *PLoS ONE, 8,* e77952. http://dx.doi.org/10.1371/journal.pone.0077952

Steffens, N. K., Haslam, S. A., & Reicher, S. D. (2014). Up close and personal: Evidence that shared social identity is a basis for the "special" relationship that binds followers to leaders. *The Leadership Quarterly, 25,* 296–313. http://dx.doi.org/10.1016/j.leaqua.2013.08.008

Steffens, N. K., Haslam, S. A., Reicher, S. D., Platow, M. J., Fransen, K., Yang, J., . . . Boen, F. (2014). Leadership as social identity management: Introducing the Identity Leadership Inventory (ILI) to assess and validate a four-dimensional model. *The Leadership Quarterly, 25,* 1001–1024. http://dx.doi.org/10.1016/j.leaqua.2014.05.002

Steffens, N. K., Munt, K., van Knippenberg, D., Platow, M. J., & Haslam, S. A. (2018). *A meta-analytic review of leader group prototypicality and leadership effectiveness*. Unpublished manuscript, School of Psychology, University of Queensland, St. Lucia, Australia.

Swann, W. B., Buhrmester, M. D., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., . . . Zhang, A. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology, 106,* 912–926. http://dx.doi.org/10.1037/a0036089

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Turner, J. C. (1982). Towards a cognitive redefinition of the social group. In H. Tajfel (Ed.), *Social identity and intergroup relations* (pp. 15–40). Cambridge, United Kingdom: Cambridge University Press.

Turner, J. C. (1991). *Social influence*. Milton Keynes, United Kingdom: Open University Press.

Turner, J. C. (2006). Tyranny, freedom and social structure: Escaping our theoretical prisons. *British Journal of Social Psychology, 45,* 41–46. http://dx.doi.org/10.1348/014466605X79840

Turner, J. C., & Haslam, S. A. (2001). Social identity, organizations and leadership. In M. E. Turner (Ed.), *Groups at work: Advances in theory and research* (pp. 25–65). Hillsdale, NJ: Erlbaum.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, United Kingdom: Blackwell.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. A. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20,* 454–463. http://dx.doi.org/10.1177/0146167294205002

Turner, J. C., & Reynolds, K. J. (2011). Self-categorisation theory. In P. VanLange, A. Kruglanski, and T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 399–417). New York, NY: Sage.

Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review, 7,* 349–361. http://dx.doi.org/10.1207/S15327957PSPR0704_07

van Dick, R., Hirst, G., Grojean, M. W., & Wieseke, J. (2007). Relationships between leader and follower organizational identification and implications for follower attitudes and behaviour. *Journal of Occupational and Organizational Psychology, 80,* 133–150. http://dx.doi.org/10.1348/096317905X71831

van Dick, R., Lemoine, J. E., Steffens, N. K., Kerschreiter, R., Akfirat, S. A., Avanzi, L., . . . Haslam, S. A. (2018). Identity Leadership going global: Validation of the Identity Leadership Inventory (ILI) across 20 countries. *Journal of Occupational and Organizational Psychology, 91,* 697–728. http://dx.doi.org/10.1111/joop.12223

Vetlesen, A. J. (2005). *Evil and human agency: Understanding collective evildoing*. Cambridge, United Kingdom: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610776

Zimbardo, P. (1992). *Quiet rage (video)*. Stanford, CA: Stanford University.

Zimbardo, P. G. (2004). A situationist perspective on the psychology of evil: Understanding how good people are transformed into perpetrators. In A. Miller (Ed.), *The social psychology of good and evil* (pp. 21–50). New York, NY: Guilford Press.

Zimbardo, P. G. (2006). On rethinking the psychology of tyranny: The BBC prison study. *British Journal of Social Psychology, 45,* 47–53. http://dx.doi.org/10.1348/014466605X81720

Zimbardo, P. (2007). *The Lucifer effect: How good people turn evil*. London, United Kingdom: Random House.

Zimbardo, P., Banks, W. C., Haney, C., & Jaffe, D. (1973, April 8). The mind is a formidable jailer: A Pirandellian prison. *The New York Times Magazine, 6,* 38–60.

Zimbardo, P., Johnson, R. L., & McCann, V. (2012). *Psychology: Core concepts* (7th ed.). London, United Kingdom: Pearson.