The dirty work of the Stanford Prison Experiment: Re-reading the dramaturgy of coercion

Stephen Scott-Bottoms

Abstract
Almost 50 years on, the Stanford Prison Experiment of 1971 remains one of the most notorious and controversial psychology studies ever devised. It has often been treated as a cautionary tale about what can happen in prison situations if there is inadequate staff training or safeguarding, given the inherent power differentials between staff and inmates. But what exactly was the ‘situation’ in the simulated prison at Stanford University, and how exactly did the participants respond to it? This article provides a new analysis of the behaviour of the nine Stanford ‘guards’, which draws on unpublished archival records and original interviews with some of the participants. It adopts an interactionist approach, whereby the individual backgrounds and personalities of the participants are seen to inform their behaviour within the situation provided, as well as vice versa. A key suggestion to emerge from this analysis is that the conduct of the three guard shifts, within the experiment, differed significantly according the interaction of the men on each shift and their influence on each other. The article is framed by the concept of ‘dirty work’ and uses theoretical sources including Goffman and Festinger to explore the extent to which the adoption of a particular social role (in this case, that of prison guard) affects change in the behaviour and beliefs of individual role players. The argument seeks to show that some of the Stanford guards adopted strategies of ‘role distancing’ to insulate themselves from the coercive demands of their positions, while others found the guard role impacting troublingly on their senses of self.

Keywords
Stanford Prison Experiment, role playing, prison officers, cognitive dissonance, archival histories, institutions

The University of Manchester, UK

Corresponding author:
Stephen Scott-Bottoms, Department of Drama, The University of Manchester, Oxford Road, Manchester M13 9PL, UK. Email: stephen.bottoms@manchester.ac.uk
Much social-psychological evidence indicates that the best way to change attitudes or beliefs is to get people to behave as if they believe in what they are doing. This finding augurs ill for the individual maintaining “self” while performing institutional roles that are dehumanizing or detrimental to others. When the “dirty work” becomes our own, then we have become truly “socialized.”

Craig Haney and Philip Zimbardo (1977: 203–204)

The disarmingly casual phrase ‘dirty work’ was first given sociological ramifications by Everett Hughes, in his 1964 article ‘Good People and Dirty Work’. Echoing Hannah Arendt’s then-recent comments on ‘the banality of evil’—in describing Adolf Eichmann’s coordination of the Nazi Holocaust (Arendt, 1963)—Hughes referred to the conduct of the SS, during the genocide, as ‘the most colossal and dramatic piece of social dirty work the world has ever known’ (Hughes, 1964: 3). The slaughter of the Jews, he argued, arose not from some dispositional sadism integral to the German people but from a familiar process of moral rationalization: ‘dirty work’ is a repugnant activity carried out in good conscience by those who have convinced themselves that their actions serve a greater good. *It’s a dirty job, but someone has to do it.*

Hughes’s provocation resonates, in part, because of the discomforting awareness that such dirty working continues to be lauded, more than critiqued, in popular culture. Its paradigmatic expression is perhaps the 1971 cop movie *Dirty Harry*, in which Clint Eastwood plays a San Francisco detective who becomes so frustrated by the failures of a too-liberal legal system (which coddles criminals and releases them on technicalities) that he takes the law into his own hands. ‘He is our martyr – stained on our behalf’, Pauline Kael wrote archly of Eastwood’s self-sacrificial commitment to torturing and executing social deviants: ‘The dirtiness on Harry is the moral stain of recognition that evil must be dealt with . . . Violence has rarely been presented with such righteous relish’ (Kael, 1972).

A more mundane form of dirty work is that of the prison officer. Unless and until prisons are abolished, somebody has to be responsible for locking others up against their will—an unpleasant task even in the best of circumstances. It is this ‘guard role’ to which Haney and Zimbardo refer in my epigraph above and which they sought to explore under controlled circumstances in their Stanford Prison Experiment (SPE)—another cultural landmark of 1971. Young, male volunteers were asked to dress up as prisoners and guards and to act out those roles, around the clock, in a simulated prison environment in the basement of Stanford University’s Jordan Hall. After six days of a possible two weeks, however, the study was famously aborted by the experimenters. Several of the prisoners, they reported, had become far more distressed by the situation than anticipated, and some of the guards had begun abusing their power. This simulated dirty work had become unsettlingly real.

As a scholar of drama and theatre studies, I have long been fascinated by the Stanford experiment. This was, after all, a role-playing improvisation—an extended variation on the kind of power games regularly used in drama workshops for exploring interactive dynamics. As such, the SPE had no methodological precedent in social psychology. Like many young researchers during that febrile period of social protest and cultural disruption around the late 1960s and early 1970s, Philip Zimbardo and his then-graduate students, Craig Haney and Curtis Banks, were pushing against the strictures of established experimental practices. Their concern was that tightly controlled, repeatable lab procedures might never capture the fluidity and volatility of interactive social situations. They thus opted, in effect, to literalize role theory—the field of sociological enquiry that had recently been popularized by scholars such as Erving Goffman and Theodore Sarbin, and which drew its animating metaphors from the theatre. For their experiment to work,
Zimbardo and colleagues needed their participants to commit to acting in the sense outlined by the great American acting coach, Sanford Meisner: ‘living truthfully under imaginary circumstances’ (Meisner and Longwell, 1987: 15).

Given all this, I have often been puzzled by the lack of detailed explanation of the participants’ behaviour – and particularly the guards’ reported aggression – in the experimenters’ various accounts of what transpired. Where Dirty Harry offers the traditional benefits of dramatic enactment, by putting viewers vicariously ‘in the shoes’ of the protagonist, enabling us to understand his subjective motivations, the Stanford study has always been discussed from the perspective of the experimenters, looking in from outside. As social psychologists, they were interested in the aggregated impacts of environmental forces on observable behaviour but far less concerned with questions of personal response. Anyone wondering what the guards themselves were thinking will find little to satisfy them in these accounts.

Curiosity about this seeming blind spot prompted me to begin researching the experiment in more detail using my training as a theatre historian. Could I, perhaps, reconstruct this long-ago performance from the point of view of its participants? This article represents part of a wider project for which I have undertaken extensive work in two separate archive locations, where Zimbardo has deposited surviving documentation of the SPE. These collections are at Stanford University Library’s Department of Special Collections and University Archives, in California, and at the Cummings Center for the History of Psychology, University of Akron, Ohio. Both contain a substantial volume of qualitative feedback from the participants (some of it duplicated across both sites), in the form of questionnaire responses, audio recordings, interview transcripts and post-experimental diaries. These reflections have been relatively underused in the psychologists’ accounts of the experiment, apparently because of their discipline’s relative disregard (particularly in the early 1970s) for qualitative data as a legitimate form of evidence. For the theatre historian, conversely, the recollections of actors and spectators are of primary value in understanding what a given performance meant, in context.

I spent two periods of one week each, in 2018 and 2020, studying the Zimbardo archives at Stanford, as well as a week in Akron (2020). On both occasions, I made digital copies of much of the material I examined, for closer scrutiny at home. My method for examining these sources has involved seeking to build as detailed a picture as possible as to what happened during those six days, in what sequence, and how the participants felt about it. What became clear is that their perspectives were often quite different not only from those of the experimenters but from each other. To supplement my understanding of this historical material, I have also sought personal interviews with those participants whose names were in the public domain and whom I was able to locate through Internet searches. I have interviewed four of the nine former SPE guards in detail, as well as three prisoners – some in person, during visits to California, and some remotely via telephone/Internet. These interviews, which were recorded and transcribed, followed my own institution’s ethics guidelines, and the text of this article has been reviewed for accuracy by those quoted. Each interview lasted between one hour and two hours, a length of time which enabled us to move past the more ‘rehearsed’ responses that some of these men have often given to journalists. Although prompted by notes, the conversations were largely unstructured, so that participants’ own recollections and concerns led the conversation, more than my questions. All respondents spoke freely and with no signs of hesitancy, although clearly, for some of them, there was still an unresolved tension between the public notoriety of the experiment and their own experiences of it – a point I will return to in conclusion. It is my hope that, by tracing the varying attempts of the SPE’s
guards to adapt to the roles assigned to them, I can shed useful light on the subjective dynamics of dirty working.

**Diagnosing the situation**

According to its creators, the SPE validated a general hypothesis about the influence of situational forces on individual behaviour. Too often in public discourse, they argued in their key, 1973 journal account, ‘the deplorable condition of our penal system and its dehumanising effects upon prisoners and guards’ are ascribed to the pre-existing dispositional traits of both prisoners and staff – since the former are criminals, by definition, and the latter are often thought to be ‘sadistic, uneducated and insensitive people’ whose underlying traits explain their attraction to the job (Haney et al., 1973: 70). The SPE sought to neutralize this dispositional argument by screening out, in advance, any volunteers with histories of aggression or volatility. The study’s outcomes, it was argued, arose from structural aspects of the situation – and particularly the power disparity between prisoners and guards.

This bold, situational message resonated strongly in 1971. On August 21, the day after the experiment ended, Black Panther leader George Jackson was shot dead by tower guards at California’s San Quentin State Prison. The following month, a four-day hostage situation at Attica Correctional Facility, in upstate New York, was brought to an abrupt end when State Troopers retook the facility with tear gas and shotguns – the bloodiest assault on a penitentiary in US history. Prisons were in crisis, and answers were being sought, so initial accounts of the SPE were quickly picked up by the national news media. By October, Zimbardo (who had never yet set foot in a real prison) found himself testifying before a Congressional enquiry into prison conditions. As is also clear from his archived papers, he was deluged that fall with letters from prisoners and correctional staff alike, from across the United States, thanking him for articulating what they already knew – that abuses can arise from bad situations, not just bad people.

The clarity and simplicity of that basic argument had, and perhaps continues to have, considerable communicative power. Considered critically, however, the experimenters’ claim that the SPE was designed ‘to separate the effects of the prison environment per se from those attributable to a priori dispositions of its inhabitants’ actually raises more questions than it answers (Haney et al., 1973: 71). There is, after all, no such thing as the prison environment. There are many, differently administered prison regimes, and a central question in penology is why the differences between institutions make some of them so much more harmful than others. This being the case, it is necessary to ask what specific aspects of the SPE’s environmental design produced the negative effects reported. What, specifically, prompted the individuals involved to act as they did?

This question of situational prompts was at the core of recent, critical attacks on the SPE by the French writer Thibault LeTexier, who in 2018 published a book accusing the experimenters of intellectual dishonesty. *Histoire d’un Mensonge* (History of a Lie) prompted extensive discussion online and was eventually condensed into an English-language journal article for *American Psychologist*. Unfortunately, however, LeTexier’s intemperate attempts to ‘debunk’ the SPE largely ignored the wider questions posed by the experimenters about the impact of institutional roles on those socialized into them. Instead, his underlying assumption seems simply to be that any prompting of the participants represented inappropriate manipulation of a psychological study. LeTexier presents evidence which, he claims, ‘directly contradicts’ the experimenters’ suggestion that they ‘did not give any formal or detailed instructions about how to be an effective guard’ (LeTexier, 2019: 8).
The most widely discussed of LeTexier’s examples is an archive recording of a conversation between the simulated prison’s ‘Warden’, David Jaffe, and one of the volunteer guards, John Mark, from the second day of the study. Jaffe had been asked by Zimbardo to tell Mark that he was being too ‘soft’ in his treatment of prisoners: ‘The guards have to know that every guard is going to be what we call a “tough guard,”’ Jaffe insisted, because ‘the success of this experiment rides on the behavior of the guards to make it seem as realistic as possible’. Mark responded by citing his own encounters with police and campus security during anti-war demonstrations: ‘real life experience has taught me that tough, aggressive behavior is counterproductive’. Jaffe overruled him, however, by emphasizing that the experiment required him to comply with the role as conceived: ‘We need you to react as you imagine the “pigs” would’ (qtd. Zimbardo, 2007: 65).

LeTexier’s ‘discovery’ of this tape has been treated by some commentators as a smoking gun. And yet the Jaffe/Mark exchange, far from being a new revelation, is openly discussed by Zimbardo in his book-length account of the experiment, *The Lucifer Effect* (as my page referencing indicates). Elsewhere in the book, he states that other guards also sometimes had to be reminded to be ‘more assertive’ (2007: 81). Clearly, Zimbardo did not feel he had anything to hide, and – as he pointed out to journalists over the LeTexier controversy – he would not have made his archives available for consultation if he had. ‘The point is, it’s a drama’, he told a reporter for the *Stanford Daily*, in response to questions about the Jaffe/Mark dialogue: ‘You have to play the role in order for the whole thing to work’ (Knowles, 2018). He elaborated further in *Vox*:

These are people we’ve hired who are doing it for a salary, $15 a day, to play the role of guard. And [John Mark] is doing nothing. He’s sitting on the sideline, doing nothing, watching. He’s gotta earn his keep as a guard. (Resnick, 2018)

Mark himself strongly refutes this characterization of his behaviour. He had attempted, he told me, to play the guard role as best he could:

I tried to do what I was supposed to do without twisting myself so that I didn’t even recognise myself. To dress like that and push people around or tell them what to do, that was already a big stretch right there. So I didn’t have to go too far to get way out of my comfort zone. I wasn’t harsh, but I did my best and I tried to maintain a persona. (Mark, 2017)

Mark’s position is perfectly understandable, and his restraint might have been appropriate in many real-world prison contexts. Yet this was not a real prison, and the experimenters were concerned to ensure that participants did not simply collect their pay cheques for standing around in a university corridor, making nice with each other. As Zimbardo’s remarks about this being a ‘drama’ indicate, the objective was to simulate a strong sense of authoritarian control within the situation. The guards were therefore required to come *out* of their ‘comfort zones’, if necessary, by *acting* as if there was a significant power differential between them and the prisoners. ‘Under what conditions’, Zimbardo asked, ‘can role-playing simulation achieve a sufficient level of reality to become more than just a game?’ (1975: 35).

LeTexier’s position seems to be that the simulation should have involved simply assigning roles and saying nothing further. Yet it is simply not possible to run a role-playing exercise in a pristine vacuum that provides no behavioural prompts: improvisation feeds on available stimuli, whether or not those stimuli are intentionally provided. From a penological perspective, moreover, total non-intervention is totally nonsensical. No prison regime in the world functions without
The Stanford researchers have always been clear that their intention was to create a temporary approximation of the kind of abrasive, degrading ‘total institution’ discussed in Erving Goffman’s influential study *Asylums* (1961a). The experiment explicitly set out to generate ‘feelings of power and powerlessness, of control and oppression, of satisfaction and frustration, of arbitrary rule and resistance to authority, of status and anonymity, of machismo and emasculation’ (Haney et al., 1973: 72).

To achieve these dynamics, Zimbardo and his colleagues ‘identified a number of relevant conceptual variables’, which were operationalized through the study’s staging design. The visual rhetoric of the costume choices was particularly clear: the guards’ khaki uniforms ‘were intended to convey a military attitude, while the whistle and nightstick were carried as symbols of control and power’. Conversely, the prisoners’ loose, dress-like smocks, worn without pants or underwear, were designed to emasculate and humiliate, serving ‘as symbols of their dependence and subservience’ (Haney et al., 1973: 75–76). The experimenters’ concern to ‘fast-track’ the required psychological dynamics, within a limited time frame, was also apparent in the oppressive physical setting. The nine prisoners were allocated, three to a cell, to three small, windowless rooms adjoining a short stretch of basement corridor. The prisoners only ever left these blank, featureless spaces when being escorted to the bathroom or the Warden’s office – and when they did so, they were shackled and had paper bags placed over their heads, to prevent them seeing other parts of the building. The prisoners were watched over day and night by three, rotating guard shifts staffed by three men each, and this ratio of one guard for every three prisoners made for a particularly-intensive form of control and surveillance. As Prisoner 416 observed in his exit interview, ‘this jail is terribly overstaffed . . . it was a constant humiliation, because the guards had nothing to do but make-work, busy-work’ (SC 9.9).

The environment of ‘Stanford County Jail’ was designed, then, as an abstracted representation of overwhelming institutional power. Earlier in 1971, Haney had visited the maximum security prison at Trenton, New Jersey, at Zimbardo’s request, and Gresham Sykes’s description of that facility in *The Society of Captives* (1958) seems an apt summary of what they were seeking to approximate:

> an attempt is made to create and maintain total or almost total social control. The detailed regulations extending into every area of the individual’s life, the constant surveillance, the concentration of power into the hands of a ruling few, the wide gulf between the rulers and the ruled – all are elements of what we would usually call a totalitarian regime. (Sykes, 1958: xiv)

It was total control of this sort that the Stanford guards were encouraged to assert, during their orientation meeting the day before the experiment began. Crucially, though, they were not told *how* to achieve this. In any dramatic exercise, there is a basic distinction to be drawn between the script, or outline scenario, and the rehearsal of that script towards performance. In the SPE, the guards were asked to play out an authoritarian prison scenario: they were instructed to enforce a given set of rules and to perform a number of set tasks, to underline their arbitrary power (e.g. waking prisoners up with whistles, subjecting them to elongated count procedures). Yet they were not rehearsed into these roles, nor provided with training or direction in *how* to perform them. Thus, *contra* LeTexier, the researchers could reasonably claim that participants ‘were intentionally given only minimal guidelines for what it meant to be a guard’ (Haney et al., 1973: 75).

As one guard put it in a retrospective diary: ‘The basic outline was fairly simple, the degrees of effectiveness were left to our individual characters’ (SC 9.15). It was in this discrepant relationship...
between institutional expectation and personal aptitude, this respondent added, that the real challenge of the guard role lay: ‘there was much more conflict than [in] the role of prisoner’. In retrospect, he was grateful for having been assigned this challenging task of dirty work, because he had been forced to ask difficult questions of himself, that he continued to reflect upon months later: ‘I’ve never stopped learning from the experiment’ (SC 9.15).

**Manufacturing dissonance**

The role demands involved in maintaining control and authority over prisoners, even in the most benign of institutions, are widely recognized as a source of psychological conflict for correctional officers. As Liebling et al. note in their ethnographic study, *The Prison Officer*, ‘role conflict’ is something of an occupational hazard, given the contradictory range of functions often expected of staff: ‘In one day, an officer can be a supervisor, custodian, disciplinarian, peacekeeper, administrator, observer, manager, facilitator, mentor, provider, classifier and diplomat . . . . Versatility and flexibility are key requirements’ (2011: 48). At Stanford, the artifice of the prison simulation involved a narrowing down of these role functions to the coercive tasks of custodian and disciplinarian – a limiting of variables that was arguably appropriate to the scientific context. In effect, the researchers sought to isolate the form of role conflict in which they were most interested – the discrepancy between ‘good people’ and their ‘dirty work’.

The prison study can thus be read as a further exploration of *cognitive dissonance*, the theoretical framework that Zimbardo had spent most of the 1960s investigating. His predecessor at Stanford, Leon Festinger, had developed the concept during the late 1950s, and for a decade, it became a key area of enquiry for social psychologists. Festinger argued that when a person is obliged to live with two conflicting cognitions, she will be driven to reduce the dissonance between them by altering one or other of those cognitions. In *A Theory of Cognitive Dissonance* (first published in 1957), he outlined four distinct paradigms under which this phenomenon was apparent, one of which – *forced compliance* – addressed conditions which appear ‘to elicit overt behavior that is at variance with private opinion’ (Festinger, 1962: 261). If, for example, a person’s job requires her to make public statements which she does not personally believe to be true, she will likely seek to reduce the dissonance arising from this conflicted position. She can do this either by quitting the job or (if she needs the money) by convincing herself that the public statements are in fact justified: thus, ‘attitude change occurs when one behaves in a manner discrepant with his private attitudes and values’ (Zimbardo, 1968: 76–77). In the case of the prison study, the question arose as to whether the guards’ forced compliance with an authoritarian regime would prompt them to rationalize their behaviour as appropriate and necessary. As Gresham Sykes had put it: ‘How does living in a system of complete social control affect the personality of the rulers and the ruled?’ (1958: xvi).

Zimbardo’s assumption, based on Festinger’s theory, seems to have been that prison guards will naturally tend to reduce any dissonance experienced in role by becoming harsher in their attitudes towards prisoners: *if I have to treat them this way, then they must deserve it*. This rationale seems to underpin his boldly sweeping claims to Congress in 1971: ‘With regard to prisons, we can state that the mere act of assigning labels to people, such as “prisoners” and “guards” . . . . is sufficient to elicit pathological behavior’ (Zimbardo, 1973: 156). Yet the invocation of pathology does little to further our understanding of guard behaviour within in the SPE. Zimbardo’s reliance on dissonance theory, moreover, meant that he overlooked another rich theoretical framework, developed by Erving Goffman. In his essay ‘Role Distance’, published in the same year as *Asylums*...
(1961a), Goffman argued that social roles are rarely taken on unquestioningly by their performers. Rather, they are subjected to various forms of personal negotiation – just as an actor makes decisions about how best to play a part. There is always something of a gap, Goffman argued, between the normative role and a particular player’s attempts to inhabit it. Normative expectations present an outline scenario that must be adhered to, yet for any given role, there is a range of viable adaptation strategies. Goffman labelled these the typical modes of adherence to a norm.

In the case of the SPE, the experimenters reported three typical responses to the assigned guard role: ‘some guards were tough but fair (“played by the rules”), some went far beyond their roles to engage in creative cruelty and harassment, while a few were passive and rarely instigated any coercive control over the prisoners’ (Haney et al., 1973: 81). This tantalizing summary seems to map usefully onto Michael J Gilbert’s (1997) analysis of typical ‘work style descriptors’ for US prison officers. The four guard archetypes examined by Gilbert include ‘The Professional’, who seems to correspond to the SPE’s ‘tough but fair’ guards, and ‘The Enforcer’, who aggressively demands compliance with orders, ‘and is quick to use threats, verbal coercion and physical force’ (Gilbert, 1997: 50). Gilbert’s other two types, ‘The Reciprocator’ and ‘The Avoider’, offer further granulation of the third SPE category – those guards who, like John Mark, remained relatively non-coercive despite inducements to the contrary.

To understand this typology in more detail, however, we need to look closer at the personal role styles adopted by individual guards. As Goffman stresses, ‘typical role must of course be distinguished from the actual role performance of a concrete individual in a given position. Between typical response and actual response we can usually expect some difference’ (1961b: 93). He adds that, while social science tends to seek generalizing patterns and theories, a nuanced understanding of role behaviour depends on drilling down to individual specifics: thus, one of his stated objectives for ‘Role Distance’ as an essay was to adapt role theory ‘for use in close studies of moment-to-moment behavior’ (Goffman, 1961b: 95). The SPE offers a usefully contained model within which to study such interactions, and the qualitative feedback materials preserved in Zimbardo’s archives provide a vividly contextualized sense of how each guard adapted to the normative role demands.

### The morning shift (2 a.m. to 10 a.m.)

Any consideration of the performances of the SPE’s guards must begin with the micro-dynamics of the different shifts. There were three teams of three guards, operating in eight-hour stints starting at 2 a.m., 10 a.m. and 6 p.m., respectively. Let us call them the morning shift, the day shift and the evening shift. (There were also two standby guards who were occasionally called on, but their involvement was minimal.) During their orientation meeting with Zimbardo and Jaffe the day before the experiment began, the guards were invited to decide among themselves who would staff which shift. The only stipulation was that each trio should feature at least one taller man, to avoid them being physically outmatched by the prisoners. The resulting allocation was as follows:

| 2 a.m. to 10 a.m. | 10 a.m. to 6 p.m. | 6 p.m. to 2 a.m. |
|-------------------|------------------|------------------|
| Ceros, Andre      | Arnett, Terry    | Burdan, Chuck    |
| Van Orsdol, Karl  | Loftus, John     | Eshleman, Dave   |
| Varnish, Mike     | Mark, John       | Loftus, Geoff    |
The morning shift, staffed by guards Van Orsdol, Ceros* and Varnish*, faced a distinctive set of challenges. Although they could relax while prisoners were asleep, they were also tasked with waking them up twice daily, with whistles and shouting – first at 2.30 a.m. for a middle-of-the-night count procedure designed to disorientate, and then again at 6 am. Unsurprisingly, this group had to deal with some real resentment from prisoners, yet they were also perhaps the least sure of themselves as individuals. Karl Van Orsdol, fresh out of high school that summer, seems to have had shift leadership thrust upon him simply by virtue of being the tallest. Yet as he now recalls: ‘I think we were all making it up. I certainly didn’t feel like I was the leader’ (Van Orsdol, 2018).

This group coped by trying to follow the experimenters’ instructions as literally as possible. In effect, they adopted Gilbert’s ‘Professional’ staff mode, but in a context where by-the-book compliance meant being routinely coercive. Van Orsdol recalls becoming very angry with resistant prisoners on his first couple of shifts, when they proved reluctant to get out of bed quick enough, because this resistance seemed to undermine his tenuous grasp on authority. This was classic ‘rookie’ behaviour: inexperienced prison officers are often the most willing to comply with management orders to get tough, because ‘with rare exception, they [see] no alternative’ (Kauffman, 1988: 198).

In a retrospective interview, Guard Varnish provided a vivid insight into his personal attempts at adhering to role demands:

The first day I was on the job, I knew it was going to be tough because … we were supposed to try to make prisoners lose all sense of identity, and we were supposed to continually eliminate any personal contact and any feeling that they might have toward us or us toward them, and I really didn’t think I was going to be able to do it. So the first day, I sort of, oh, watched the others on my shift really. It wasn’t until the second day that I began to … force myself to really go about this in the right way … I began to treat them as coldly as possible, as harshly as possible – verbally, and would not in any way let any feelings that they might [show] affect me in the least. And by the third day, I had done such a good job convincing myself that they were of a distinct inferior order than myself [that] I just herded them together when I needed to have them do something, and put them back in their cells when I decided they needed to be … . The actual reason I acted this way I think was imposed from outside. I honestly didn’t feel that I was like that. [But] I actively sought to put myself in a frame of mind that was detrimental to getting along with prisoners in a good way. (M3484)3

Here, we see a clear example of what Goffman calls role embracement: ‘To embrace a role is to disappear completely into the virtual self available in the situation’ (1961b: 106). Just so, Varnish wrote, ‘there was no realization that this was “mere” role-playing while it was going on. Rather I treated it as a job with certain basic rules to follow (in particular, maximum harassment of prisoners)” (SC 9.23). Varnish’s willingness to behave as instructed, without questioning role requirements, illustrates Haney and Zimbardo’s observation that ‘institutional socialization is facilitated by a general tendency to grant social institutions a “presumption of rationality”: We simply assume that institutional behavior is founded on rational grounds and considered purpose’ (1977: 204).

Notice, too, that Varnish’s accommodation to the guard role was dependent on ‘convincing myself that [the prisoners] were of a distinct inferior order’. His shiftmate, Ceros, also reported attempting ‘to dehumanize them in order to make it easy for me’ (SC 9.15). Van Orsdol concurred that ‘I had to subdue my “humanitarianism” and think of the prisoners as lower than human’ (Van
Orsdol; SC 9.22). These comments suggest textbook examples of dissonance reduction: by derogating others as lesser beings, the exertion of power becomes more personally tolerable. ‘To the degree that inmates are defined as not-fully-adults’, Goffman suggests in *Asylums*, ‘staff need not feel a loss of self-respect by coercing deference from their charges’ (1961a: 108). For Varnish, embrace of the power role resulted in an unsettling internalization of the affrontive, impersonal attitude he had assumed towards the prisoners: ‘I was genuinely surprised to discover that I was capable of considerable irrational delight at the suffering of others’, particularly since, by the third or fourth day, ‘prisoners obeyed my merest whim with no apparent hesitation’ (SC 9.23). He gives the example of an incident, often referenced by Zimbardo, in which he arbitrarily instructed a prisoner to clean a toilet bowl with his bare hands, and the prisoner obediently complied. This episode illustrates the experimenters’ assertion that guard hostility, though ‘initially [used] only in response to perceived threats’, gradually became ‘a “natural” consequence of being in the uniform of a “guard” and asserting the power inherent in that role’ (Haney et al., 1973: 92, 94). ‘As I adopted the role’, Van Orsdol concurs, ‘the barbarism on my part became almost natural’ (SC 9.22).

These guards, however, were not entirely successful in their attempts at reducing cognitive dissonance. Far from simply relaxing into their dirty work, the morning shift seems to have found it increasingly difficult to avoid a creeping sense of guilt and self-disgust when reflecting on their actions. Following the toilet bowl incident, Varnish reports, he felt ‘physically uncomfortable and emotionally upset for remainder of experiment’ and was thus hugely relieved when it was called off on the sixth day (SC 9.23). Far from reducing the cognitive gap between self and role, he seems to have experienced on ongoing state of dissonance. This taught him a valuable lesson, he noted, about how to ‘differentiate between my “real” nature and any role-playing I might be required to do’ (SC 9.23). Van Orsdol, who had caught himself ordering his own parents around at home when not on shift, also described the experiment as teaching him a greater degree of self-awareness: ‘I feel that now, because of my ability to better understand it [authoritarian behaviour], I have become less demanding and bossy than I was before the experiment’ (SC 9.22).

Of the three morning shift guards, Andre Ceros seems to have been most heavily affected by role conflict. He had wanted to quit, he reported, but had accepted the inducement to comply: ‘I needed the money . . . so I became what I believe was expected of me’ (SC 9.15). In embracing the guard role, Ceros experienced moments of vicious pleasure in his power over others: ‘I laughed at his groping hands’, he recalled of leading a blindfolded prisoner to the bathroom; ‘Something sadistic surged within me’. And yet, he emphasizes, ‘I was afraid of that feeling, that laugh. I couldn’t believe that it came from me. I wondered about my authority, about my accepting it so readily’ (SC 9.15). As the week of the experiment wore on, Ceros found himself losing sleep over his inability to resolve this internal conflict: ‘I tried various ways of escaping myself – e.g. daydreaming, over-eating and a hard, cold attitude’ (SC 9.15). The worst of it, he suggests, was that attempts to tone down his initial, aggressive posture were thwarted by the prisoners’ established perceptions of him. In one incident of shouting and shoving, he sensed that the resentful prisoner hated me as the guard. He was reacting to the uniform. [So] I had no choice but to defend myself as a guard. It shocked me . . . I realized that I was just as much of a prisoner as they were. I was just a reaction to their feelings. (SC 9.15)

The discomforts expressed by the morning shift guards suggest a clear limit to dissonance theory. Role embracement does not lead straightforwardly to dissonance-reduction but may also
prompt reassessment and modification of behaviour. Certainly, the fact that some guards toned down their coerciveness over the course of the experiment did not go unnoticed by the prisoners. ‘I think people in general assumed their roles very strongly at first but then gradually became themselves’, prisoner Glenn-3401 stated in feedback (SC 9.3). According to Paul-5704, the guards ‘taught me that if something just isn’t in you, it shows when you have to do it’ (SC 9.2). These comments resonate with the extensive ethnographic research conducted by Liebling et al. for their book *The Prison Officer*: ‘the need for prison officers to “be themselves” as far as possible’ arises from the fact that it shows when role performance feels forced: ‘Officers were certain that you could not be effective if you were not yourself’ (2011: 60; emphasis original). Gilbert also suggests that an ability to moderate and adapt normative role demands is typical of the ‘Professional’ officer type, who is ‘reasonable, innovative, able to make exceptions’, and capable of resolving ‘ethical or moral dilemma[s] concerning the use of coercion’. He adds, however, that officers who cannot resolve such quandaries ‘are said to have a “conflicted morality” that prevents them from using coercion without damaging their self-image’ (Gilbert, 1997: 52). He could be describing the SPE’s morning shift.

### The day shift (10 a.m. to 6 p.m.)

The experiences of the day shift were shaped by a very different set of interpersonal dynamics. Where the morning shift had looked to each other for guidance, this trio had a clear leader in Terry Arnett*. A graduate student in sociology, Arnett was one of the oldest, most experienced participants in the study, as well as one of the tallest, and he exerted these status advantages from the outset. He had few qualms about providing an example of strong, coercive authority, in part because he recognized that he could still ‘be himself’ while doing so: ‘being superficially tough came easily to me. For one thing, I am an authoritarian person in some ways (even though I strongly dislike the trait in myself and others)’ (SC 9.13).

In *The Lucifer Effect*, Zimbardo describes Arnett as one of the most tyrannical of the SPE’s guards, an impression confirmed by evening-shift guard Geoff Loftus, who

> particularly disliked the way Terry handled them. [In] terms of being a guard, and doing guard functions, you’d have to say that Terry is a good guard, you know? He doesn’t get as much shit. But he is oppressive, he does get them to knuckle under. (SC Audiotape 3)

Such accounts suggest that, in his treatment of the prisoners, Arnett exemplified Gilbert’s Enforcer mode. He did not, however, perceive his behaviour as cruel and maintained that he ‘never experienced any particular sense of power or elation while punishing people or ordering them about’ (SC 9.13). Unlike the morning shift guards, whose role conflicts affected them emotionally even when off-duty, Arnett found that he was able to compartmentalize offstage life from onstage role. He distanced himself from the tough guard role by treating it simply as an ‘act’ – one necessary to the viability of the psychologists’ study. ‘I consciously felt that for the experiment to be at all useful, “guards” had to act something like guards’, he wrote later: ‘I didn’t know what that was but my two shift mates seemed incapable of dealing with this problem . . . at all’ (SC 9.13).

Arnett’s shiftmates were John Mark and John Loftus (Geoff’s younger brother), both of whom were far less comfortable with exerting authority. They were, in Gilbert’s terms, the experiment’s most notable Reciprocators: they instinctively sought to work with the prisoners rather than treating them as pawns. Ironically, Arnett’s harsher style provided them with some cover in this regard. As Mark now recalls: ‘I was aware and I was kind of relieved, in a way, that one guard
actually did kind of take it seriously and get with the program, because that took pressure off of me’ (Mark, 2017). Despite being ordered by Jaffe to get tougher in the guard role, Mark continued to follow his instincts. ‘I really didn’t want to give them a hard time’, he recalls of the prisoners. Mark opted to quietly subvert the coercive regime, by (for example) smuggling small comfort items to the prisoners. Two years later, in a longitudinal questionnaire response, Glenn-3401 still fondly remembered the moment when Mark had slipped him a plum.

‘The reciprocator wants to help people’, Gilbert notes, ‘and tends not to use coercive authority or physical force even when it is justifiable’. In applying a necessarily improvisatory sense of personal discretion, however, this officer type ‘may be inconsistent when making exceptions’ (Gilbert, 1997: 50). Prisoner feedback suggests some discomfort with Mark precisely because of such inconsistency: ‘He seemed to favor me’, noted Paul-5704, a nicotine addict whose cravings Mark occasionally mitigated by passing him cigarettes. In response to a standardized question about what defined a ‘bad guard’, Paul wrote: ‘someone who extended to me more than just the cursory courtesy and privileges’ or who was ‘a free rider’ (SC 9.2). The latter remark apparently refers to Mark’s reluctance to do the dirty work of the guard shift, so that others were obliged to. Stew-819, however, identified both John Loftus and John Mark (‘the cat with the pony tail’) as ‘good guards’, because they were ‘often quite compassionate to our needs’ (SC 9.7).

In sociometric feedback, John Loftus was consistently ranked as the most liked person in the experiment. This seems to reflect the comparative consistency with which he played the moderating role of Reciprocator. Although he made more visible attempts than Mark to act tough, ‘I never stopped thinking of the prisoners as people like me, you know? There was definitely a feeling at the time that we were trying to get through this together’ (Loftus, 2020). Loftus thus found subtle ways to signal a sense of camaraderie back to the prisoners, even while performing his duties. In feedback, several prisoners listed such subtle mitigations of the regime’s harshness as ‘the best thing’ about the experiment: there were, Jerry-5486 recalled, ‘short glances, touches or words between me and other individuals that said “we’re in this together,” and “I’m on your side.”’ This happened between guards and prisoners as well as pris. + pris.’ (SC 9.10). Such close-up signals were probably not apparent to the experimenters themselves, whose view of proceedings was via video relay, from a single camera at one end of the ‘prison yard’ corridor. Yet as Goffman’s discussion of role distance emphasizes:

face-to-face interaction provides an admirable context for executing a double stance – the individual’s actions unrebelliously adhere to the official definition of the situation, while gestural activity ... shows that he has not agreed to having all of himself defined by what is officially in progress. (Goffman, 1961b: 133)

This kind of discretionary role-distancing is famously alluded to in Society of Captives, where Sykes writes of the quiet reciprocity between officers and prisoners that can mitigate institutional inflexibility: ‘The unissued order, the deliberately ignored disobedience, the duty left unperformed – these are cracks in the monolith just as surely as are acts of defiance in the subject population’ (1958: 53). And yet, Sykes adds, it is precisely such unofficial softening of tough regimes that makes them more tolerable and thus more sustainable. Looking back, John Loftus blamed himself for complicity in perpetuating oppression, because he had never challenged Arnett’s more authoritarian behaviour: ‘I felt it would jeopardize the experiment and my participation if I questioned any actions by the guards’ (M3484). In a post-experimental diary, he noted that such rationalizations had failed to salve his conscience. ‘I knew inside I was a shit’, he wrote of his dissonant state:
that while I was good and just to the prisoners I failed myself. I let cruelty happen and I did nothing except feel guilty and be a nice guy. I honestly didn’t think I could do anything. I didn’t even try. (M3484)

It is possible, however, that in providing an ever-present example of ethical moderation, Lof-tus proved more of an influence on Arnett than he realized. The latter insisted in feedback that he too had maintained careful limits to his coercive persona: ‘the main influence on my behavior was the feeling . . . that real prison is brutal in that it is dehumanizing. [But] I could do this only in a limited way’ (SC 9.13). Arnett maintained that, if he could see an individual prisoner was suffering distress, he would consciously ease up on that person, even while being careful not to break character. That claim might appear self-favouring, were it not that Arnett’s discretion is corroborated by prisoner feedback. Clay-416, for example, singled him out for praise in an exit interview: ‘The tall guy with dark hair on the day shift is a good man. He’s gone over that a few times, but basically he’s a pretty good man’ (SC 9.8). Similarly, Tom-2903 ‘admired greatly the head guard on the afternoon shift. I pictured him as being probably one of the best types of guard that a prison could have in order to facilitate rehabilitation, maintain discipline and everything’ (SC 9.11). These expressions of respect point to the consistency and predictability with which Arnett maintained his strict persona. As Liebling et al. also found, in their UK study, ‘prisoners preferred officers to be “straight,” even if they were giving unwelcome news or instructions’ (2011: 93).

Evening shift (6 p.m. to 2 a.m.)

Dave Eshleman, the lead guard on the evening shift, was also a model of consistency. ‘[He] never played favorites’, noted Paul-5704: ‘he had it in for everyone’ (SC 9.2). Eshleman was the swaggering ‘John Wayne’ of SPE legend, who used his charisma and height (6’ 2’’) to assert dominance over shiftmates as well as prisoners: ‘I was definitely going for top dog’, he recalls, ‘and I think they simply followed’ (Eshleman, 2016). Nobody in the experiment ever questioned Eshleman’s leadership, and his shorter, slighter colleague, Chuck Burdan*, slipped readily into the role of aggressive sidekick. This left Geoff Loftus as the odd man out, aware that his own, gentler style did not fit. ‘While a team-performance is in progress’, Goffman observes, ‘any member of the team has the power to give the show away or disrupt it by inappropriate conduct’ (1959: 88). As if to sidestep this eventuality, Loftus took any opportunity he could to absent himself from the prison yard – by escorting individual prisoners to the bathroom, for example, or to meetings with the Warden or Counsellor (Haney). The Avoilder, Gilbert notes in his officer typology, ‘avoids confrontations and coercion’, and ‘uses the mechanical aspect of security and control to reduce contact with inmates’ (1997: 50, 55).

Loftus’s avoidance of role demands went as far as shedding parts of the guard uniform itself whenever possible. Challenged on this directly by Warden Jaffe, late in proceedings, he protested that ‘a couple of the nights there weren’t enough glasses to go around’; ‘the only times I took [my shirt] off is because it was too small for me’. Nevertheless, he conceded, ‘I do not like the idea of the uniform’. Loftus told Jaffe that he saw the guard role as

almost like a prison that you create yourself. [To] keep up and put on a face like that is just really one of the most oppressive things you can do . . . You just want to be able to tell everyone that this isn’t really me at all, and I’m not the person that’s confined in there . . . I’m not the sadistic type of person that enjoys this kind of thing. (SC Audiotape 3)
The strength of this avoidance reaction suggests that the elder Loftus had an instinctive aversion to the kind of dirty work demanded by the experiment. Not everyone, it seems, can ‘be themselves’ while locking up their peers. Dave Eshleman, however, experienced no such conflict with the guard role. Like Arnett, he treated it as an act that could be clearly delineated from his sense of self: it was, indeed, as an actor that he defined himself. Eshleman had just completed his freshman college year studying theatre, music and communications, and his reputation as a performer is recalled by both John Mark and Karl van Orsdol, from their days at Palo Alto’s Gunn High School. He thus chose to orientate himself within the experiment by treating the guard role as an opportunity to stretch his improvisation skills. Taking particular inspiration from the mirrored sunglasses assigned to all the guards, which referenced the Paul Newman movie, Cool Hand Luke (1967), Eshleman affected a Deep South accent like that of Strother Martin’s prison warden in that film and even mimicked some of his dialogue. ‘Throughout the entire experiment I was an actor’, he explained in an exit interview,

and I was hamming it up, and this was just my way of dealing with the whole thing. So, I purposely put on this attitude of being the mean, impersonal, drill-sergeant type of thing, when you’re really hounding the guy... I thought it would be better for the study if I presented what I thought to be a realistic relationship between guard and prisoner. (SC Audiotape 3)

The elision here between ‘ham acting’ and ‘realism’ is mirrored in the ambivalence with which Eshleman’s performance was received by the prisoners. On the one hand, as their John Wayne nickname suggests, most were fully aware that his posturing machismo was an act: prisoner feedback repeatedly comments on both Eshleman and Burdan ‘over-acting’ and ‘overplay[ing] the part’. ‘We enjoy in a way their ridiculousness’, Rich-1037 noted on day 3: ‘it’s funny because we know it’s a fake Southern accent’ (SC 9.10). Even so, Eshleman’s relentlessly uncompromising performance of harassment eventually led this same prisoner to wonder whether ‘that was his true nature... I was really upset by the fact that he could be that way’ (M3483.1.2). ‘The atmosphere around John Wayne was so depressing’, concurred Paul-5704: ‘It was like watching a bad movie and then realizing that I was a part of it and had to participate’ (M3483.1.3). Eshleman says that it never occurred to him, at the time, that the prisoners took his act so seriously: ‘I was probably not the most sensitive of teenagers, with regard to other people’s feelings’ (Eshleman, 2016). Geoff Loftus, though always loyal to his teammate, nonetheless acknowledged to Jaffe that, to maintain such a performance so consistently, ‘you really have to be able to almost ignore the fact that you might in some way be causing the prisoners a lot of anguish’ (SC Audiotape 3).

Eshleman recalls rationalizing his performance as being simply what the researchers needed from him: ‘I was helping them to create an uncomfortable experiment so they could get some results’ (2016). Yet his willingness to create such discomfort on their behalf speaks directly to Festinger’s ‘forced compliance’ paradigm. One way to reduce cognitive dissonance, Zimbardo writes, is ‘to assert a difference between the real you and the role-playing you’, and thus ‘to abdicate personal responsibility for your role-instigated behavior’. In these circumstances, one may feel morally licensed ‘to degrade, brutalize and dehumanize other human beings’ (Zimbardo, 1975: 54–55). According to Michael Gilbert, such ‘cynical’ self-distancing is typical of the Enforcer type among prison officers, whose ‘“John Wayne” behaviors [permit] little or no empathy for the human condition of inmates’ (1997: 55). Goffman employs the same term to make an even more damning suggestion: ‘It should be understood that the cynic, with all his professional
disinvolved, may obtain unprofessional pleasures from his masquerade, [because] he can toy at will with something his audience must take seriously’ (Goffman 1959: 28–29).

Eshleman insists that while he may have taken pleasure from his acting, he did not enjoy the coerciveness itself: indeed, he says, maintaining the John Wayne posture for eight hours at a time was exhausting. Having established this character at the start of the experiment, he felt obliged to keep it going, but like his fellow Enforcer, Arnett, he maintained a sense of personal limits: ‘There’s some point where you say okay, this is an act, and I can’t take it any further. So I could never have physically harmed anybody. It’s not in my nature to do that’ (Eshleman, 2016). Eshleman also dropped the act immediately after walking off set. Indeed, his proficiency in delineating between onstage and offstage behaviour clearly irritated shiftmate Chuck Burdan: ‘Outside the block he is friendly and just another rather straight college kid’ (M3484). Burdan himself struggled with such separation, and like the morning shift guards, he felt the guard role affecting his sense of self almost from the outset. In a post-experimental diary, he described himself ‘as a pacific and non-aggressive individual’ who could not have foreseen ‘a time when I might guard and/or maltreat other living things’. Yet as early as day two, he found himself enjoying his new sense of power: ‘Walking from my car, I suddenly wanted people to notice me, my uniform (“hey look what I’m doing”)’ (M3484).

Burdan might well have responded differently to the experiment had he been placed on a different shift. The immediate influence of Eshleman, however, proved decisive for him, as he sought to mimic the Enforcer role. ‘[Dave’s] behavior I thought was, in the context of the experiment, really rather brilliant’, he wrote; ‘quite upsetting at times, and yet I felt that it was proper in the context of the experience’ (M3484). Burdan thus tried to develop his own spin on this aggressive authoritarianism, and in doing so, he discovered an unexpected liking for it. During visiting periods, for example, when prisoners enjoyed brief meetings with friends and family, Burdan made a point of hovering over these conversations, and gratuitously interrupting them, even as his shiftmates stayed discreetly out of the way: ‘this was the first chance for the type of manipulative power that I really like’, he later wrote, ‘being a very noticed figure with almost complete control over what is said or not’ (M3484). It was also Burdan whom Craig Haney had to chastise for being overly physical with a prisoner at the end of a ‘Counsellor’ meeting: ‘The psychologist rebukes me for handcuffing and blindfolding a prisoner before leaving the office, and I resentfully reply that it is both necessary security and my business anyway’ (M3484).

Burdan, in short, became the guard to whom the term ‘sadistic’ might be most accurately applied. As Steven Pinker observes, ‘sadism is literally an acquired taste’, which requires two things to develop: ‘motives to enjoy the suffering of others, and a removal of the restraints that ordinarily inhibit people from acting on them’ (2011: 668, 663). Just so, Burdan seems to have discovered a taste for abusing power not just because of the situation in which he found himself but because of the shift to which he was assigned. Yet if his inspiration came from Eshleman, his performance became identifiably different in tone. Doug Korpi, a career clinical psychologist who was also the first of the Stanford prisoners to be released, recalls that among the guards,

Eshleman wasn’t the biggest asshole. He was playing a part. [But] the guy with the long dark hair that said he was a pacifist – he was genuinely in his heart, in his soul, a sadist. You knew it. When you’re in such close quarters, you can feel people. (Korpi, 2018; emphasis original)

According to psychoanalyst Erich Fromm, one should expect to find a small ‘percentage of unconscious sadists’ in any given work setting (1974: 94). Yet if Burdan was the exception that
proved this rule, in the context of the SPE, then the role-distancing of other guards perhaps underlines Fromm’s own, early critique of the experiment. ‘The difference between behaviour and character matters very much in this context’, he wrote – again invoking the notion of a discrepancy between role and self. ‘It is one thing to behave according to sadistic rules and another thing to want to be and to enjoy being cruel to people’ (Fromm, 1974: 93). Zimbardo seems to have worked from the assumption that such self/role distinctions are ultimately insupportable, because they become blurry through dissonance reduction: ‘beliefs change following a commitment to behavior discrepant with the original beliefs’ (Zimbardo and Ebbesen, 1969: 13). Yet as Fromm argued, the experimenters’ own account of the guards’ reactions falling into three, distinct categories suggested that the absorption of dirty working into one’s sense of self is far from inevitable: ‘[If] two thirds of the guards did not commit sadistic acts for personal “kicks,” the experiment seems rather to prove that one can not transform them so easily into sadists by providing them with the proper situation’ (Fromm, 1974: 93).

Conclusion

What conclusions can be drawn from this examination of the SPE’s shift-pattern dramaturgy? First, to the extent that it corroborates and adds flesh to Michael Gilbert’s typology of work-style adaptations for US prison officers, the Stanford simulation does seem to have reproduced some of the decision-making dilemmas faced by those who play guard roles ‘for real’. Second, the experiences of the Stanford guards suggest that personal role adaptations arise from the ways in which individual personalities transact both with situational requirements and with each other. Less self-confident guards, such as those clustered on the SPE’s morning shift, tended to be less critical of the role expectations presented to them and had to make a greater effort to play the role required. Their dissonance-reduction efforts thus resulted in a degree of personal ‘attitude change’ – including mentally dehumanizing the prisoners to make coercive rule enforcement more palatable. Conversely, the most self-assured of the guards – the postgraduate, Arnett, and the stage actor, Eshleman – proved relatively at ease with performing strict regime enforcement, because they felt able to distance themselves from their dirty work and to treat it as a necessary evil. Both were perceived by some as taking the Enforcer posture too far, yet rationalized their behaviour by determining personal moral limits. Their bold performances also afforded some of their shiftmates the cover needed to develop more resistant or subversive role adaptations.

As we have seen, the behaviour of guards on all three shifts sometimes exceeded what was necessary simply for the maintenance of the prison regime. Yet the increasing application of personal discretion in role performance meant that, over time, two of the three shifts moved towards a state of relative equilibrium between staff and prisoners. Thus, while the experimenters were justified in claiming that the simulated situation brought about troubling behaviour in ‘normal’ individuals, their assumptions arising from dissonance theory need some modification. Forced compliance with coercive role requirements can affect a realignment of personal values, as well as ‘cynical’ self-distancing and – in certain cases – the discovery of sadistic impulses. Yet the evidence of the SPE archive is that moral compasses can also reassert themselves, with surprising resilience, when personal limits are reached or breached. It is thus by no means a given that ‘dirty workers’ must end up rolling around in the mud. Ultimately, it would seem incumbent on prison authorities to provide training and structural conditions that legitimate and enable such moral resilience.

Wider society, too, has responsibilities in regard to supporting the work and well-being of prison officers. ‘I am finding’, Terry Arnett wrote to Zimbardo in November 1971, ‘that whenever
I talk with friends about the experiment I always end up feeling very defensive, as I am invariably accused in one way or another of having callously played with people’s minds’ (SC 9.13). Ever the sociologist, Arnett extrapolated from this personal experience to consider the situation of ‘real guards, whose identity in the eyes of their friends and community is intimately tied in with their occupation’ (emphasis original). If prison officers are made to feel defensive about their professional roles, he surmised, they might naturally gravitate towards socializing only with each other and their families. This, too, would be a dissonance-reduction method, corresponding with Festinger’s paradigm of social support — whereby we seek out the friends who tell us what we want to hear. Yet such professional inwardness might result, Arnett concluded, in ‘an intricate process of mutual reinforcement and legitimation of their [guard] role and behavior’, which could have ‘direct and profound effects on prisoners’. These reflections offer a salutary reminder that, if we believe prisons to be necessary at all (and there could surely be far fewer of them), then it is incumbent on all of us not only to demand high standards of officers but to show empathy and understanding for the difficult work they do. If it is a dirty job, and someone has to do it, then society as a whole remains collectively responsible.

Acknowledgements
I am grateful for the support and advice of the Special Collections staff at Stanford University Libraries, and of Lizette Barton at the Cummings Center for the History of Psychology, University of Akron. Special thanks, also, to my interviewees.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research for this article was supported by a Major Research Fellowship from the Leverhulme Trust.

Notes
1. SC references indicate material from the Philip G Zimbardo Papers (SC0750) held at the Department of Special Collections and University Archives, Stanford University Libraries. Subsequent numerals indicate the relevant box and file numbers. Thus, 9.9 indicates Box 9, Folder 9.
2. Guard names asterisked upon first use indicate the aliases adopted by Zimbardo in The Lucifer Effect, to protect anonymity. Non-asterisked names are the real names of participants, which are used with their permission. (In the case of Geoff Loftus, who is deceased, I have his brother John’s permission.)
3. Archive references headed M indicate the catalogued box numbers for the collection of Zimbardo papers held at the Cummings Center for the History of Psychology, University of Akron.

References
Arendt H (1963) Eichmann in Jerusalem: A Report on the Banality of Evil. New York, NY: Viking.
Eshleman D (2016) Personal interview with the author, Saratoga, 7 April.
Festinger L (1962) A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press.
Fromm E (1974) The Anatomy of Human Destructiveness. Harmondsworth, UK: Penguin.
Gilbert MJ (1997) The illusion of structure: a critique of the classical model of organization and the discretionary power of correctional officers. *Criminal Justice Review* 22(1): 49–64.

Goffman E (1959) *The Presentation of Self in Everyday Life*. New York, NY: Anchor.

Goffman E (1961a) *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. New York, NY: Anchor.

Goffman E (1961b) *Encounters: Two Studies in the Sociology of Interaction*. New York, NY: Bobbs-Merrill.

Haney C and Zimbardo P (1977) The socialization into criminality: on becoming a prisoner and a guard. In: Tapp JL and Levine FJ (eds), *Law, Justice, and the Individual in Society*. New York, NY: Holt, Rinehart and Winston, pp. 198–223.

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

Hughes E (1964) Good people and dirty work. *Social Problems* 10(1): 3–11.

Kael P (1972) Dirty Harry: Saint Cop. *The New Yorker*, 15 January. Available at: https://scrapsfromtheloft.com/2017/12/28/dirty-harry-saint-cop-review-by-pauline-kael/

Kauffman K (1988) *Prison Officers and Their World*. Cambridge, MA: Harvard University Press.

Knowles H (2018) Unchaining the Stanford Prison Experiment: Philip Zimbardo’s famous study falls under scrutiny, *Stanford Daily*, 13 November. Available at: https://www.stanforddaily.com/2018/11/13/unchaining-the-stanford-prison-experiment-philip-zimbardos-famous-study-falls-under-scrutiny/

Korpi D (2018) Personal interview with the author, Oakland, 3 April.

LeTexier T (2019) Debunking the Stanford Prison Experiment. *American Psychologist* 74(7): 823–839.

Liebling A, David P and Guy S (2011) *The Prison Officer*. London, UK and New York, NY: Routledge.

Lofus J (2020) Telephone interview with the author, 10 March.

Mark J (2017) Telephone interview with the author, 20 December.

Meisner S and Longwell D (1987) *Sanford Meisner on Acting*. New York, NY: Vintage.

Pinker S (2011) *The Better Angels of Our Nature: A History of Violence and Humanity*. London, UK: Penguin.

Resnick B (2018) Philip Zimbardo defends the Stanford Prison Experiment, his most famous work. *Vox*, 28 June 2018. Available at: https://www.vox.com/science-and-health/2018/6/28/17509470/stanford-prison-experiment-zimbardo-interview

Sykes G (1958) *The Society of Captives: A Study of a Maximum Security Prison*. Princeton, NJ: Princeton University Press.

Van Orsdol K (2018) Personal interview with the author, Palo Alto, 4 April.

Zimbardo PG (1968) The effect of effort and improvisation on self-persuasion produced by role-playing. In: Swingle PG (ed.), *Experiments in Social Psychology*. New York, NY: Academic Press, pp. 76–93.

Zimbardo PG (1973) The psychological power and pathology of imprisonment. In: Milton O and Wahler RG (eds), *Behavior Disorders*. Philadelphia, PA: Lippincott, pp. 151–161.

Zimbardo PG (1975) Transforming experimental research into advocacy for social change. In: Deutsch M and Hornstein HA (eds), *Applying Social Psychology*. New York, NY: Wiley, pp. 33–66.

Zimbardo PG (2007) *The Lucifer Effect: How Good People Turn Evil*. London, UK: Rider.

Zimbardo PG and Ebbe BE (1969) *Influencing Attitudes and Changing Behavior*. Reading, MA: Addison-Wesley.