The Psychological Record, 2012, 62, 133–154

Theoretical Article

HISTORICAL PSYCHOLOGY AND THE MILGRAM PARADIGM: TESTS OF AN EXPERIMENTALLY DERIVED MODEL OF DEFIANCE USING ACCOUNTS OF MASSACRES BY NAZI RESERVE POLICE BATTALION 101

Douglas J. Navarick

California State University, Fullerton

In Milgram's (1963, 1965a, 1965b, 1974/2004) experiments on destructive obedience, an authority figure repeatedly ordered a resistant participant to deliver what seemed to be increasingly painful shocks to a confederate victim who demanded to be released. A three-stage behavioral model (aversive conditioning of contextual stimuli, emergence of a decision point, and a choice between immediate and delayed reinforcers) proposes that participants withdraw to escape personal distress rather than to help the victim. The model explains significant details in accounts of the 1942 massacres of some 3,200 Jewish civilians at Józefów and Lomazy, Poland, by Nazi Reserve Police Battalion 101. The use of historical analyses to test nomothetic psychological theories offers unique opportunities for advancing understanding of destructive obedience.
Key words: destructive obedience, Holocaust, violence, delay discounting, prospect choice, self-control, impulsivity, negative reinforcement, moral judgment, anti-Semitism

Stanley Milgram's landmark experiments on destructive obedience (Milgram, 1963, 1965a, 1965b, 1974/2004) are widely seen as a seminal demonstration of the power of situational variables to induce individuals to harm others on the orders of a person who occupies a position of institutional authority. At the same time, Milgram's experiments have also come to represent a prototypic case of the kind of research that can no longer be conducted under the ethical standards applied by institutional review boards. For example, whereas participants today are always assured of their right to withdraw at any time through informed consent procedures, Milgram's participants, often visibly shaken and expressing concern for the welfare of the ostensible victim, were repeatedly ordered to continue by the authority figure, an actor who played the stereotypical role of a calm, efficient, dedicated scientist (see Table A of the Appendix for procedural details).

Although Milgram's findings and insights continue to be discussed in connection with a variety of social pathologies, including suicide terrorism (Atran, 2003) and military prisoner abuse (Fiske, Harris, & Cuddy, 2004; Zimbardo, 2007/2008), empirical research on the dynamics of destructive obedience has nearly ceased. Full replications of the Milgram paradigm ended in the United States by the mid-1970s and in Europe by the mid-1980s (Blass,

The author would like to thank Richard Lippa and two anonymous reviewers of this article for their many helpful suggestions.

Correspondence should be addressed to Dr. Douglas J. Navarick, Department of Psychology, P.O. Box 6846, California State University, Fullerton, Fullerton, CA 92834-6846. E-mail: dnavarick@fullerton.edu

1991, p. 6). Nevertheless, researchers have continued to explore alternative approaches in the form of simulations (Geller, 1978; Slater et al., 2006) and most recently in the form of a partial reproduction of Milgram's original procedures (Burger, 2009), an approach that required participants to be screened by clinical psychologists to minimize the risk of harm from the levels of stress that were anticipated.

While Burger's (2009) innovative approach has the potential to revive laboratory research using procedures similar to Milgram's, ethical considerations could prevent its widespread adoption by other researchers. In a commentary accompanying Burger's article, Miller (2009) doubted that many other institutional review boards would be as receptive as Burger's to an approach that pressured participants to remain in the study or subjected the participants to risks that required clinical screening (cf. Darley's, 1995, risk/benefit assessment, pp. 151–152).

An alternative approach involving mildly aversive, nondestructive obedience (i.e., no one is harmed) in a conventional laboratory situation has been explored by Navarick (2009) and by Navarick and Bellone (2010), but its connections to the defining emotional and moral dimensions of the Milgram paradigm have not been elaborated. In the present article, the broader theoretical context is provided in the form of a three-stage model of defiant behavior. Although the model was derived from experimental research, it was assessed by examining detailed historical accounts of destructive obedience for evidence related to hypothesized processes and predicted effects.

This methodology can be seen as an example of *historical psychology*, a term introduced by Runyan (1988, pp. 270–278) to refer to a relatively undeveloped branch of psychology that uses historical data to test the generality of nomothetic psychological theories (cf. *psychohistory*, the field that applies psychological theory to the interpretation of historical figures and events). A classic example would be McClelland's (1961) cross-era tests of his hypotheses concerning the relationships between achievement motivation, entrepreneurial behavior, and economic growth. McGuire (1976) viewed cross-era testing as a complement to cross-cultural and cross-species research, and he suggested that certain well-documented "archetypal episodes" (p. 170) could serve as a source of hypotheses and insights for the development of psychological theories.

The current model focuses on the following question: Once an individual accepts the role of being a subordinate, what factors in the immediate situation promote his or her withdrawal from that role as a means of defying the authority's command to harm innocents? Each of the model's three stages has clear experimental support, as will be discussed. External validity is assessed by analyzing analogous processes in what is reasonably regarded as an archetypal episode of destructive obedience at its earliest stage when subordinates receive their first orders to commit murder: the massacre in July 1942 of some 1,500 Jewish civilians in Józefów, Poland, by Nazi Reserve Police Battalion 101. This analysis draws on a detailed account of the events by Holocaust historian Christopher Browning in his book, *Ordinary Men* (1993/1998). Also examined is Browning's account of the battalion's next massacre approximately one month later at Lomazy, involving some 1,700 Jewish victims, in which commanders took steps to reduce the levels of defiance that they had witnessed at Józefów.

The central tenet of the model, supported by these two historical accounts as well as by experimental research, is that subordinates defy destructive authority to escape personal distress arising from an aversive situation. They do not withdraw to help the victim. Moral judgment (either reasoned or intuitive; Haidt, 2001) plays a subsidiary role, if any, by contributing to this distress, and the victim may then be helped as a by-product of the subordinates' escape. Therefore, analysis of the escape process itself (Navarick, 2009; Navarick & Bellone, 2010) can elucidate the dynamics of defiance, even though the origins of the driving emotions remain to be clarified (de Waal, 2009; Greene & Haidt, 2002; Haidt, 2001; Prinz, 2006).

The Incongruous Behavior of Milgram's Defiant Participants

In choosing to escape distress through withdrawal, participants in the Milgram experiments acted in their own interests rather than the victim's. The primacy of self-interest is

supported by a peculiar, ancillary fact about the participants who withdrew. Philip Zimbardo (Zimbardo, Maslach, & Haney, 2000) reported a conversation that he had with Milgram in which Zimbardo asked Milgram whether any of these participants went to help the victim after they withdrew or at least requested that the experimenter do so. Milgram's reply, according to Zimbardo, was "Not one, not ever!" (p. 196). Zimbardo commented that at a fundamental level Milgram had actually demonstrated 100% obedience. He noted that the submissive behavior of the defiant participants was reminiscent of that of elementary-school pupils who wait obediently in their seats for the teacher to dismiss them.

A distinguishing feature of the current model is that it is the only account of defiance that addresses the behavior of Milgram's participants *after* they withdrew. The participants' failure to follow through on their defiance is at odds with the notion that they withdrew to help the victim. To explain this incongruity in the context of prosocial motivation, one would need to make additional assumptions. For example, one could say that the teachers failed to go to the learner's aid because they feared retribution from the learner during an angry personal encounter. But if that were the case, why would the teachers not ask the experimenter to look in on the learner? Another assumption would be needed here.

The failure of Milgram's defiant participants to help the victim after they withdrew is entirely consistent with the premise of the current model that the participants acted to minimize their own distress, which would entail avoiding further conflict with the experimenter. The implication is that a subordinate's defiance of destructive commands through withdrawal may be analyzed as a form of escape behavior rather than as prosocial, helping behavior.

Applying History to Psychology: Overview of Issues

The present analysis assesses and interrelates three analogies. The central one is between the model and the Milgram-type obedience experiments from which the model is derived. The model is tested through the second analogy between it and the Nazi massacres. This second analogy, in turn, depends upon the plausibility of the third—between Milgram's experiments and the massacres.

The basic connection between Milgram's experiments and the massacres is that an authority figure ("experimenter"/commander) gave orders to a reluctant, distressed subordinate ("teacher"/policeman) to harm an innocent victim ("learner"/civilian). As will be discussed, similar temporal patterns of withdrawal among the teachers and the policemen further support this analogy by suggesting a common underlying process of escape. However, the many glaring differences between a social psychology experiment and a massacre perpetrated within an extraordinary historical context have led some historians to reject any insights that Milgram's research may offer into episodes of destructive obedience during the Holocaust (Cesarani, 2006; Goldhagen, 1996/1997, 2009; see Browning, 1993/1998, for an opposing view). A difference that is particularly relevant to the present analysis (specifically to Stage 2, the decision point) and that will be addressed is the fact that the ostensible victim in Milgram's experiments initially agreed to serve in his assigned role and then, after receiving the "shocks," demanded to be released, an event that prompted some participants to withdraw, whereas the victims of the massacres were forcibly subdued from the outset, so there was no analogous point at which the policemen might be prompted to reassess the moral implications of their acts.

The premise of the current approach is that the processes that govern escape from aversive stimulation are the same whatever the source of the stimulation (e.g., electric shock, empathic distress, moral judgment), the situation, or the historical era. Considerations of parsimony favor a single model that is consistent with both the experimental research and the historical events over a separate model for each domain. The goal, therefore, is to build a model based on the commonalities between the experiments on obedience and the accounts of the massacres while at the same time making provision for each situation's unique circumstances.

Each analogy poses its own challenges. For example, in applying the model to an historical episode, alternative interpretations inevitably arise based on the unique circumstances

of that event, such as the possibility that the behavior of the policemen at the second massacre at Lomazy could have been influenced by social pressures emanating from the first massacre at Józefów, rather than by the events at Lomazy. Evidence will be provided that will permit the current interpretation of historical events to be weighed against alternatives.

The Three-Stage Model of Defiance

The processes that govern obedience and defiance appear to be asymmetrical. Factors that promote the development of destructive obedience are described in a comprehensive, integrative model of inhumane conduct developed by Bandura (1991, 1999). The model attributes destructive obedience to the disengagement of deeply rooted self-regulatory mechanisms that ordinarily inhibit inhumane conduct and encourage prosocial conduct. Such "moral disengagement" is said to result from a variety of processes that can induce the individual to misinterpret or distort the harmful act (e.g., through euphemistic labeling), the severity of the harm done, and characteristics of the victim (e.g., through dehumanization or blaming the victim for the harm inflicted).

Once a subordinate has begun to implement destructive commands, moral judgment apparently becomes a secondary or incidental process in terminating obedient behavior. Milgram's hypothesis (1974/2004) and a related mathematical model developed by Rochat, Maggioni, and Modigliani (2000) propose that the basis for defiance is escape from strain or tension rather than moral judgment per se. As stated by Modigliani and Rochat (1995), participants' "overarching objective is *to alter or escape the unpleasant situation that faces them* [italics in original]" (p. 110).

The authenticity of the escape process in a subordinate's defiance of destructive commands is strikingly illustrated by Browning's (1993/1998) account of the events at Józefów. A distillation of that account is provided in the following section, after which the behavior of the perpetrators and the behavior of Milgram's participants will be directly compared and used as a frame of reference for the presentation of the current model.

Defiance by Shooters in Nazi Germany's Reserve Police Battalion 101

Most of the 500 members of Reserve Police Battalion 101 came from working-class backgrounds in the Hamburg area, a city that reputedly was particularly resistant to the regime's racist ideology. With an average age of 39, the men's formative years predated the Nazi era. Thus they were exposed to and potentially internalized moral values contrary to the regime's virulent racism, an ideology that transformed centuries-old, anti-Semitic prejudices into a policy of mass murder (Browning, 1993/1998; Goldhagen, 2009). The men were assigned to a police battalion because they were too old to be drafted into the army, and *most of them had no prior military experience.*

Battalion members' initial assignments during the period of May 1941 to June 1942 were deporting prisoners from Hamburg, guarding collection points, and escorting trains. However, beginning in July 1942 the battalion's assignment was abruptly changed from support duty to mass murder. From July to November 1943, these ordinary men shot a minimum of 38,000 defenseless victims: men, women, children, and infants.

The battalion's introduction to mass murder came in the town of Józefów, located 15 km from Warsaw, where over the course of a single day at least 1,500 people were shot. It was an assignment that most of the men had not anticipated and for which they were emotionally unprepared. Before a gathering of the battalion at dawn, the commander described the men's horrific assignment, and in acknowledgment of its distressing impact he announced that any of the older men could refuse involvement by stepping forward. Approximately 10 to 12 men out of the almost 500 men assembled accepted his offer and turned in their rifles.

The assignment consisted of three phases. The first was to round up Jewish residents and march them to a marketplace. Those who resisted or who were unable to walk—the sick, the frail, and infants—were to be shot on the spot. Virtually all of the men defied the

order to shoot infants and small children and instead allowed them to be taken by their mothers to the marketplace. The second phase was to transport the victims in trucks and by foot to prepared execution sites in the forest. Each man in a firing squad met an assigned victim at the edge of the forest and took him or her to an execution site. The interaction was "*face to face* [italics in original]" (Browning 1993/1998, p. 61). The third phase was to force the victims to lie in a row and to shoot them in a prescribed manner.

Browning estimates that 10 to 20% of the men in the firing squads quit at some point in the execution sequence. Some men quit at the market after they received instruction from a physician on how to shoot for maximum effect. Most men quit after performing one or more executions. Consistent with the pattern of withdrawals found in Milgram's (1965b, 1974/2004) experiments (the most relevant being the *touch proximity* condition, Experiment 4; procedural details of this and other cited experiments by Milgram are provided in the Appendix), Browning (1993/1998) stated that "Most of those who found the shooting impossible to bear quit very early" (p. 68). For example, after one shooter completed his first round of executions and returned to the trucks for his next victims, he encountered a mother and daughter from Germany. In a postwar interrogation he stated that he then became so sickened by the executions that he sought to be released and was reassigned to guard duty at the marketplace, a case that highlights the basic process hypothesized here to govern withdrawals: escape from distress.

Assessment of Moral Processes

During postwar interrogations, what emerged as an explanation for the men's withdrawals was "sheer physical revulsion" (Browning, 1993/1998, p. 74). Additional factors had to be involved because the men who completed the mission exhibited as much distress as the men who withdrew. But those additional factors did not include conscious moral judgment. As quoted by Browning, one policeman who killed as many as 20 people before quitting stated, "Truthfully I must say that at the time we didn't reflect about it at all. Only years later did any of us become truly conscious of what had happened then. . . . Only later did it first occur to me that [it] had not been right" (p. 72). Although accounts of events by Holocaust perpetrators during judicial proceedings are notorious for their self-serving misrepresentations, Browning commented that many of the testimonies from this battalion were unique for their apparent "candor and frankness" (p. xvii). Interpretations drawn from such testimony would seem worthy of consideration for insights into perpetrators' psychological state at the time of their crimes.

It is possible, of course, that while the men reported truthful and accurate recollections, their internal representations of their psychological state during the massacre were distorted by dissonance-reduction processes or self-serving bias tendencies. One therefore needs to weigh the plausibility of two possible interpretations: (1) All the men who withdrew consciously viewed their actions as morally wrong at the time but then through various psychological mechanisms distorted their memories in the same way; or (2) all the men who withdrew had the same experience during the crimes—disgust without conscious moral evaluations—and accurately reported what they experienced. While the first interpretation cannot be excluded, the second has the advantages of parsimony and of being the basis for the major historical analyses of this episode (Browning, 1993/1998; Goldhagen, 1996/1997).

Moral judgment has been defined by Haidt (2001) as a positive or negative evaluation of an individual's actions or characteristics based on values widely shared within a culture or subculture. At Józefów, the apparent absence among the men who withdrew of a conscious moral judgment that killing innocents was wrong argues against the involvement of two basic forms of moral appraisal as defined by Haidt: *moral reasoning* and *moral intuition*. Moral reasoning is said to be a conscious, multistep process in which a person intentionally uses select information to reach a moral judgment. Moral intuition is said to be a quick, unconscious reaction to an eliciting event that results in "the sudden appearance in consciousness of a moral judgment" (p. 818) accompanied by a positive or negative feeling.

Even in the absence of conscious moral evaluations, the quitters' feelings of disgust are potentially interpretable in moral terms depending on what one considers to be the source of their feelings. Haidt (2001) maintained that moral intuitions are partly innate. He traced the cognitive and emotional components of the human moral sense to primate protomorality. Particularly relevant to Józefów would be behavior of chimpanzees and other primates indicative of empathic concern for close others (de Waal, 2009). Police shooters who withdrew with feelings of disgust over the harm they inflicted on their victims may have felt a violation of a biologically rooted social norm that engenders empathy for close others (e.g., the case of the policeman who withdrew after encountering the mother and daughter from Germany). Greene and Haidt (2002) referred to this kind of moral violation that is understandable even by chimpanzees as "Me hurt you" (p. 519). Indeed, Browning (1993/1998) similarly suggested that some of the men may have acted on "humane instincts" (p. 74).

An alternative explanation of the men's revulsion and withdrawal is that they were reacting to observation of the horrific wounds that they inflicted at close range and that this was the first time that they had been called upon to kill. Goldhagen (1996/1997), writing about Józefów, rejected the notion that the men's emotional reaction and withdrawal reflected a moral judgment on the killing of innocents. He noted that novice soldiers in combat also "often feel sickened, throw up, and lose their appetites" (p. 221). In addition, at Józefów, in contrast to combat operations, commanders tolerated withdrawals and the men "easily got themselves excused from the killing" (p. 220). That the shooters' reactions of disgust and withdrawal resulted from experiencing the physical consequences of killing for the first time is further supported by the men's enhanced tolerance for killing innocents at the next massacre in Lomazy and by the absence of withdrawals after Józefów.

Moral Sentiments as Reflexes

Even if one favors the moral interpretion, the Józefów episode supports two conclusions related to the current model:

1. Withdrawals occurred without evident conscious moral reasoning or judgment.
2. Any moral sense that may have been involved likely took the form of a *moral sentiment* as conceptualized by Adam Smith in 1759: a mirror-like emotional reaction to the emotional expressions of others. In current terms, moral sentiments could be said to function as a type of unconditioned reflex, implying that at Józefów a victim's expressions of pain elicited in the policemen an instinctive, empathic response of distress that was potentially associable with environmental stimuli through classical conditioning. Primatologist Frans de Waal (2009) argued that such sentiments in humans do not themselves constitute moral judgments. To equate them with moral judgments would be to imply that "monkeys and apes are moral beings" (p. 8), a notion he rejected. Rather, de Waal held, as did Darwin, that human moral sentiments derive from animal social behavior and set the direction for moral reasoning and judgment.

Reinforcement Theory and Prospect Choice

The current model draws on reinforcement theory to elucidate the withdrawal process, with emphasis on the concept of *delay discounting*, the continuous decrease in the present value of a reinforcer as a function of the delay until its delivery (Ainslie, 1975; Fantino, 2000; Navarick & Fantino, 1976; Rachlin & Green, 1972). A reinforcer is defined as an event that follows a response and increases the probability that the response will be repeated. When Milgram's participants quit, and when the shooters at Józefów quit, it was the first time that they had defied their respective authorities. Therefore, they had not yet experienced the relief that came with escaping their aversive situations, which was the event that potentially functioned as reinforcement for withdrawal (in this case, it would be

negative reinforcement as it entailed removal of a "stimulus"). How then could reinforcement theory apply when reinforcement had not yet occurred?

Reinforcement effects are often studied using questionnaires that ask participants to make *prospect choices* between hypothetical alternatives that differ in reinforcer size and delay, and these specific reinforcer size/delay combinations are unlikely to have been experienced previously by the participants (e.g., Green, Myerson, & McFadden, 1997; Johnson, Bickel, Baker, Moore, Badger, & Budney, 2010; Lagorio & Madden, 2005; Navarick, 2004; Rachlin, Raineri, & Cross, 1991). Regardless of the type of reinforcer described or whether or not a reinforcer was actually delivered, the form of the discounting function has closely resembled the hyperbolic function that is typically found in animal research (Lagorio & Madden, 2005). Verbal representations of immediate and delayed outcomes can therefore produce choices that approximate the choices that individuals would make if they had previously experienced the outcomes.

Although Milgram's participants and the shooters at Józefów had not previously confronted destructive authority, escape from other sorts of stressful situations could have served as the basis for assigning values to immediate and delayed relief from stress and for making prospect choices. The specific escape responses that they employed were not previously reinforced, but initial approximations to those responses could have generalized from conventional authority–subordinate situations involving aversive, nondestructive commands, such as those by parents, teachers, and employers. Supporting a generalization process is Milgram's (1974/2004) view that an individual's predisposition to obey authorities is rooted in his or her early family life and subsequently strengthened in the institutional settings of school and work. Similar views have been expressed from the vantage point of other perspectives, for example, Alice Miller's (1980/2002) psychoanalytic theory that traces the mindset of detached, callous obedience that was typical of Nazi Germany to common child-rearing practices such as scolding, spanking, and beating.

The Paradigm of Signaled Avoidance

When Milgram's participants withdrew from their assigned role as teachers, and when the men of Battalion 101 withdrew from their assigned role as shooters, they avoided performing further harmful acts that were distressing to themselves. This emotional reaction set in motion the dynamic that facilitated their transition to defiance. Each prior act of obedience added to the aversiveness of the situation. By withdrawing from their roles, both the teachers and the shooters quickly escaped from aversive stimulation that foreshadowed another spike in distress.

This sequence of events closely resembles the paradigm of signaled avoidance in behavioral research (Herrnstein, 1969; Kamin, 1956; Mowrer, 1947): A warning cue signals the impending presentation of an aversive stimulus (typically, shock). In the absence of the designated avoidance response (the analog to quitting), the signal continues until the shock is presented. But if the avoidance response is performed before the shock occurs, the signal is terminated immediately and no shock is presented on that trial.

In Milgram's experiments, the warning signal (*conditioned stimulus*) would be the protocol that the participants followed when they administered the shocks. The *unconditioned stimulus* (analogous to shock for a rat) was the victim's expressions of pain. The *unconditioned response* was the participant's spike in distress, which was likely mediated by instinctive empathy for the victim (de Waal, 2009; Haidt, 2001). The removal of the warning signal (*termination of the conditioned stimulus*) was the point at which the experimenter announced the end of the experiment.

Mower (1947) proposed a two-factor theory of avoidance that became the basis for subsequent theoretical developments. The first factor was classical conditioning: The warning signal was paired with shock and became an aversive conditioned stimulus. The second factor was operant conditioning: The avoidance response was negatively reinforced by the removal of the warning signal. Although the two-factor theory made the now-dated

claim that avoidance responding was reinforced only by escape from the warning signal and not by the omission of shock, it has been shown that immediate removal of the warning signal facilitates avoidance responding (Kamin, 1956).

What is gained by adopting this conditioning analogy over previous interpretations expressed in terms of tension and strain (Milgram, 1974/2004; Rochat, Maggioni, & Modigliani, 2000)? By moving the locus of causation to the external warning signal (the negative reinforcer), attention is directed toward variables that may otherwise be overlooked, especially the prospective duration of the aversive task at its outset (based on initial instructions) and the progressive decrease in the time until completion of the task as obedience continues. Both withdrawal and task completion offer prospective negative reinforcement by terminating the aversive protocol, and the changing temporal relationship between these two sources of reinforcement is a potentially powerful influence on the choice to withdraw. Based on these and related considerations, tasks may be structured in such a manner as to reduce the incidence of withdrawals, as will be illustrated by the case of Nazi Reserve Police Battalion 101's second massacre at Lomazy.

Three Stages in the Process of Withdrawal

Table 1 presents a diagram of the model proposed here. Withdrawal results from three subprocesses: priming, recognition of a decision point, and choice. *Priming*, a kind of classical conditioning, entails associating the discomfort arising from obedient acts (triggered by the victim's expressions of pain or signs of injury) with contextual stimuli; the *decision point* is an event or circumscribed period that initiates a choice process that potentially leads to immediate escape from the aversive contextual stimuli but does not itself impel escape; and *choice* entails acceptance of one of two alternatives: immediate escape or delayed escape plus completion of the task, in other words, a choice between immediate negative reinforcement or delayed negative reinforcement plus positive reinforcement. Each of these stages has clear empirical support, as will now be discussed.

Table 1
Model of Withdrawal From Destructive Subordinate Roles

Stages leading to withdrawal		
Priming	Decision point	Choice
Distress arising from obedient acts is associated with contextual stimuli through classical aversive conditioning (Modigliani & Rochat, 1995; Slater et al., 2006). →	An event initiates a choice process that results in withdrawal or continuation, with most withdrawals clustered around the time of the cue (Meeus & Raaijmakers, 2006). →	There are two alternatives: immediate escape, which avoids further harmful acts, or delayed escape + task completion (i.e. immediate negative reinforcement) or delayed negative + positive reinforcement (Navarick, 1982, 2009).
Milgram participant		
When a lever is pressed, the protocol (CS) is paired with the learner's expression of pain (US), producing a spike in the participant's distress (UR). →	Learner demands to be released at 150 V (initiates moral reasoning process); or before 150 V, repeated exposures to US initiates outcome-sampling in some participants. →	Teacher quits; immediately escapes protocol; no further spikes in distress.
Shooter in Nazi Reserve Police Battalion 101 (Browning, 1993/1998)		
Execution site is paired with a spike in distress whenever a victim is shot. →	One or more executions (outcome sampling). →	Shooter asks officer for release; is reassigned to guard duty at collection point; immediately escapes execution site; no further spikes in distress.

Stage 1: Priming

In Milgram's (1974/2004) *voice-feedback* condition, in which the learner was heard but not seen, verbal interchanges between the participant and the experimenter provide evidence of a buildup of aversive stimulation prior to the participant's withdrawal.

Modigliani and Rochat (1995) analyzed sequences of verbal interchanges between the participant and experimenter using audiotapes that were made in Milgram's (1974/2004) Experiment 10, which was conducted in an office building in Bridgeport, Connecticut, under the auspices of an organization that seemingly had no affiliation with Yale University. (Additional details on this and other conditions can be found in the Appendix.)

Comments made by the participants were assigned to six categories that represented increasing degrees of opposition to the experimenter's commands. This scale was seen as a reflection of increasing stress or tension. Modigliani and Rochat (1995) found that the earlier in the shock sequence that participants made a comment in one of the three strongest categories of verbal protest, the more likely they were eventually to withdraw.

Direct physiological evidence for the buildup of stress prior to withdrawal was found by Slater et al. (2006) in virtual-reality simulations of Milgram's *proximity* condition, in which the learner could be seen and heard, and of the *remote* condition, in which the learner was neither seen nor heard (see additional procedural details in the Appendix). Stress was measured by assessing heart rate and skin conductance levels. Twenty "shocks" of increasing intensity could be administered, with the learner first demanding to be released after receiving the ninth shock (as compared with the 10th shock, 150 V, out of a maximum of 30 in Milgram's experiments).

Participants in the proximity condition showed higher skin conductance levels and heart rates than participants in the remote condition, and the skin conductance levels of these two groups increasingly diverged as the session progressed and the learner's reactions became more severe. Only one of 24 participants quit when the learner first demanded to leave (one had quit previously), which was the primary trigger for departures in Milgram's experiments (Packer, 2008). Most left just before the end of the session: three at Level 19, one at 18, and one at 16. It appeared that stress levels rose too slowly to produce withdrawals at the usual point, an indication that participants must be sufficiently primed by discomfort to take action when the situation calls for a change in their orientation to authority.

Stage 2: The Decision Point

In Milgram-type experiments, the learner's first demand to be released appears to initiate the process of choosing between continuing and withdrawing. It is also the point at which moral appraisal is most clearly implicated.

Packer (2008) conducted a meta-analysis of the withdrawal frequencies at different shock levels across eight of Milgram's (1974/2004) basic conditions (Experiments 2–6 and 8–10). Overall obedience rates varied from .65 to .30. However, as obedience rates dropped, the number of participants who quit increased only at the 150-V level. Although the learner repeatedly pleaded to be released at higher voltages, the only subsequent reaction that appeared to increase withdrawal rates, if only slightly, was his declaration at 300 V that he would no longer provide answers on the test.

Packer (2008) characterized the learner's initial plea to stop as a "critical decision point" (p. 303) at which some participants apparently came to the conclusion that the learner's right to leave outweighed the experimenter's authority to order continuation of the procedures. Packer suggested that participants who continued beyond 150 V may have justified their actions through dissonance-reduction processes or through a foot-in-the-door compliance process whereby acquiescence to earlier commands with less severe consequences constrained future decisions to be consistent with the previous ones.

A series of experiments on administrative obedience conducted in the Netherlands by Meeus and Raaijmakers (1986, 1995) provides additional evidence that a victim's first

request to be released from a harmful protocol is an effective withdrawal cue in Milgram-type experiments, but its potential influence may be revealed only in the absence of the authority. In other words, even when virtually all participants who are under the authority's surveillance are fully obedient, we can infer the existence of a latent withdrawal cue that will become functional for some participants when the surveillance is removed.

In the Meeus and Raaijmaker (1986) study, participants from the community were recruited by newspaper advertisements to assist the psychology department in a task of selecting applicants for governmental positions. The participant was to read a series of test questions to the applicant, who had to pass the test to get the position. In addition, the participant was asked to assist the department in an unrelated research project on the effects of stress on test performance. At 15 designated points in the test, the participant was to make scripted disparaging comments ("stress remarks") about the applicant's answers that could interfere with test performance. Stress levels were to be measured by electrodes placed on the applicant's skin. With the participant present, the applicant (a confederate) agreed to participate but was deceptively informed that the comments would provide accurate feedback. This conditional consent made the condition most similar to Milgram's (1974/2004) Experiment 9, where the victim "enters with prior conditions" (p. 61; see the Appendix for further details). The applicant was located in another room and, in response to the false disparaging remarks, expressed increasingly anguished protests that were accompanied by numerical displays on a monitor showing increasing levels of stress as measured by the electrodes. The applicant first asked the participant to stop making the comments after the eighth stress remark, and these requests continued with greater vehemence after subsequent stress remarks.

When the experimenter was present, 91.7% of the participants made all of the stress remarks (versus 40.0% in Milgram's Experiment 9 and 62.5% in the basic voice-feedback condition), even though during the debriefing most participants reported a high level of stress and indicated that they strongly disliked making the remarks (Meeus & Raaijmakers, 1995). But when the experimenter was absent, the obedience rate dropped to 36.4% (versus 20.5% in Milgram's similar Experiment 7, "experimenter absent").

The pattern of withdrawals in the experimenter-absent condition in the Meeus and Raaijmakers (1986) study replicated the pattern found in Milgram's experiments. Most withdrawals (36%) occurred at Stress Remark 8, with the remainder clustered around this point (29%, 21%, 0%, and 14% after Remarks 9–12, respectively). By inference, the cue for withdrawal was present but latent in the condition where the experimenter suppressed defiance through surveillance (combined with implicit disapproval of any request to deviate from the protocol).

As previously discussed, the analogy between Milgram-type experiments and the massacres is the basis for using the accounts of the massacres to test the model. The decision point at 150 V highlights a discrepancy between the two situations that requires resolution.

In Milgram-type experiments, the victims initially consent to participating in a situation that eventually harms them. Their initial demand to be released would change the situation from one of consent to one of subjugation if the participant continued to follow orders. The studies by Packer (2008) and by Meeus and Raaijmakers (1986, 1995) show that some participants will withdraw soon after this point rather than become an agent of the authority's oppression (assuming that the participants are sufficiently primed by distress, i.e., Stage 1). But this scenario, with its "critical decision point" (Packer, 2008; cf. Gilbert, 1981), has no direct counterpart in Józefów or in countless other genocidal massacres (Goldhagen, 2009) where victims are forced into submission from the outset.

Any choice process has a starting point. In operant studies of choice in pigeons, the starting point is readily identified: The animals begin to make choice responses when a discriminative stimulus is presented by illuminating two discs that the animals have been trained to peck. In Milgram-type experiments, the discriminative stimulus that begins the process of choosing between obeying and withdrawing is also readily identified because

(a) it is a discrete environmental event that changes the situation, (b) it appears to initiate a process of moral reasoning that was not relevant during the preceding period (when the victim implicitly consented to being shocked), and (c) it produces a sharp increase in withdrawals consistent with the change in circumstances.

At Józefów, a form of appraisal also appears to have occurred during an early, circumscribed period of the operation and to have produced most of the withdrawals, but this appraisal was not about the moral implications of shooting innocents: It was about the shooters' tolerance for distress. One policeman explained the dynamics of the men's withdrawals as follows: "It was one thing to refuse at the beginning . . . and quite another to try to shoot but not be able to continue" (Browning, 1993/1998, p. 72).

This statement has two implications: (1) For most of the men, the choice process started after they shot their first victim rather than before. The initiating event was not some dispassionate moral judgment that they made after they received their orders; and (2) the phrase "try to shoot" implies a process in which the perpetrator sampled the horrendous consequences of his obedient acts and projected a repetition of those consequences until the end of the operation. Rather than a discrete event as in Milgram-type experiments, the decision point appears to have been a circumscribed period of *outcome sampling*. In both cases, a buildup of aversive stimulation before or during the appraisal process preceded the choice to withdraw.

An analogous case of outcome sampling is suggested by an experiment by Navarick (2009) on nondestructive obedience in which college students were instructed to repeatedly choose a schedule of brief reinforcement followed by a long time-out over a schedule of extended reinforcement followed by a brief time-out. Participants could tap a desk bell at any time to signal their choice to withdraw. The function relating the percent of total withdrawals to the number of trials completed was unimodal, with a peak at two trials. It closely resembled Milgram's (1974/2004) distributions of "breakoff points" (p. 28), which had sharp peaks at 150 V. Apparently, the students in Navarick's (2009) experiment, like the shooters at Józefów, sampled the aversive consequences of their obedience and soon found that they would "not be able to continue."

Stage 3: Choice

Why do most withdrawals occur close in time to the decision point? Although a pattern of quick withdrawal may be consistent with intuition, a process is needed to fill the gap between the decision point and the act. A plausible link can be found in behavioral studies of impulsive choice and self-control (Ainslie, 1975).

The basic paradigm involves repeatedly offering opportunities to choose between an immediate, small reinforcer and a delayed, large reinforcer. Consistent choice of the former schedule represents impulsivity, whereas consistent choice of the latter represents self-control. Nonhuman species are typically impulsive (Ishii & Sakagami, 2002; Navarick & Fantino, 1976).

Adult humans show wide individual differences when responding for positive reinforcers, such as playing a video game (Millar & Navarick, 1984) or watching videos of cartoons (Navarick, 1998, 2001). However, most participants make consistent impulsive choices when in the presence of an aversive stimulus (white noise delivered through headphones) that they can terminate either immediately for a short period of relief or after a very long delay for a greater period of relief (Navarick, 1982). At shorter delays until offset of the noise, the larger amount of reinforcement is preferred. Such situations in which impulsive choice is induced by aversive stimulation are particularly relevant to the Milgram paradigm.

At each step in the voltage sequence in the Milgram paradigm, the participants had a choice between withdrawing and obtaining negative reinforcement through escape immediately or continuing and obtaining negative reinforcement through escape after a delay. If they continued until the end of the session, they would additionally receive the positive

reinforcement of completing the task (Kish, Woody, & Frankel, 1977) and satisfying the demands of the experimenter.

It should be emphasized that Milgram's participants were all volunteers from the community who answered a newspaper ad or direct-mail solicitation, which suggests that they attached a higher value to assisting with scientific research than did most others in the community. For these participants, completing the task could be conceptualized as prospective positive reinforcement because task completion was implied by the solicitation and normally would have consummated the chain of behaviors that the participants agreed to undertake.

To the extent that escape and task completion were sources of reinforcement, the participants were choosing between a relatively small amount of reinforcement that they could obtain immediately and a larger total amount of reinforcement that they could obtain only after a delay. Withdrawal would thus be a form of impulsive choice and full obedience would be a form of self-control.

A reverse form of self-control could also occur in which a subordinate could choose to produce an immediate increase in distress through confrontation with a disapproving authority because of the prospect that this would bring greater positive self-regard in the future, thereby making withdrawal a form of self-control and obedience a form of impulsivity (cf. Ainslie, 1975). There are several considerations that make reverse self-control an implausible interpretation of subordinates' behavior both in Milgram's experiments and at Józefów.

First, as previously discussed, delay-discounting functions for prospective reinforcers have been demonstrated in studies using questionnaires that describe the alternatives. Verbal representations of future outcomes are part of the choice process. At Józefów there was no evidence of moral reasoning that might have created a verbal representation of future positive self-regard when the choice was being made between shooting and not shooting. Second, if Milgram's defiant participants acted on moral principle and looked forward to the greater self-esteem that withdrawal would bring them in the future, why did they not help the victim after they withdrew? Additional assumptions would be needed to explain why they engaged in behavior that was incongruent with their moral judgment after the experiment but engaged in behavior that was congruent with their moral judgment during the experiment. The current model requires no additional assumptions to explain participants' behavior during and after the experiment. Third, whether one interprets the choice to withdraw as impulsivity or as self-control, the fact remains that subordinates obtained relief from distress immediately after their confrontation with the authority. It is difficult to justify the claim that the prospect of immediate relief played no role in the subordinates' choice to withdraw and that their choice was entirely due to the prospect of future positive self-regard.

The current model also helps to explain the upside-down attitudes of the shooters who killed all of their assigned victims. According to Browning (1993/1998, p. 185), the fully obedient shooters viewed their killing of defenseless men, women, and children as a sign of toughness, not cowardice. Yet they reacted emotionally to the shootings in much the same way as the men who quit—with horror and revulsion. (It may be noted here that whatever anti-Semitic prejudices the fully obedient men may have had, these biases did not add up to making murder a source of positive reinforcement or sadistic pleasure *during the initial massacre*. As discussed by Browning, the men's motivations changed over the course of repeated massacres, and many of them eventually volunteered or sought out opportunities to commit murder rather than waiting for orders.) Having accepted the legitimacy of their mission, the men who completed it rationalized their actions as evidence of their fortitude (i.e., self-control) and scornfully looked down upon the relatively few men who defied the orders as being weak. This disapproval by peers is a complicating factor that needs to be considered in the analysis of obedience at the next massacre in Lomazy.

One implication of the model for Milgram's experiments is that as participants pressed more and more shock levers, the delay preceding the larger amount of reinforcement progressively decreased, which reduced the temporal advantage associated with withdrawal.

In Navarick's experiment (1982), decreases in the delay to the longer period of relief produced fewer choices of the immediate, short period of relief. The trend in obedience should thus have been toward fewer and fewer choices to withdraw as the end of the session approached, which was the typical pattern that Milgram found. Such an escape process is vividly described by Zimbardo (2007/2008) in his interpretation of the situation from the point of view of the participant who nears the end of the session: "A few more lever presses is the fast way out . . . a simple matter of up and then out" (p. 272).

Delay discounting and choice. In the choice stage of the three-stage model, a key process promoting withdrawal is delay discounting (Ainslie, 1975; Fantino, 2000; Navarick, 2004), the continuous decrease in the present value of a reinforcer as a function of the delay until its delivery. In the choice between immediate escape and delayed escape plus task completion, the present value of delayed escape is assumed to be discounted (Navarick, 1982). However, the present value of task completion is assumed to be constant insofar as it functions as a stable individual difference characteristic (Kish, Woody, & Frankel, 1977; cf. Harber, Zimbardo, & Boyd's, 2003, characterization of individuals who have a predominantly future time perspective).

In other words, the prospect of completing a task tomorrow could be as motivating as completing the task right now. But the prospect of getting relief from a toothache tomorrow is likely to be much less motivating than is getting rid of it right now.

Without considering the precise form of the discounting function for negative reinforcement, the basic choice process of the model may be summarized as follows:

$$V_w = V_e - (V_{e*}/D + Vc) \quad (1)$$

where, at the time of choice, V_w is the value of withdrawal, V_e is the value of escaping now, V_{e*} is the value of escaping after completing the task, D is the duration of the delay until the task is completed, and Vc is the value of task completion.

Participants will withdraw sooner rather than later if the value of delayed escape is discounted sufficiently so that the value of immediate escape exceeds the combined value of delayed escape and task completion. However, as the session progresses and D decreases, the combined value of delayed escape and task completion rises and eventually surpasses the value of immediate escape.

Confronting authority. Equation 1 does not take into account the stress that would be produced by confronting a disapproving authority figure. Milgram (1974/2004) suggested that a potential deterrent to a subordinate's breaking with authority was the "awkwardness of withdrawal" (p. 7). As noted by Ross (1988), Milgram's procedure gave the participants no specific means for withdrawing and left them to their own devices to find a way out. Most participants did at times step out of their subordinate role to express concerns for the learner or to challenge the experimenter's judgment, and in that sense they were at least partially defiant, but they often could not find a way to translate their interruptions of the protocol into a complete break with it. Ross surmised that obedience rates would have been much lower if the participants could simply have pressed a button to withdraw instead of arguing with the experimenter.

Economist George Akerlof (1991) proposed a procrastination model of obedience (his Equations 4 to 6) based on the assumption that withdrawal would have greater negative utility as an immediate, impending event than it would as a future prospect. Therefore, as subordinates continued to perform the task, they would tend to keep choosing the less negative, future prospect and end up being fully obedient. Akerlof's model does not explain withdrawals.

In terms of the current model, the prospect of an aversive confrontation that exerts its maximum effect at the present moment could be represented as follows:

$$V_w = (V_e - A) - (V_{e*}/D + Vc) \quad (2)$$

where A, the aversiveness of confronting authority, incorporates the combined effects of all impediments to withdrawal, such as protracted argumentation with the authority or a

subordinate's having to act among peers who continue to obey rather than among peers who periodically withdraw (Milgram, 1965a, 1974/2004, Experiment 17). In other words, the prospect of immediate relief through withdrawal is partially or fully offset by the prospect of an immediate increase in distress from confrontation with the authority.

At Józefów, the major who commanded the battalion set a tone of tolerance for withdrawal by announcing at the outset that the men could opt out of the shootings. Officers at execution sites took their cues from the major and always (if grudgingly) granted requests for release. As in Milgram's experiments, the authority figures at Józefów deterred withdrawals through argumentation, not through threats of bodily harm, which further supports the analogy that is the basis for the current historical test. In fact, for the entire period of the Holocaust there has never been a documented case of serious punishment for any subordinate who refused orders to murder an unarmed civilian (Browning, 1993/1998).

Additionally, the shooters at Józefów could see other men leaving the execution sites, which would be expected to undermine the norms that supported obedience based on Milgram's findings (1965a, 1974/2004, Experiment 17) for a condition in which two confederate peers of the teacher withdrew relatively early in the shock sequence. This arrangement produced the lowest level of obedience of any of the situational variables that Milgram investigated.

Predictions. Equation 1 suggests several variables that could be used to predict or control withdrawals from destructive subordinate roles. Of particular importance is the *time frame* specified for the task or mission (cf. Fantino, 2000). The model assumes that subordinates define their alternatives as escaping now versus escaping upon completion of the task. A shorter task would reduce the value of D and increase the value of V_{e^*}. Most of the Nazi policemen quit after shooting at least one victim. Had the mission been framed as shooting several victims rather than as shooting victims for an indefinitely long period of time, the percentage of withdrawals would plausibly have been lower at the beginning of the massacre when most withdrawals occurred. In other words, *the same objective duration since the start of the mission should be tolerated differently depending upon whether the projected end to the mission is near or distant.* This principle was applied in the police battalion's second massacre at Lomazy, to be discussed in the next section, where commanders divided a longer killing mission into several shorter ones with rest periods between bouts of shooting.

In an experimental context, the time-frame prediction would serve to distinguish the current model from the "dynamical system" model developed by Rochat et al. (2000), which explicitly uses the time since the beginning of the task to predict obedience or withdrawal. The passage of time is emphasized because it is taken as a useful marker correlated with the participant's rising tension, which stems both from the repeated demands of the experimenter and the increasingly severe expressions of pain by the learner. The outcome is said to depend on several other variables, including internal resistance and two thresholds, but no provision is made for the *prospective* duration of the task. The probability of a participant withdrawing *at the current time* should be the same whether there are 5 minutes to go until the experiment is completed or 50 minutes to go until the experiment is completed. The current model predicts a higher probability of withdrawal at the current time with longer delays until task completion. This is not to say that elapsed time is irrelevant. Rather, withdrawal needs to be considered as function of *both* the elapsed time and the remaining time.

A second variable potentially linked to Equation 1 would be questionnaire measures of *task-completion motivation* (Kish et al., 1977; cf. Harber et al.'s, 2003, discussion of correlates of a future time perspective), an indication of the value of Vc in Equation 1. Higher values of Vc would reduce the value of withdrawal, V_w, at all delays. As compared to a generic questionnaire pertaining to all tasks, a method that could more precisely reveal differences in Vc for a particular task is a comparison between participants who were randomly assigned to the task and participants who volunteered for it. Milgram's (1974/2004) participants were volunteers from the community who presumably attached a particularly

high value to assisting with scientific research, a procedure that allowed for self-selection bias in the samples with respect to the value that participants assigned to completing the experiment. An analogous form of self-selection bias is discernible in Nazi Reserve Police Battalion 101 in that some men reportedly planned to become career policemen while others planned to return to their previous occupations (Browning, 1993/1998, p. 75). Those men who planned to become career policemen were cited in interrogations as being more likely to complete their missions.

Social Engineering at Lomazy

Lomazy, a town in eastern Poland, served as a concentration camp for some 1,700 local Jewish residents and deportees who were slated for execution. In August 1942, several weeks after the Józefów massacre, a company of Reserve Police Battalion 101 received orders to assist in the killing operation by rounding up the victims and delivering them first to a schoolyard that served as a collection point, then to undressing areas at the edge of the forest where clothing and valuables of the victims were collected, and finally in groups (not individually as at Józefów to minimize personal contact) to a mass grave in the forest. Because of the men's emotional reaction to the previous massacre, the shooting at the gravesite was to be carried out by a unit of auxiliary killers known as Hiwis (*hilfswilligen*, auxiliary volunteers), men from eastern European POW camps who were recruited, trained, and supervised by the SS.

The Hiwis reduced any inhibitions they may have had by consuming alcohol before and during the shootings. As they became increasingly intoxicated, they started to shoot wildly and had to be given time to sober up. The policemen were then ordered to continue the shooting temporarily while the Hiwis recovered. Several steps were taken by the commanders to reduce the emotional strain on the shooters, with the result that no man who started to shoot asked to be released. These steps and their connections to the current model are as follows.

1. *Shortening the time frame of the mission.* In addition to the fact that the policemen were required to shoot fewer victims than at Józefów, a system of rapid rotations was instituted wherein squads of eight to 10 men took rest periods after shooting five or six victims, a task requiring approximately 30 minutes to complete (Goldhagen, 1996/1997, p. 229). According to Browning (1993/1998, p. 85), under the rotation system "the men were spared the sense of unremitting, endless killing" that they experienced at Józefów. The time-frame effect is predicted by the delay discounting factor, V_{e^*}/D, in Equation 1.
2. *Increasing physical and psychological distance from the victim.* At Józefów, the men shot their victims by placing their rifles against the backs of the victims' heads. At Lomazy, they shot from a distance of approximately 27 m, with the victim lying inside the grave on one side and the shooter standing above the grave on the opposite side. (Hiwis shot at close range.) In addition, the personal tie between killer and victim was broken by having the victims forced into the grave as a group by other policemen rather than pairing up the killer with his victim as at Józefów. As a result, Browning (1993/1998) notes that in the interrogations the men usually could not recall the identities of their victims. Equation 1 incorporates distance effects through V_e, the value of immediate escape. Greater distance reduces distress, which reduces the aversiveness of the negative reinforcer and therefore the value of escaping it.
3. *Changing the demand characteristics related to withdrawal.* In contrast to Józefów, no explicit option for withdrawal was ever announced at Lomazy, and while officers would have permitted withdrawal, the shooters would have had to "exert themselves" (Browning, 1993/1998, p. 86). Equation 1 represents this counterdynamic to withdrawal by the parameter A.

4. *Habituation*. During interrogations, the men recalled the events at Lomazy without mentioning the shock, horror, and revulsion that were typical of their recollections of Józefów (Browning, 1993/1998, p. 85). It appeared that the men had habituated considerably through mere repetition of the shootings. However, their lower levels of distress also could have resulted from viewing the victims' wounds from a greater distance and by having the victims' individuality obscured through the elimination of personal contact. The model would handle habituation effects in the same way as distance effects, by reducing V_e.

In terms of the current model, the cumulative effect of the above factors would be to reduce the level of distress associated with obedience and to increase the level of distress associated with withdrawal. These are the kinds of steps that one would take to increase obedience based on the model's dynamics. However, since the techniques were confounded, it is not possible to establish which techniques (if any) were effective. Indeed, the broader historical context suggests that entirely different factors could have been involved. In particular, intense peer pressure was brought to bear on the men who withdrew at Józefów. It is conceivable that this pressure to conform, rather than any of the techniques used by the commanders, increased obedience at Lomazy.

What can be concluded with confidence about Lomazy is that the commanders—the observers closest to the situation—concluded that peer pressure was not sufficient to ensure an acceptable level of obedience; otherwise, why bring in a special unit of auxiliary killers or take elaborate precautions when the men were unexpectedly called upon to shoot? At a minimum, the current model makes explicit the processes that the commanders implicitly assumed to be relevant to their efforts to control the behavior of their subordinates.

Summary of Key Contributions

1. The current model is the only one that addresses the anomalous behavior of Milgram's defiant participants after they withdrew. It explains their failure to help the victim after they withdrew in the same terms as their withdrawal.
2. The present analysis is the first to relate the behavior of the policemen at Józefów to contemporary concepts of moral psychology and to apply the distinction between moral judgment and moral sentiment. In addition, the analysis integrates moral psychology with classical conditioning by proposing that a victim's expressions of pain elicit in the perpetrator an empathic, unconditioned response of distress. This instinctive response of distress is the basis for classical aversive conditioning of contextual stimuli, which in turn sets in motion the dynamics of escape.
3. The current analysis is the first one to connect the pattern of early withdrawals at Józefów to the pattern of early withdrawals in Milgram's experiments. Browning (1993/1998) discussed Milgram's (1974/2004) findings on proximity effects but did not make the connection between the withdrawal patterns he discovered and Milgram's break-off point distributions. The decreasing probability of withdrawal as obedience continued was one of Milgram's major findings and the basis for his theory of the *agentic state*. In the present analysis, the similar withdrawal patterns strengthen the analogy between Milgram's experiments and the Józefów massacre that is the basis for testing the current model.
4. The analysis shows how the concept of delay discounting can be applied to the interpretation of circumscribed historical episodes.
5. The current model is the only interpretation of an episode from the Holocaust in terms of behavioral processes derived from experimental research, especially the literature on choice, impulsivity, and self-control. Issues are introduced that can be assessed both experimentally and by historical analyses of

massacres from other eras, such as the My Lai massacre during the Vietnam War, an episode that Milgram (1974/2004) explicitly tied to his research on obedience.

Conclusions

Historical narratives are a form of anecdotal evidence and, unlike experiments, generally do not allow for the separation of potential causes of behavior (e.g., Lomazy). However, the Józefów account shows that the reverse can also occur. Whereas Milgram-type experiments confound moral judgment and distress, Józefów demonstrates distress without evident moral judgment. It is still possible to retain moral judgment as a proximate cause of withdrawals by having one model for Milgram-type experiments and another model for Józefów, Lomazy, and similar historical episodes of destructive obedience. But the current model, which is consistent with both experimental evidence and historical accounts, has the clear advantages of parsimony and social relevance.

Integrating psychological analysis with historical analysis offers unique opportunities for advancing understanding of destructive obedience both within and beyond these two disciplines. George Santayana famously warned, "Those who cannot remember the past are condemned to repeat it." Psychological science has a role to play in ensuring that humanity remembers—and learns from—the past.

References

AINSLIE, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82,* 463–496.

AKERLOF, G. A. (1991). Procrastination and obedience. *American Economic Review, 81,* 1–19.

ATRAN, S. (2003). Genesis of suicide terrorism. *Science*, *299*, 1534–1539.

BANDURA, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 45–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

BANDURA, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3,* 193–209.

BLASS, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology, 60,* 398–413.

BROWNING, C. R. (1998). *Ordinary men: Reserve police battalion 101 and the final solution in Poland.* New York: HarperCollins. (Original work published 1993)

BURGER, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64,* 1–11.

CESARANI, D. (2006). *Becoming Eichmann: Rethinking the life, crimes, and trial of a "desk murderer."* Cambridge, MA: Da Capo Press.

DARLEY, J. M. (1995). Constructive and destructive obedience: A taxonomy of principal-agent relationships. *Journal of Social Issues, 51,* 125–154.

DE WAAL, F. (2009). *The age of empathy: Nature's lessons for a kinder society.* New York: Three Rivers Press.

FANTINO, E. (2000). Delay reduction theory: The case for temporal context. Comment on Grace and Sevastano (2000). *Journal of Experimental Psychology: General, 129,* 444–446.

FISKE, S. T., HARRIS, L. T., & CUDDY, A. J. (2004). Why ordinary people torture enemy prisoners. *Science*, *306*, 1482–1483.

GELLER, D. M. (1978). Involvement in role-playing simulations: A demonstration with studies of obedience. *Journal of Personality and Social Psychology, 36,* 219–235.

GILBERT, S. J. (1981). Another look at the Milgram obedience studies: The role of a graduated series of shocks. *Personality and Social Psychology Bulletin, 7,* 690–695.

GOLDHAGEN, D. J. (1997). *Hitler's willing executioners: Ordinary Germans and the Holocaust.* New York: Vintage Books. (Original work published 1996)

GOLDHAGEN, D. J. (2009). *Worse than war: Genocide, eliminationism, and the ongoing assault on humanity.* New York: PublicAffairs.

GREEN, L., MYERSON, J., & MCFADDEN, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, *25*, 715–723.

GREENE, J., & HAIDT, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6,* 517–523.

HAIDT, J. (2001). The emotional dog and its rational tail: A social intuitionism approach to moral judgment. *Psychological Review, 108,* 814–834.

HARBER, K. D., ZIMBARDO, P. G., & BOYD, J. N. (2003). Participant self-selection biases as a function of individual differences in time perspective. *Basic and Applied Social Psychology, 25,* 255–264.

HERRNSTEIN, R. J. (1969). Method and theory in the study of avoidance. *Psychological Review, 76,* 49–69.

ISHII, T., & SAKAGAMI, T. (2002). Self-control and impulsiveness with asynchronous presentation of reinforcement schedules. *Behavioural Processes, 59,* 25–35.

JOHNSON, M. W., BICKEL, W. K., BAKER, F., MOORE, B. A., BADGER, G. J., & BUDNEY, A. J. (2010). Delay discounting in current and former marijuana-dependent individuals. *Experimental and Clinical Psychopharmacology*, *18*, 99–107.

KAMIN, L. J. (1956). Effects of termination of the CS and avoidance of the US on avoidance learning. *Journal of Comparative and Physiological Psychology, 49,* 420–424.

KISH, G. B., WOODY, M. M., & FRANKEL, A. (1977). The development of a scale to measure task completion motivation. *Journal of Clinical Psychology, 33,* 128–133.

LAGORIO, C. H., & MADDEN, G. J. (2005). Delay discounting of real and hypothetical rewards: III. Steady-state assessments, forced-choice trials, and all real rewards. *Behavioural Processes*, *69*, 183–187.

MCCLELLAND, D. C. (1961). *The achieving society.* Princeton, NJ: Van Nostrand.

MCGUIRE, W. J. (1976). Historical comparisons: Testing psychological hypotheses with cross-era data. *International Journal of Psychology*, *11*, 161–183.

MEEUS, W. H. J., & RAAIJMAKERS, Q. A. W. (1986). Administrative obedience: Carrying out orders to use psychological-administrative violence. *European Journal of Social Psychology, 16,* 311–324.

MEEUS, W. H. J., & RAAIJMAKERS, Q. A. W. (1995). Obedience in modern society: The Utrecht studies. *Journal of Social Issues, 51,* 155–175.

MILGRAM, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378.

MILGRAM, S. (1965a). Liberating effects of group pressure. *Journal of Personality and Social Psychology, 1,* 127–134.

MILGRAM, S. (1965b). Some conditions of obedience and disobedience to authority. *Human Relations, 18,* 57–76.

MILGRAM, S. (2004). *Obedience to authority: An experimental view.* New York: Perennial Classics. (Original work published 1974)

MILLAR, A., & NAVARICK, D. J. (1984). Self-control and choice in humans: Effects of video game playing as a positive reinforcer. *Learning and Motivation, 15,* 203–218.

MILLER, A. (2002). *For your own good: Hidden cruelty in child-rearing and the roots of violence.* New York: Farrar, Straus and Giroux. (Original work published 1980)

MILLER, A. G. (2009). Reflections on "Replicating Milgram" (Burger, 2009). *American Psychologist, 64,* 20–27.

MODIGLIANI, A., & ROCHAT, F. (1995). The role of interaction sequences and the timing of resistance in shaping obedience and defiance to authority. *Journal of Social Issues, 51,* 107–123.

MOWRER, O. H. (1947). On the dual nature of learning: A reinterpretation of "conditioning" and "problem-solving." *Harvard Educational Review, 17,* 102–148.

NAVARICK, D. J. (1982). Negative reinforcement and choice in humans. *Learning and Motivation, 13,* 361–377.

NAVARICK, D. J. (1998). Impulsive choice in adults: How consistent are individual differences? *The Psychological Record, 48,* 665–674.

NAVARICK, D. J. (2001). Control of impulsive choice through biasing instructions. *The Psychological Record, 51,* 549–560.

NAVARICK, D. J. (2004). Discounting of delayed reinforcers: Measurement by questionnaires versus operant choice procedures. *The Psychological Record, 54,* 85–94.

NAVARICK, D. J. (2009). Reviving the Milgram obedience paradigm in the era of informed consent. *The Psychological Record, 59,* 155–170.

NAVARICK, D. J., & BELLONE, J. B. (2010). Time of semester as a factor in participants' obedience to instructions to perform an aversive task. *The Psychological Record, 60,* 101–114.

NAVARICK, D. J., & FANTINO, E. (1976). Self-control and general models of choice. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 75–87.

PACKER, D. J. (2008). Identifying systematic disobedience in Milgram's experiments: A meta-analytic review. *Perspectives on Psychological Science, 3,* 301–304.

PRINZ, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9,* 29–43.

RACHLIN, H., & GREEN, L. (1972). Commitment, choice, and self-control. *Journal of the Experimental Analysis of Behavior, 17,* 15–22.

RACHLIN, H., RAINERI, A., & CROSS, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior, 55,* 233–244.

ROCHAT, F., MAGGIONI, O., & MODIGLIANI, A. (2000). The dynamics of obeying and opposing authority: A mathematical model. In T. Blass (Ed.), *Obedience to authority: Current perspectives on the Milgram paradigm* (pp. 161–192). Mahwah, NJ: Lawrence Erlbaum Associates.

ROSS, L. (1988). Review of Arthur G. Miller, The obedience experiments: A case study of controversy in social sciences. *Contemporary Psychology, 33,* 101–104.

RUNYAN, W. M. (1988). Reconceptualizing the relationships between history and psychology. In W. M. Runyan (Ed.), *Psychology and historical interpretation* (pp. 247–295). New York: Oxford University Press.

SLATER, M., ANTLEY, A., DAVISON, A., SWAPP, D., GUGER, C., BARKER, C., ET AL. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PlosS One,* 1: e39. Retrieved October 28, 2010 from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0000039

ZIMBARDO, P. (2008). *The Lucifer effect: Understanding how good people turn evil.* New York: Random House. (Original work published 2007)

ZIMBARDO, P. G., MASLACH, C., & HANEY, C. (2000). Reflections on the Stanford Prison Experiment: Genesis, transformations, consequences. In T. Blass (Ed.), *Obedience to* a*uthority: Current perspectives on the Milgram paradigm* (pp. 193–237). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix

Summary of Procedural Features of the Milgram (2004) Experiments

Basic Procedure (Voice-Feedback Group)

The "experimenter" (authority figure) instructed the "teacher" (participant) to administer a shock (elaborately simulated) to the "learner" (victim, a confederate) as punishment for every wrong answer that the learner gave on a memory test. The teacher sat facing a "shock generator" with a row of 30 switches that were labeled with voltages increasing in steps of 15 V from 15 V to 450 V. The learner, who was located in an adjacent room, communicated answers by pressing switches that illuminated numbered panels above the shock generator. After a wrong answer the teacher had to administer a shock, starting with the weakest shock for the first wrong answer and moving up the scale to the next higher voltage for each subsequent wrong answer. Beginning at 75 V, prerecorded expressions of discomfort from the learner were keyed to each lever that the teacher pressed. Whenever the teacher hesitated, the experimenter stated a scripted command (referred to as a "prod") until the teacher either continued or quit. There were four main prods of increasing severity, beginning with "Please continue" and ending with "You have no other choice. You *must* go on" (Milgram, 1974/2004, p. 21).

Table A
Key Voltages, Prerecorded Reactions, and Percent Withdrawals in the Voice-Feedback Group (Milgram, 1965b, 1974/2004: Experiment 2)

Voltage	Reaction of learner	Percentage of group (n = 40) that withdrew
75	First expression of discomfort (grunt).	0
120	Shouts that shocks have become painful. Expressions of pain intensify through 330 V.	0
150	Demands to be released ("Experimenter, get me out of here!"; Milgram, 2004, p. 23). Similar demands are made at each voltage through 330 V.	12.5
180	Cries out that the pain is unbearable.	2.5
300	Screams; states that he will refuse to give any more answers.	2.5
315	Intense scream; again states that he will refuse to give answers.	7.5
330	Screams.	0
345	Silent. Silence continues through 450 V.	2.5

Glossary of Experiments, Listed by Number and Title (Milgram, 1974/2004)

Each description highlights the essential way in which the indicated experiment differed from the voice-feedback group described above.

1, Remote. No sounds were heard from the learner except that at 300 V he pounded on the wall as an expression of protest. After 315 V he stopped answering (vs. 300 V in voice-feedback).

3, Proximity. The learner was seated in the same room and several feet away from the teacher, and he displayed visual cues of discomfort as well as auditory cues.

4, Touch-Proximity. The arrangement was the same as in the proximity condition except that to administer the shock the teacher had to place the learner's hand on a shock plate. At 150 V, in addition to demanding to be released, the learner refused to place his hand on the plate, at which point the experimenter ordered the teacher to force it into position.

5, New Baseline. As the experimenter explained the procedures to the learner, with the teacher looking on, the learner reported that he had a "slight heart condition" (Milgram, 2004, p. 56). At 150 V, 195 V, and 330 V, the learner complained that his heart was "bothering" him.

7, Experimenter Absent. The experimenter left the laboratory after he described the procedures to the teacher and learner, and he gave the teacher his orders by telephone.

9, Enters with Prior Conditions. Before the learner entered the room where he received the shocks, he informed the experimenter, with the teacher present, that he would agree to be in the experiment only if he would be released upon his request due to a heart condition. When he demanded to be released at 150 V, the experimenter ordered the teacher to pay no attention to the appeal and to continue the experiment.

10, Office Building, Bridgeport. The experiment was conducted by a fictional research organization with no stated academic affiliation in a "somewhat rundown commercial building" (Milgram, 2004, p. 68).

17, Two Peers Rebel. Two confederate teachers assisted the participant by performing the preliminary tasks in the shock sequence: One read the list of words and the other told the learner whether his answers were right or wrong. The first confederate quit after the lever was pressed at 150 V, and the second confederate quit after the lever was pressed at 210 V. Both confederates remained seated near the teacher while he assumed their duties.