Shadd Maruna and Heith Copes

Excuses, Excuses: What Have We Learned from Five Decades of Neutralization Research?

ABSTRACT

Sykes and Matza’s neutralization theory, though a popular framework for understanding deviant behavior, remains badly underdeveloped in the criminological literature. In particular, few attempts have been made to connect it to narrative and sociocognitive research in psychology and related fields. From the perspective of this wider context, it appears that one reason neutralization theory has received only mixed empirical support is that it has been understood as a theory of criminal etiology. This makes little sense (how can one neutralize something before they have done it?). It also makes the theory difficult to test. Neutralization should instead be seen as playing a role in persistence in or desistance from criminal behavior. Additionally, the theory’s central premises need to be substantially complicated. The notions that all excuses or justifications are “bad” and that reform involves “accepting complete responsibility” for one’s actions are not tenable. Sykes and Matza were right to draw attention to self-understandings and attributions, but criminology was wrong in seeing neutralization theory as the last word on cognition in criminal behavior.

And oftentimes excusing of a fault doth make the fault the worse by the excuse. (Shakespeare, King John)

Shadd Maruna is lecturer in criminology, University of Cambridge, and Heith Copes is assistant professor, Department of Justice Sciences, University of Alabama, Birmingham. We thank Michelle Butler, Moshe Hazani, Andy Hochstetler, Susanne Karstedt, Anna King, Lois Presser, Judith Rumgay, Neal Shover, Hans Toch, Michael Tonry, Volkan Topalli, Tony Ward, and three anonymous reviewers for their extremely helpful comments on a very long first draft.

© 2004 by The University of Chicago. All rights reserved.
0192-3234/2004/0032-0004$10.00
The question “Why did they do it?” is central to the criminologist’s quest, and posing it to offenders themselves has been a part of criminology since its origins (see esp. Bennett 1981). Perhaps the best known response is the account that makes up Clifford Shaw’s *The Jack-Roller: A Delinquent Boy’s Own Story* (1930). Traditionally, life stories such as these were interpreted somewhat literally by criminologists in the Chicago School as indications of the social processes that might lead to criminal behavior (see Hollway and Jefferson 2000). Shaw “made no attempt to pursue the implications of the Jack-Roller’s idiosyncratic point of view for an understanding of his involvement in delinquent conduct” (Finestone 1976, p. 101). Instead, these life histories were collected to provide “concrete and vivid” illustrations of the lives of young people in disadvantaged areas (Shaw 1929, p. 124) or to focus attention on the social factors involved in criminal etiology (Shaw and McKay 1931, p. 4). To some extent, this literalist tradition continues today in oral history research in criminology (see, e.g., Laub and Sampson 2003).

Yet, this is not the only way to interpret the accounts people give for their actions. As C. Wright Mills (1940, p. 904) suggested, “The differing reasons men give for their actions are not themselves without reasons.” The interest in life narratives among many contemporary social scientists is not so much in the substantive events these stories depict but the meanings the person attaches to such facts. How people choose to frame the events of their lives says as much about the psychology of the individual—his or her personality, identity, or self—as it does about the events and structural conditions experienced (McAdams 1985, 1993; Bruner 2002). Donald Cressey put the argument this way: “Listening to people tell you why they did it does not give you explanations of why they did it. When you ask people why they commit crime, they make sounds. I call them verbalizations. These are data. You study them” (Laub 1983, p. 139, emphasis added).

The study of offender verbalizations as “data” in criminology might have originated with Cressey’s (1953) study of embezzlers’ excuses, but it has since become associated primarily with Gresham Sykes and David Matza’s (1957) article, “Techniques of Neutralization: A Theory of Delinquency.” The important idea at the heart of this argument (hereafter, “neutralization theory”) was that the excuses and justifications that deviants use to rationalize their behaviors might themselves be implicated in the etiology of deviant behavior. As Sykes
and Matza (1957, p. 667) note, “It is by learning these techniques that the juvenile becomes delinquent.”

The influence of this creative insight has been unquestionable. Sykes and Matza’s article has been one of the most frequently cited and influential explanations of criminal behavior through the first part of the twenty-first century. According to the Social Science Citation Index, the original article was cited over 700 times between the time it was published and the end of 2003. Citations have gradually increased; it was cited an average of 9.2 times per year from 1958 to 1983, 19.6 times per year from 1984 to 1993, and 26.9 times per year from 1994 to 2003. Moreover, the central concept has been integrated into theoretical frameworks as different as learning theory (Akers 1985, p. 60), control theory (Hirschi 1969, p. 199), reintegrative shaming theory (Braithwaite 1989, p. 24), and rational choice theory (Clarke and Cornish 1985, p. 160). The theory has also influenced criminal justice innovations ranging from cognitive therapy (Ross and Fabiano 1983) to reintegrative shaming and restorative justice (Braithwaite 1989, 1999).

Perhaps the greatest testament to the importance of neutralization theory is that it is no longer confined to the study of juvenile delinquents, or even adult offenders. Neutralization theory has been used to help understand issues as serious as rape (Bohner et al. 1998), murder (Levi 1981), and genocide (Alvarez 1997), and to explain participation in less serious deviant behaviors such as playing bingo (King 1990; Chapple and Nofziger 2000), Sunday shopping among Mormons (Dunford and Kunz 1973), and entering preteen daughters into beauty pageants (Heltsley and Calhoun 2003). Neutralization theory has been used to explain how contemporary German youths avoid the stigma of the Holocaust (Hazani 1991b) and survivors of domestic violence cope with their victimization (Ferraro and Johnson 1983). Finally, somewhat oddly for a theory based on juvenile delinquency, it has found its most receptive audience in studies of organizational and white-collar crime (e.g., Benson 1985; Jesilow, Pontell, and Geis 1993; Simon and Eitzen 1993; Shover and Bryant 1993).

It is clear that neutralization theory “transcends the realm of criminology” (Hazani 1991b, p. 146). As Hazani quite accurately states, neutralization techniques are “universal modes of response to inconsistency” that reveal widely shared “modes of reduction of disequilibrium” (Hazani 1991b, p. 146). Thus, neutralization theory has “universal applicability” (Hazani 1991b, p. 135) as it can be applied
to any situation where there are inconsistencies between one’s actions and one’s beliefs.

Now that neutralization theory is nearing its golden anniversary, it is appropriate to stop and reflect. Where is the theory going, and where has it been? The world of social science research has changed immeasurably since neutralization theory was first formally outlined. Yet, neutralization theory has hardly evolved in this time. Even though neutralization techniques continue to be invoked (e.g., Laub and Sampson 2003), these discussions tend to preserve the theory in its original form rather than refine it in response to strategic research and developments in related areas. As a result, neutralizations have been widely recognized by criminologists but badly undertheorized.

This is ironic for several reasons. First, the study of crime and deviance is a natural arena for the study of explanatory accounts, excuses, and rationalizations. As Scott and Lyman (1968, p. 62) point out, “Since it is with respect to deviant behavior that we call for accounts, the study of deviance and the study of accounts are intrinsically related, and a clarification of accounts will constitute a clarification of deviant phenomena.” The scientific study of personal accounts and self-narratives has advanced substantially (see esp. McAdams’s [1999] review), but this development has been largely isolated from the study of deviance.

Second, criminology as a discipline can stake a small claim to having “gotten there first” with one of the major movements in the social sciences in the second half of the twentieth century: the so-called cognitive revolution (Gardner 1985; Baars 1986). The cognitive revolution emerged as a challenge to the prevailing behaviorist model of social functioning and its dismissal of “interior” mental processes in favor of explanations based on operant and classical conditioning (see, e.g., Chomsky 1959). According to one of the movement’s chief architects, the cognitive revolution was, in its original form,¹ “an all-out effort to establish meaning as the central concept of psychology—not stimuli and responses, not overtly observable behavior, not biological drives and their transformation, but meaning” (Bruner 1990, p. 2). The rise of cognitive and narrative-based research extended well

¹ Some say that the cognitive revolution has given way (or paved the way) for the “narrative turn” in the social sciences (see Bruner 2002; Angus and McLeod 2003). The idea characterizing this trend is that central aspects of our lives, if not life itself, are “storied”—i.e., we live in narrative and are defined and constituted by our stories of ourselves. Even critics of this narrative trend say that it has “come to dominate vast regions of the humanities and human sciences” (Strawson 2004, p. 15).
beyond psychology, with important advances made in linguistics, communication research, anthropology, neurology, theology, legal research, artificial intelligence, computer science, and philosophy.

Criminology is notably absent from that list. Yet, with its focus on inner speech, self-attributions, and locus of control, neutralization theory is (from a contemporary vantage point) one of the earliest, fully articulated sociocognitive or narrative accounts of deviant behavior. Moreover, neutralization theory was developed before or around the same time as most of the authoritative works in the cognitive movement—including Festinger’s (1957) theory of cognitive dissonance, Heider’s (1958) psychology of interpersonal relations, and seminal works on attribution theory by Jones and Davis (1965) and Kelley (1967). Neutralization theory even predated the highly influential cognitive therapies of Albert Ellis (1962) and Aaron Beck (1963), which suggested that “cognitive distortions” or thinking errors were at the root of much psychopathology. As such, neutralization theory might be considered one of the most creative and visionary (if flawed) theoretical developments in twentieth-century criminology.

However, neutralization theory’s influence has differed substantially from these other developments. Consider, for instance, Festinger’s (1957) theory of cognitive dissonance. Both theories were broadly outlined in 1957, and both focus on modes of response to inconsistency. Later versions of cognitive dissonance theory (e.g., Wicklund and Brehm 1976) focus specifically on the inconsistency central to neutralization theory: the conflict between one’s self-concept as a moral person and one’s morally questionable behaviors.

The similarities between the two theories largely end there. Here is Elliot Aronson’s assessment of the impact of Festinger’s theory on the discipline of psychology: “In 1957, dissonance theory sounded the clarion call for taking cognition seriously in social psychology; dissonance theory produced experimental research that demonstrated convincingly, like no other theory before it, that people think: we are not simple reinforcement machines. And because we think, we frequently get ourselves into a tangled muddle of self-justification, denial, and distortion” (Aronson 1992, p. 304).

Neutralization theory provides an early and widely appreciated example of how criminologists might begin to explore this same “tangled muddle” of justifications as it relates to offending behavior. Yet, subsequent work never bloomed into a full-fledged cognitive or narrative movement. Thirty years after its appearance, Clarke and
Cornish (1985, p. 160) suggested that “the cognitive revolution has passed largely unnoticed by criminologists.” Neutralization theory itself, and its famous five neutralization techniques, appear with remarkable regularity in criminological writing, but that never prompted criminology to take cognition or offender self-accounts seriously.

Criminological research examining neutralization theory since 1957 has rarely tried to incorporate even the most basic lessons from psychological research into refinement of this clearly cognitive theory (but see Agnew 1994). Likewise, discussions of what appear to be neutralization techniques among psychologists often ignore the comparatively fledgling criminological literature on neutralizations. No less a figure than Albert Bandura (1990), for instance, recently developed an important cognitive theory of “moral disengagement.” He describes the following “techniques” for avoiding self-sanction: displacement of responsibility, diffusion of responsibility, distorting the consequences of an action, dehumanizing the victim, and assuming the role of victim for one’s self. No mention is made of Sykes and Matza or the research that followed (see also the review of techniques for overcoming moral inhibitions in Anderson and Bushman [2002]).

The result has been what Howard and Levinson (1985, p. 191) call a “wasteful duplication of effort that follows from mutual interdisciplinary ignorance.”

What follows is a theoretical reappraisal of the neutralization idea in light of the five decades of research emerging out of the “cognitive revolution” in social science. We hope to show that neutralization theory, like many other fifty-year-olds in criminology, can be as relevant to criminological research today as it ever was. Yet, to remain relevant, it needs to catch up with the times, be more flexible, admit to its shortcomings, and accept that it cannot explain all of the deviance phenomenon by itself.

In the pages that follow, we argue that neutralization theory probably fails as an explanation of “primary deviations” (Lemert 1951) or
criminal etiology. There is little empirical evidence that individuals ascribe to neutralizations in advance of behaving criminally, and it is difficult to imagine how evidence of this could be reliably collected. Further, the psychological research on account making and self-attributions is clear: there is nothing pathological about neutralizing negative behaviors or contextualizing one's faults in broader circumstances. Taking full responsibility for every personal failing does not make a person "normal," it makes them extraordinary (and possibly at risk of depression).

We argue, however, following Hirschi (1969) and others, that neutralization techniques may play an important role in maintaining persistence in crime or Lemert's "secondary deviance." This conceptual shift would allow researchers to overcome what has been the thorniest methodological problem to date: how to measure the acceptance of neutralization techniques prospectively rather than simply in retrospect. It also makes sense theoretically in terms of the psychological literature, situating neutralization techniques as part of the narrative process through which individuals make meaning out of their lives.

If neutralization theory is to survive, it must develop a far more nuanced understanding of cognition. For instance, excuses and justifications that rely upon highly stable and global attributions (e.g., "That is just the way the world works"); "This is just who I am") and attributions of a hostile nature (e.g., "It is because everyone is against me") are most likely to be associated with persistent criminality. Excuses that separate past offending behavior from the individual's "core self" (e.g., "It was a complete accident") may be more commonly associated with maintaining desistance from crime. These remain important, lingering questions, not just for their value in working with ex-offenders but also because they allow us better to understand desistance or persistence from the individual offender's point of view.

We begin our review by dissecting Sykes and Matza's formulation, tracing its origins to the work of Mills, Cressey, Sutherland, and Redl, and discussing the influence that neutralization theory has had on criminology and criminal justice practice. We suggest that neutralizations are included in almost every major theoretical tradition in criminology (sometimes as little more than an afterthought) and play a substantial role in the applied world of corrections. Yet, the richness of the theory in its original formulation (especially its roots in Redl's psychoanalytic work) is often missed.
In Section II, we attempt to situate the neutralization idea in the broader context of social psychological research on attribution, explanatory style, cognitive dissonance, and, most importantly, narrative psychology. It becomes clear that the treatment of neutralization techniques as automatically “bad things” in criminology and corrections is an oversimplification of a complex and substantial body of literature.

This mischaracterization probably accounts for the mixed and inconsistent support neutralization has received in the empirical literature. The empirical evidence is reviewed in Section III, where we analyze the strengths and weaknesses of existing work. This research tends to fall in one of two camps: illustrative qualitative studies or survey-based efforts to test the theory. We argue in favor of taking advantage of the strengths of both approaches—the validity and depth of qualitative interviewing, and the rigor of the hypothesis-testing approach. Such designs become possible when neutralization theory is reconceived as an explanation of persistence or desistance rather than of criminal etiology. We make this case in Section IV and review the scattered bits of evidence in favor of this revision of the theory.

Finally, we conclude with some suggestions for the next era of neutralization research. Section V reviews things we already know (or should know) about neutralization theory. These are lingering confusions and arguments that should have been largely settled by now, but continue to haunt current research, for instance, that not all neutralizations are “bad,” and that the relationship between neutralizing and offending is probably not a causal one. In Section VI, we outline those things that we still do not know, but should, regarding neutralization theory. These are the more substantial questions that we hope will characterize the next generation of neutralization research.

I. Overview of Neutralization Theory

Sykes and Matza’s (1957) influential article began with a critique of subcultural theorists of the time. Subcultural theorists, such as Cohen (1955), argued that delinquent boys rebelled against the dominant social order by rejecting middle-class standards and replacing them with a new, often delinquent, set of values. Sykes and Matza (1957) disagreed, contending that subcultural theorists overstated the extent to which delinquents rejected conventional values. They believed that
everyone, even lower-class delinquent gang members, retained some commitment to the dominant value system of society.

They based their argument on four key points. First, if delinquent subcultures do exist, then delinquents should view their criminal behavior as morally correct. Therefore, they should not experience guilt or shame for engaging in the act or for being caught. Second, delinquents should value the opinions and lifestyles of those promoting similar delinquent lifestyles and dismiss the opinions of conventional others. Third, if offenders unconditionally accept crime, we would expect them to treat all victims equally. Fourth, offenders should be immune to the demands of conformity. Critiquing each of these claims, Sykes and Matza argue that delinquents do often feel guilt and shame for participating in illegal behaviors; show respect and admiration for honest, law-abiding others; make clear distinctions about who can and cannot be victimized; and participate in the same social functions that law-abiding citizens do (including church, school, and family activities). Together, these factors suggest that delinquents do distinguish between right and wrong and are subject to influences of both conventional and delinquent subcultures. According to Sykes and Matza, young offenders are well aware of the wrongfulness of their actions.

Yet, if delinquents maintain at least minimal commitments to the dominant social order, as Sykes and Matza claim, how are they then able to violate its norms? If people are committed to the social order, they typically experience guilt or shame for violating, or even contemplating violating, social norms. This guilt, and its potential for producing a negative self-image, helps dissuade us from engaging in criminal or deviant acts most of the time. Therefore, in order to participate in deviant behavior under such conditions, we must find ways to rationalize the actions or neutralize the guilt associated with it. Sykes and Matza argue that “much delinquency is based on what is essentially an unrecognized extension of defenses to crimes, in the form of justifications for deviance that are seen as valid by the delinquent but not by the legal system or society at large” (Sykes and Matza 1957, p. 666).

1 Sykes and Matza use these terms interchangeably (1957, p. 664), so we follow suit. A substantial theoretical literature has developed around the differentiation of shame and guilt (e.g., Tangney 1995), but subsequent empirical research has "provided only limited support" for this conceptual distinction (J. Braithwaite and V. Braithwaite 2001, p. 75).
These linguistic devices, when invoked by offenders, blunt the moral force of the law and neutralize the guilt of criminal participation. Essentially, one can maintain a sense of morality and violate it too by “reconstruing the conduct, obscuring personal causal agency, misrepresenting or disregarding the injurious consequences of one’s actions, and vilifying the recipients of maltreatment by blaming and devaluing them” (Bandura et al. 1996, p. 364). Through the use of these techniques, social and internal controls that serve to check or inhibit deviant motivational patterns are “rendered inoperative, and the individual is freed to engage in delinquency without serious damage to his self-image” (Sykes and Matza 1957, p. 667). The offender can remain “committed to the dominant normative system and yet so qualifies its imperatives that violations are ‘acceptable’ if not ‘right’” (Sykes and Matza 1957, p. 667). What is culpable can be “made righteous through cognitive reconstrual” (Bandura et al. 1996, p. 365).

Sykes and Matza (1957, p. 669) argue that all of these neutralizations are “extensions of patterns of thought prevalent in society rather than something created de novo” (see also Matza and Sykes 1961). Matza (1964) later argued that delinquent neutralizations are legitimated by a most unlikely source—the juvenile justice system itself. When agents of convention, from social workers to judges, argue that delinquents are the helpless products of their environment, they unwittingly contribute to the internalization of neutralizing excuses. Moreover, neutralization should not be confused with outright deceit. Mills (1940, p. 907) writes, “Verbalizations are not lies merely because they are socially efficacious.” Every event is subject to multiple interpretations. As Mills pointed out, one person’s rational explanation is another’s rationalization. If neutralizations are to carry any psychological weight, they must, at least partially, be believed by the person using them.

Finally and most important, Sykes and Matza (1957, p. 666) claim that neutralizations “precede deviant behavior and make deviant

[Neutralization techniques share many of the same properties as rationalizations, “verbalizations,” accounts, motives (in Mills’s sense of the word), micronarratives, explanations, minimizations, stories, self-attributions, excuses, and justifications. There is also considerable overlap between this concept and psychological notions such as denial, self-deception, “moral disengagement” (Bandura 1990), and “techniques of ego defense” (Redl and Wineman 1951). To avoid confusion, we mostly use “neutralization” when describing verbalizations designed to relieve the speaker of culpability or censure and the more generic “account” to describe “a statement made by a social actor to explain unanticipated or untoward behavior” (Scott and Lyman 1968, p. 46) when exculpability is not implied. In places, we use the other, near synonymous terms listed above either because we are quoting from other sources or else as a means of avoiding repetition.]
behavior possible.” In this brief phrase, they make two crucial claims that are often overlooked in empirical work. First, they suggest a specific chronological sequence: neutralizations are not just a posteriori rationalizations; they precede delinquency and make deviant behavior possible. Second, and just as important, they emphasize that this order is not meant to imply a deterministic or causal relationship. Neutralization “enables crime but does not require it” (Minor 1981, p. 300).

Matza (1964, p. 29) develops this argument much more explicitly in his later solo work with his concept of “drift.” He defines “drift” as a temporary period of irresponsibility or an episodic relief from moral constraint. Neutralization enables drift by freeing the individual from the moral bind of law and order. Once in a state of drift, Matza argues, a young person is likely willfully to choose to commit a crime under circumstances of “preparation” (or familiarity with the particular offense type) or “desperation.” Matza’s concept of desperation is linked to the delinquent’s central neutralization technique, the denial of responsibility, or what is referred to as a “mood of fatalism.” In the mood of fatalism, common to the experience of drift, delinquents believe that they have been “helplessly propelled into new situations” like a “billiard ball” (Sykes and Matza 1957, p. 667). This feeling of helplessness simultaneously relieves the individual from the binds of morality and also encourages the delinquent to want to “take control” of his or her situation and prove that he or she can “make something happen.” Considering the limited options available to adolescents, this frequently means committing a new type of offense in order to regain a sense of being in control of the environment (see Brezina [2000] for an innovative analysis of this aspect of Matza’s theory).

A. The Famous Five Neutralizations (and Beyond)

Sykes and Matza (1957) outlined five techniques of neutralization that allow offenders to engage in wrongdoing: denial of responsibility, denial of injury, denial of the victim, condemnation of condemners, and the appeal to higher loyalties. The subsequent expansion of the theory to different types of offenders and offenses has led to the “discovery” of several new neutralization techniques. Schönbach (1990) devoted nearly an entire book to cataloging the different accounts that individuals offer for their actions.

As Cohen (2001, p. 61) and others have pointed out, the denial of responsibility is clearly the “master account.” If offenders can define
the situation in a way that relieves them of responsibility for their actions, then they can mitigate both social disapproval and a personal sense of failure. Offenders deny responsibility by claiming their behaviors are accidental or due to forces beyond their control. They see themselves as victims of circumstance or as products of their environment. Psychotherapists who engage in improper sexual relations with clients, for instance, tell “sad tales” about mounting family and personal problems that diminished their ability to cope with the situation properly, causing them to act inappropriately (Pogrebin, Poole, and Martinez 1992). For corporate offenses, the hierarchical structure of organizations diffuses responsibility for actions that may harm consumers or the public. For example, when asked to inform the chief engineer of the reporting of false documents at B. F. Goodrich, Vandivier’s superior told him, “I learned a long time ago not to worry about things over which I have no control. I have no control over this... Why should my conscience bother me?” (cited in Vandivier 1996, p. 216). Organizational offenders also deny their responsibility by claiming they were “forced” to engage in the illegal acts to keep their jobs or for their businesses to survive.

A second technique, denial of injury, focuses on the extent of harm or injury caused. The wrongfulness of one’s behavior is determined by the amount of harm done and by the intentions of the actor. Offenders can excuse their behavior if they believe no one is “really” harmed. Auto thieves, especially young joyriders, claim that they were merely “borrowing” the vehicle (Copes 2003). They argue that since the vehicle is brought back to the rightful owner no true harm took place. Other auto thieves argue that vehicle owners have insurance, or at least should, and can easily replace the vehicle at little to no cost. Similarly, marijuana users claim that smoking is neither socially destructive nor harmful to the individual (Peretti-Watel 2003). Many marijuana smokers contend that the world would be a better place if more people smoked (Priest and McGrath 1970). Offenders who use these techniques may believe or state that their behavior is inappropriate in general but in this particular instance it is acceptable because no real harm was done.

Sometimes offenders admit that their actions cause harm but neutralize moral indignation by denying the victim. This can be done in one of two ways. First, one might contend that some victims act improperly and thus deserve everything that happens to them. Offenders define their own actions as a form of rightful retaliation or punishment, thereby claiming the victim does not deserve victim
status. Delinquents may see their victims as wrongdoers and perceive themselves as avengers of the wronged. Some robbers target drug dealers as a form of street justice or “righteous retribution for the destruction dealers wrought on persons and entire communities by hawking their evil wares” (Jacobs 2000, p. 33). Denial of the victim also occurs if the victim is absent, unknown, or abstract. In these situations the offender can easily ignore the rights of victims because the victims are not around to stimulate the offender’s conscience. Many employees are able to insulate themselves from guilt associated with theft by defining the objects they steal as being of “uncertain ownership” (Horning 1970). Items found on company property that are not clearly owned by other employees or the company can be taken without the accusation of harm or theft (Dabney 1995).

The fourth technique involves the condemnation of the condemner. Instead of focusing on their own actions, delinquents shift the focus of attention to the motivations or behaviors of the people expressing disapproval. One might claim that their critics “are hypocrites, deviants in disguise, or impelled by personal spite” (Sykes and Matza 1957, p. 668) and have no right to pass judgment on others. Delinquents accuse police of being corrupt and teachers of being unfair. Deer poachers often argue that game wardens frequently hunt illegally and should not be given the authority to make arrests or issue citations (Eliason and Dodder 1999). Doctors convicted of defrauding Medicaid point to irrationality in the health care system (Jesilow, Pontell, and Geis 1993).

The final technique described by Sykes and Matza is the appeal to higher loyalties. Offenders neutralize internal and external controls by claiming that their behaviors are consistent with the moral obligations of a specific group to which they belong. This does not imply that offenders reject the norms they are violating. Instead, other norms are seen as more pressing or deserving of precedence. This technique is common among young members of fraternities or gangs who put loyalty to their peer groups above all else. It is also common among corporate offenders who argue that their actions were conducted for “higher” goals including profit for their stockholders and financial stability for their families. Others, usually lower in the organizational hierarchy, might blame family responsibilities for their participation in illegal behaviors. Accounting for an unethical decision he made while working for B. F. Goodrich, one worker argued, “Hell, I’ve got two sons to put through school” (cited in Vandiver 1996, p. 215).
Sykes and Matza’s original list of five offender justifications was not the final inventory, nor should it be. Scholars have expanded greatly on these initial techniques. It appears that “new” techniques emerge with each new exploration into a deviant group. Qualitative studies of white-collar offenders have produced several new techniques including the defense of necessity, the claim of normality, and the claim of entitlement (Benson 1985; Coleman 2002; Conklin 2004). Studies of property offenders have introduced the techniques of the metaphor of the ledger (Klockars 1974), justification by comparison, and postponement (Cromwell and Thurman 2003). This list no doubt will lengthen as research in the area continues.

B. Origins of Neutralization Theory

Neutralization theory is the product of a rich and somewhat mixed lineage. Although it emerged out of the “uniquely American tradition of symbolic interactionism” (Hamlin 1988, p. 425), it can also trace its origins to the more Austrian-flavored psychoanalytic work of Fritz Redl, August Aichorn, and Anna Freud. Sykes and Matza do not include an extensive literature review in their short article. They refer no more than eleven previous works, only two of which centered on offender accounts. Most obviously, they fail to cite C. Wright Mills’s (1940) work on “vocabularies of motive” or Kenneth Burke’s Rhetoric of Motives (1950). The predecessors they do acknowledge are Sutherland and, in back-to-back footnotes, Cressey’s (1953) Other People’s Money and Redl and Wineman’s Children Who Hate (1951).

Sykes and Matza’s concept builds directly on Sutherland’s (1947) notion of “definitions favorable to violations of law,” but argued that too little effort had been made to delineate the “specific content” of these definitions (Sykes and Matza 1957, p. 664). Cressey (1953) started to catalog the substance of these definitions in his work on embezzling, however. Cressey found that the typical embezzler described his offense in a language that enabled him to look upon his trust violation as essentially noncriminal, justified, or else a part of a “general irresponsibility for which he is not completely accountable” (p. 93). Cressey stated that the white-collar offender utilizes this “vocabulary of adjustment” in order to maintain “conceptions of himself as a trusted person” (p. 94). He posited that although these verbalizations were clearly used as an ex post facto justification for behavior, it was equally possible that “the person may prepare his rationalization before he acts” (p. 94). Essentially, he argues, “the rationalization is his motivation” (p. 94).
Cressey grounded his argument in the basic tenets of symbolic interactionism (esp. Mead 1934) and the work of C. Wright Mills (1940) on “vocabularies of motive.” The symbolic interactionists’ central mantra, “if [people] define situations as real, they are real in their consequences,” has, since its inception, linked deviant behaviors to the situated worldviews of offenders. Similarly, Mills (1940, p. 905) developed the bold argument that “motives are words,” that is, “the vocalized expectation of an act, its ‘reason,’ is not only a mediating condition of the act but it is a proximate and controlling condition for which the term ‘cause’ is not inappropriate” (p. 907). Rather than invoking “psychological terms… as explanatory,” Mills argued that motivation could be understood as the anticipated answer to the “question” (of why) that follows our actions. Arguing against what might now be seen as a “rational choice” explanation of behavior (that individuals make choices on the basis of the “differential consequences which they anticipate”), Mills argued that “It is more adequate to say that individuals act in terms of anticipation of named consequences” (pp. 905–6). The “real motive” (or deeper, underlying reason for an action), according to Mills, “is not something different in kind from the verbalization or the ‘opinion.’ They turn out to be only relatively and temporally different” (p. 909).

Although failing to recognize the parallels to Mills’s groundbreaking work on the sociology of motives (or to Mills’s predecessor Kenneth Burke), Sykes and Matza acknowledge their argument’s debt to the psychoanalytically oriented work of Fritz Redl. Although the sociologist Mills might be somewhat uneasy with this link, the parallels between neutralization theory and Redl’s research on ego defense mechanisms is (pardon the pun) undeniable. Redl and Wineman (1951) sought to analyze both the “ego strengths as well as ego disturbances” of a sample of delinquents in their care at a residential treatment program. Their examination is largely a corrective on earlier psychoanalytic work, which assumed that delinquents had weak or deficient egos and/or underdeveloped or nonexistent superegos. To the contrary, Redl and Wineman found that their clients had “hyperfrothically developed” ego skills that are simply “in the service of

5 In the original passage containing this famous phrase, Thomas and Thomas (1928, p. 572) are describing the criminal career of a convict at Dannemora prison. The prisoner had murdered several persons because they “had the unfortunate habit” of talking to themselves on the street. “From the movement on their lips he imagined that they were calling him vile names, and he behaved as if this were true.”
the wrong cause.” Far from suffering from pathological deficits in cognition, “what [delinquents] seem to ‘know’ about people and how to handle them—including [treatment practitioners]—often puts current research in such matters to shame” (pp. 181–82).

Redl and Wineman (1951, pp. 145–46) make almost precisely the same argument that Sykes and Matza (1957) would later develop. They write, even the “ego of the toughest delinquent” has “quite a job to perform in order to keep all phases of that behavior from being ‘tax exempt’ from feelings of guilt” (1951, p. 147). Writing three decades prior to the initial discussion of “cognitive distortions” in criminal thinking and six years prior to Sykes and Matza, Redl and Wineman describe, in elaborate detail, the “system of delusions” (1951, p. 146) or “techniques of a delinquent ego” that seem to characterize young offenders. This taxonomy included familiar-sounding defenses such as “He did it first,” “Everybody else does such things anyway,” “We were all in on it,” and “He had it coming to him.” These so-called guilt evasion tricks, Redl and Wineman conclude, are often “the only tangible evidence for the existence of values, for the duping of which the whole machinery had to be invented to begin with” (1951, p. 147).

Sykes and Matza’s formulation owes a substantial debt to this work, but the psychoanalytic influence on neutralization theory is all but forgotten in later work on the theory (Barriga et al. 2000 is one exception). Yet, ego psychology, unlike neutralization theory, provides a coherent framework for explaining how neutralizations work. To the ego psychologist, defense mechanisms like denial, projection, reaction formation, and sublimation defend the integrity of the ego when internal or external events occur that violate a preferred view of the self (A. Freud 1936; Fenichel 1945). This research also “locates” neutralizations securely outside of conscious thought. “Like self-deception generally, defense mechanisms must involve some motivated strategy that is not consciously recognized, resulting in a desirable conclusion or favorable view of self that is conscious” (Baumeister, Dale, and Sommer 1998). When later readings of neutralization theory interpret these techniques as purely conscious and deliberate maneuvers, performed for a specific audience, neutralizations lose something important.

Redl and Wineman’s work is still influential in psychoanalytic work and in research involving adolescents in group care but has been neglected in criminology (but see Toch 1997).
C. Neutralization Theory’s Place in Criminology

Neutralization theory is usually understood as a single component of a larger theory. Alone, the theory provides no sufficient explanation for differences in crime across cultures, groups, genders, or the like. Matza is explicit on this point, arguing that there is no point to research that isolates neutralization acceptance alone and out of the context of the other aspects of an individual’s life that contribute to delinquency: “It makes little sense to take each element out of context, to gaze at it and to reject it because it does not significantly differentiate delinquents from other boys. That the subcultural delinquent is not significantly different from other boys is precisely the point. He is marginally different and only in process is there a cumulation sufficient to sometimes culminate in infraction” (Matza 1964, p. 89).

The theory’s value is rightly understood as enhancing or developing existing theoretical frameworks for understanding offending. Neutralization theory has been linked to so many different wider traditions of criminological thought over the years that it is difficult to know how to classify it. Introductory textbooks consider it variously as a part of control theory (Williams and McShane 2004), psychological theories (Bohm 2001), learning theory (Lanier and Henry 1998), and subcultural theory (Winfree and Abadinsky 2003). For good measure, Hirschi (1969, p. 24) suggests that the accounts by Cressey (1953) and Sykes and Matza (1957) can be understood as strain theories. In addition, neutralization techniques have been incorporated into reintegrative shaming theory (Braithwaite 1989), rational choice theory (Cornish and Clarke 1986), and even as a small component of life course theory (Laub and Sampson 2001, p. 164), leaving very few areas of contemporary criminological theory outside its reach.

As initially proposed, Cressey’s (1953) and Sykes and Matza’s (1957) writings on neutralizations were seen as an extension and refinement of Sutherland’s differential association theory. Sutherland argued that, through interacting with others, offenders learned not just the techniques of crime but also the definitions (i.e., motives and rationalizations) favorable to crime. Sykes and Matza argued that researchers had ignored the content of what was learned, preferring instead to focus on the process by which delinquency was learned. Thus, their techniques of neutralization were thought to “make up a crucial component of Sutherland’s ‘definitions favorable to violation of law’” (Sykes and Matza 1957, p. 667).
Eventually neutralization theory began to be viewed as more than a refinement of differential association theory and became an independent theory of crime and deviance. Matza’s (1964) drift theory was instrumental in this process as neutralization took a primary role in the theory. The incorporation of neutralizations into Matza’s theory of delinquency and drift led others to classify neutralization theory as a component of control theory (e.g., Lilly, Cullen, and Ball 2002). For instance, Ball (1966) argues that neutralization theory should be considered as one component of Reckless’s (1961, 1967) containment theory or Reckless and Shoham’s (1963) norm erosion thesis. Containment theory argues that refraining from criminal behavior requires a blend of self factors (inner containment) and social factors (outer containment). Strong inner and outer containments insulate individuals from becoming involved in crime. Reckless and Shoham argue that “norm erosion”—ignoring the moral significance of norms, the neutralization of what “ought” to be done, and emancipation from internalized norms—is an important factor in the breakdown of inner containments.

To a lesser degree, neutralization theory has been incorporated into the writings of rational choice theorists (Clarke and Cornish 1985; Cornish and Clarke 1986). Contemporary rational choice theorists have moved away from early economic models, preferring models of behavior that recognize bounded decision-making processes. Rational choice theorists now devote much of their time to modeling the various stages of criminal decision making, including initiation, continuance, and desistance (Cornish and Clarke 1986). Neutralization is thought to play a significant part in the decision-making process at each of these stages, and therefore investigators frequently take them into account when modeling criminal decision making.

The theory is also firmly established within work dealing with account making in sociology. Scott and Lyman’s (1968) sociology of accounts, for instance, borrowed heavily from neutralization theory. Accounts can also be seen as an important refinement of the original neutralization formulation, although it is not always incorporated into contemporary discussions. A brief review of Scott and Lyman’s

---

7 There have been numerous sociological perspectives that explain aligning action, including Mills’s (1940) motive talk, Scott and Lyman’s (1968) accounts, Hewitt and Hall’s (1970, 1973) quasi-theories, Hewitt and Stokes’s (1975) disclaimers, Hunter’s (1984) acclaimers, and Pestello’s (1991) discounts.
accounts is presented here because of the undeniable influence it has had on neutralization theory and research.\(^8\)

Accounts are forms of “aligning actions” (Stokes and Hewitt 1976, p. 838) that are meant to “verbally bridge the gap between action and expectation” when an individual behaves in a way that is inconsistent with normative expectations. Scott and Lyman make an important distinction between two types of accounts: justifications and excuses. They define justifications as “accounts in which one accepts responsibility for the act in question, but denies the pejorative quality associated with it” (1968, p. 47). Excuses are “accounts in which one admits that the act in question is bad, wrong, or inappropriate but denies full responsibility” (1968, p. 47).

Sykes and Matza’s techniques of neutralization make up a large part of the justifications described by Scott and Lyman. They contend that denial of injury, denial of victims, condemnation of condemners, and appeal to higher loyalties can be viewed as a “tentative list” of types of justifications (Scott and Lyman 1968, p. 52). The remaining (and probably most central) technique, “denial of responsibility,” is included in their schema under “appeal to defeasibility,” which Scott and Lyman consider an excuse. Subsequent research suggests that actors use justifications and excuses depending on the deviant act they are engaging in. For example, actors tend to provide justifications for violent offenses (Felson and Ribner 1981; Henderson and Hewstone 1984), but excuses for property crimes. Violent crimes are often the product of a dispute between two parties, and offenders frequently interpret their role as one of self-defense or a reasonable reaction to hostile provocation. Property crimes can rarely be interpreted in this way, and so are more frequently excused.

\(D.\) Policy Applications of Neutralization Theory

Neutralization theory is intended to help explain the occurrence of certain kinds of deviant and criminal behavior; it is natural that its adherents see implications for criminal justice or correctional policy. This was not true of all of the theory’s originators. Cressey, for instance, argues that his ideas about verbalizations have “few practical implications either for prevention and detection of trust violation or

\(^8\) For a more thorough overview, see Orbuch (1997) and Fritzsche (2002). Schönbach (1990) has expanded Scott and Lyman’s classification scheme with his nearly exhaustive taxonomy of accounts.
for treatment of apprehended offenders” (1953, p. 153). Some implications of the theory are too broad or too vague to be of much practical use. Bohm, for instance, states, “A policy implication of neutralization theory would be to delegitimate neutralizations, that is, make them unacceptable. One way that might be accomplished is by reducing or eliminating social injustices and double standards” (Bohm 2001, p. 52). Nonetheless, the theory has found its way into the rationales (rationalizations?) of numerous criminal justice innovations in recent years, including restorative justice conferencing, situational crime prevention, and correctional therapy.

Restorative justice interventions, such as family group conferencing—where offenders sit down with family members, community elders, and their victims in a reintegrative shaming process—are largely premised on social-cognitive principles, with the explicit aim of undermining offender neutralizations. John Braithwaite, for instance, frequently explains the social psychological basis behind restorative justice (Braithwaite 1999, pp. 47–51) by reference to neutralization theory: “At a victim-offender mediation or conference when the victim is present, it is hard to sustain denial of victim or denial of injury. . . . Condemnation of the condemners is also more difficult to sustain when one’s condemners engage in a respectful dialogue about why the criminal behavior of concern to them is harmful. . . . [Finally,] appeals to higher loyalties like loyalties to one’s mates [can be addressed by bringing] together the audiences the criminal would most want to be segregated [e.g., one’s peers and one’s parents]” (Braithwaite 1999, pp. 47–49). Even more directly, Thomas Scheff (1998, p. 105) argues that one of the great advantages of mediation is in “making sure that all of the shame connected with the crime is accepted by the offender . . . acknowledging his or her complete responsibility for the crime.”

Neutralization ideas have also been used in developing crime prevention programs. The theory is that by learning the linguistic devices that offenders use to make their crimes palatable, program designers can actively attack these belief systems. By “neutralizing the neutralizations,” potential offenders would not be able to define their actions as noncriminal and thus would refrain from criminal behavior (Clarke 1997; Clarke and Homel 1997). True to situational crime

---

9 True to his roots in differential association theory, Cressey suggests that since verbalizations are learned from others who define some situations as appropriate for crime, deviant behavior can be minimized by isolating persons from individuals sharing such definitions.
prevention’s roots, “removing excuses” does not entail making long-term changes in the disposition of the offender. Instead, situational crime prevention theorists argue that programs geared toward removing excuses should still focus on highly specific forms of crime and should be presented at the time criminal decisions are being made. The idea is to “stimulate feelings of conscience at the point of contemplating the commission of a specific kind of offense” (Clarke 1997, p. 24).

Researchers have suggested numerous programs to help reduce crime based on the “removal of excuses.” Thurman, St. John, and Riggs (1984) found that neutralizations used to justify tax evasion can block the potential inhibiting effects of guilt. They suggest that campaigns designed to make tax cheaters feel guilty about their behaviors can reduce the prevalence of tax fraud. Bohner and colleagues (1998, p. 266) suggest that the prevention of some forms of crime, such as rape, can be prevented by exposing cultural stereotypes that rapists use as neutralizations and by replacing these myths with interpretations of reality that are not as conducive to crime. Most commonly, though, these interventions are targeted to stop deviant behavior that occurs within formal organizations such as workplaces and schools (Pelfrey 1984; Greenberg 1990; Lim 2002). For instance, organizational managers are encouraged to discuss openly the neutralizations that wayward employees use. Bringing these neutralizations into the open is thought to force employees to consider consciously their actions when stealing from the company (Cressey 1953).

Finally, nowhere is the impact of neutralization theory more pronounced than in the psychological treatment of offenders. Confession is central to “almost all of the proliferating systems of psychotherapy and counseling” (Rose 1996, p. 96). Nearly every form of offender treatment—from the “12 steps” model of Alcoholics Anonymous to the confrontational techniques of therapeutic communities—involves strategies for “overcoming denial” and challenging offender rationalizations.

The ascendancy of cognitive-based treatment in correctional settings (e.g., Bush 1995; Gibbs, Potter, and Goldstein 1995; McGuire 1995) has effectively guaranteed a second life for neutralization theory. In terms of “what works” in recidivism reduction, meta-analysts like Gendreau, Goggin, Cullen, and Andrews (2000, p. 13) have reached the “inescapable conclusion” that cognitive-based interventions may be the “only game in town.” Cognitive-based programs—with titles like “Reasoning and Rehabilitation” (Ross and Fabiano 1985), “Mind
over Matters: Corrective Thinking Treatment Model” (Tru-Thought 2000), and “Thinking for a Change” (Bush, Glick, and Taymans 1997)—have “dominated practice developments” in correctional programming in the past decade (Vanstone 2000).

The premise behind much cognitive programming owes a considerable debt to the neutralization idea: offending is partially facilitated by a cognitive mindset that justifies and rationalizes criminal behavior (see, e.g., Sharp 2000). 10 For instance, White and Walters argue that offenders suffer from a “psychology of disresponsibility” or “the intellectual process by which a person’s actions are attributed to factors other than the person himself” (1989, p. 259). Further, as in many other formulations, White and Walters lay much of the blame for this sense of “disresponsibility” on the shoulders of social science for providing offenders with ready-made, socially sanctioned excuses for their behavior (see also Matza 1964).

The best developed and most sophisticated research and implementation in this area has been in sex offender treatment (e.g., Stermac and Segal 1989; Polaschek, Ward, and Hudson 1997; Ward et al. 1997; Marshall et al. 2001). However, some models of “criminal thinking” are thought to apply more broadly to the general offender population. In their multivolume work on The Criminal Personality, psychiatrists Samuel Yochelson and Stanton E. Samenow (1976) listed fifty-two thinking errors associated with offending. John C. Gibbs and his colleagues (Gibbs, Potter, and Goldstein 1995; Barriga et al. 2000) developed a more coherent and rigorously evaluated framework of criminal thinking patterns including “blaming others,” “minimizing” or mislabeling, and “assuming the worst” (see also Slaby and Guerra 1988). Clearly, there are “certain potential overlaps” between these beliefs and neutralization techniques (McGuire 2000, p. 36).

Drawing from the cognitive therapies of Beck (1963) and Ellis (1962), therapists working with groups of offenders are trained to focus on self-referent “inner speech,” elicit “thinking reports,” and challenge offender attributions. The focus of cognitive restructuring work is on understanding the relation between thinking, beliefs, and evaluations in the offense process or chain. The aim is to understand an individual’s goals and what exactly he or she is seeking via offending.

10 In this framework, “cognitive distortions” (or sometimes “dysfunctional thoughts,” “cognitive deficits,” or “irrational beliefs”) are the terms used to describe transgressions from normative information processing or processing filters that are thought to be distorted or maladaptive. Like neutralization theory, it is not known if these biases are the cause or consequence of pathological behavior.
behavior. There is more of an attempt to tease out the broad views that individuals have of the self, others, and the social world and the way these are embedded in practices and ways of living (see Polaschek, Ward, and Hudson 1997; Ward 2000). Considerable evaluation evidence suggests that such interventions can be effective in changing cognitive patterns (Leeman, Gibbs, and Fuller 1993).

In practice, however, cognitive therapeutic encounters often become bogged down in confronting and challenging the excuses and justifications offered by offenders (Kendall and Pollack 2003). In many cognitive-based programs, excuses for offending are “seen as unacceptable responses which need to be discouraged and replaced with internal, stable, global attributions of cause and responsibility-taking” (Beech and Mann 2002, p. 265). Walters (1998, p. 67), for instance, advocates interventions based on “confronting rationalizations with facts and self-deception with feedback.” Likewise, Sharp (2000, p. 3) writes, “We believe that optimum opportunity for success in a treatment program requires that clients be held accountable for all their actions, past, present and future.”

In an ethnographic study of one prison-based cognitive treatment program, Kathryn Fox argues that “the obligation to confess” is used as a form of “cognitive social control” (Fox 1999a, p. 91). She found that the “somewhat sociological” accounts used by prisoners to explain the criminal violence they committed in the past were rejected by therapists as an example of “criminal thinking” and replaced by “ideology of moral autonomy” (Fox 1999b, p. 442). “These accounts are regarded as ‘cognitive distortions’ by (treatment) program logic and framed instead as evidence of extraordinary pathology” (Fox 1999b, p. 436). Alternatively, treatment discourse worked to decontextualize inmates’ past actions, leaving the individual with little choice but to accept the dominant therapeutic discourse of pathology.

Indeed, Fox writes, the following was listed as a “thinking error” in a workbook for a cognitive treatment program for prisoners: “The criminal believes that he is a good and decent person. He rejects the thought that he is a criminal” (cited in Fox 1999b; see also chap. 10 of Samenow 1984). This becomes something of a “catch-22” for treatment participants—if they claim to be decent, that is proof that they are criminally minded; if they admit to being criminally minded, that also is proof that they are criminally minded (see also Beech and Mann 2002).

Sykes and Matza (1957) might have been shocked to learn their theory would end up serving as a potential rationale for programs like
Yet, without this development in the applied world of corrections, interest in neutralization theory might have been more limited than it has been. One reason neutralization theory remains popular is because of its applicability to therapeutic practice. For instance, in their research on neutralization among inmates, Shields and Whitehall (1994) conclude that measurement of neutralizations may be considered “superfluous” in the prediction of continued delinquency. However, they decide that the neutralization scale “nevertheless would appear to have a place in a delinquency assessment battery, not because it will enhance the ability of that battery to predict delinquency, but because its content may point to an important treatment target” (p. 334, emphasis added). Similarly, after failing to find the expected correlation between neutralizations and offending, McCarthy and Stewart conclude, “Neutralization is an area of research that needs to be reopened and explored further. This will serve to provide more understanding about different forms of therapy that are appropriate for different groups of offenders” (1998, p. 289, emphasis added). Considering the relative effectiveness of cognitive interventions with offenders (Allen, MacKenzie, and Hickman 2001), this makes sense. However, it is unfortunate that criminology’s interests in cognition seems to begin and end with Sykes and Matza’s neutralization idea.

II. Neutralization Theory, Social Cognition, and Narrative

Sykes and Matza (1957) explicitly urged future researchers to develop a “systematic approach” and to probe “the internal structure of techniques of neutralization, as a system of beliefs and attitudes, and its relationship to various types of delinquent behavior” (p. 670, emphasis added). Unfortunately, the criminological community appears largely satisfied with the five rationalizations listed by Sykes and Matza and has not prioritized systematic analysis of the subjective worlds or belief systems of criminal actors (see Groves and Lynch 1990; Katz 1991; Toch 1997). Despite the ubiquity of cognitive-based treatment in correctional programming, cognitive-level research rarely appears in mainstream U.S. criminology (see Andrews, Bonta, and Hoge 1990; Foglia 2000, pp. 10–11).12

11 Neutralization theory is cited in the most thorough reviews of cognitive-behavior therapy (e.g., McGuire 2000). However, it is possible that many cognitive programs have been designed by psychologists who have never heard of, nor care about, the theory.

12 This is less true outside the United States, where an emergent “cognitive criminology” is more prominent (e.g., Polaschek, Ward, and Hudson 1997; Andrews and Bonta 1998). In a recent German study, for example, Karsted-Henke (1991) applies...
Cognitive psychology is the study of thinking, including perception, belief systems, memory, imagery, reasoning, and intelligence. The study of cognition is frequently broken down into impersonal and social dimensions. Impersonal cognition involves the processing of stimuli from the physical world. Social (or interpersonal) cognition has to do with interpretations and understandings in one’s interactions with others. With its focus on language, self-referent “inner speech,” and self-perception, the study of social cognition is sometimes associated with Farber’s (1963, p. 185) aphorism, “The things people say to themselves determine the rest of the things they do” (cited in McGuire 2000, p. 26).

Cognitive psychologists interested in adolescent aggression and conduct disorder have developed a well-articulated model of social cognition linking aggressive and antisocial behavior to a variety of cognitive factors, including endorsement of aggression-supporting social norms (Huesmann 1988; Slaby and Guerra 1988), deficits in information-processing skills (Dodge, Bates, and Petit 1990), and hostile attribution bias (Dodge 1993). Studies of the relation between accounts (explanations, neutralizations, etc.) and aggressive behavior, however, are “notably lacking in this area of research” (Guerra, Huesmann, and Zelli 1990, p. 348). Neutralization theory is still cited by a handful of psychologists working in the cognitive tradition (e.g., Slaby and Guerra 1988; McGuire 2000), yet for the most part, despite its overlap with this body of research, it has largely been ignored.

This section represents a preliminary attempt to catch neutralization theory up with the cognitive revolution and situate neutralizations within the wider and better-developed literature on social cognition and explanatory style. It is impossible to do justice to five decades of research in a short review like this, so this section inevitably reads like something of a caricature of a Social Cognition 101 lecture (for a far more satisfying review of this literature, see Fiske and Taylor [1991]). Further, a substantial amount of research on social cognition focuses on cognitive “skills” or the ability to process social information, rather than on the issues of “cognitive content” that are more closely related

Converse’s (1964) notion of the “belief system” and attribution theory to her “new approach” to understanding criminal deterrence.

Levine, Resnick, and Higgins (1993) argue that all cognition is a “fundamentally social activity” (p. 387), pointing out that “outside the laboratory and the school, cognition is almost always collaborative” (p. 599).
to neutralization theory. We focus only on those areas of research that have the most direct implications for neutralization theory, including explanatory style, locus of control, excuse theory, narrative psychology, cognitive dissonance, and shame management.

A. Explanatory Style

Central to the ascendance of cognitive approaches has been the study of attributions or the processes through which individuals construct causal explanations for their own behavior and the behavior of others (Heider 1958). The habitual use of neutralization techniques in accounting for one’s criminal actions is seen as an aspect of an individual’s “explanatory style.” Explanatory style is defined as a person’s tendency to offer similar sorts of accounts for different events in his or her life (Peterson, Buchanan, and Seligman 1995). Individuals spontaneously construct explanations or attributions for negative events—and less frequently, for positive events (Weiner 1985)—and these schematized interpretations are thought to be proximally responsible for the continuity of a person’s actions over time through “reactive person-environment interactions” (Dodge 1993; Caspi and Moffitt 1995). A person’s interpretation of the cause of some outcome seems to lead to a variety of affective reactions, and these emotions then influence subsequent behavior (Weiner 1985; Guerra, Huesmann, and Zelli 1993). For instance, if something good happens, and we perceive that we were responsible for causing it, we may feel confident. This confidence might then lead to future risk-taking or ambitious behaviors.

Considerable research suggests that processing biases occur in three salient dimensions: internality (“I am solely responsible for this good/bad event”) versus externality (“This event is someone else’s fault/responsibility”); stability (“The cause is going to last forever”) versus instability (“The cause will be short lived”); globality (“It is going to affect everything I do”) versus specificity (“It’s only going to influence this one thing”). Explanatory styles skewed toward one of these extremes (e.g., cases where people hold themselves solely responsible for everything that happens to them) are thought to correlate with specific behavioral patterns. Considerable work in cognitive psychology focuses on changing these thinking patterns as a means of changing the consequent behavior (see McGuire 2000). According to Seligman (1991, p. 8), “One of the most significant findings in psychology in the last twenty years is that individuals can choose the way they think.”
The best known application of this framework is in depression research and therapy. In their “revised helplessness theory” of depression, Abramson, Seligman, and Teasdale (1978) assign a central role to the influence of causal thinking and emphasize the unique contribution of each of these dimensions of explanatory style. The theory predicts that individuals who have an explanatory style that invokes substantially internal, stable, and global attributions for negative life events (and external, unstable, and specific attributions for positive events) will be most at risk for becoming depressed when faced with unfortunate circumstances like the loss of a job or a relationship breakup. People who use highly internal explanations for negative life events are thought to increase their risk for depression because of the threat this poses to their self-esteem. The stability-instability dimension of causal beliefs is thought to affect the chronicity of helplessness and depression following bad events. Finally, the globality-specificity of causal beliefs influences the “pervasiveness of deficits following bad events” (Peterson and Seligman 1984, pp. 348–49). This explanatory style promotes an expectation that no action will be able to control the outcome of similar events in the future, hence resulting in “helplessness” or a lack of hope.

The theory suggests that mentally healthy individuals tend to use highly internal, stable, and global attributions when accounting for good things that happen in one’s life. Yet, when negative events occur, they tend to make attributions in the opposite direction. In Learned Optimism, Seligman summarizes these findings: “For nondepressives, failure events tend to be external, temporary, and specific, but good events are personal, permanent, and pervasive. ‘If it’s bad, you did it to me, it’ll be over soon, and it’s only this situation. But if it’s good, I did it, it’s going to last forever, and it’s going to help me in many situations!’” (1991, p. 110). This widespread “beneffectance” (Greenwald 1980) or “self-enhancing biases that distort appraisals in the positive direction” (Bandura 1989, p. 1177) in the accounts of healthy, nondepressed adults, is one of the most robust findings in cognitive psychology (Seligman 1991; Dodge 1993).

People who display this sort of hedonic or self-enhancing bias are healthier (Peterson, Seligman, and Valliant 1988) and perform better in school (Wilson and Linville 1985), in the workplace (Seligman and Schulman 1986), and in politics (Zullow and Seligman 1990) than those who do not think as optimistically. In her groundbreaking research on “positive illusions,” Shelley Taylor (1989) has demonstrated that individuals who make overly optimistic self-evaluations are happier,
more able to care for others, and more productive and creative in their work than are “realists” who do not employ these cognitive “illusions.” By protecting one’s sense of self as a good person, the excuse maker is more likely to seek to live up to the standards she sets for herself (see also Taylor and Brown 1988). “Mental health, it turns out, depends not on being in touch with reality, but on illusion, self-deception and denial” (Cohen 2001, p. 56).

There is some suggestion in the criminological literature, contrary to neutralization theory, that offenders might think more like depressed persons than self-protecting and ego-enhancing nondepressives. In a study that is described later in this essay, Maruna (2001, forthcoming) found that the long-term, habitual property offenders in his sample (two-thirds of whom struggled with addiction) displayed thinking styles nearly identical to those of depressed individuals. This is not terribly surprising, as a large body of research suggests considerable overlap between the occurrence of depression and offending behavior (Chiles, Miller, and Cox 1980; McManus et al. 1984; Capaldi 1992; McLeod and Shanahan 1993). While O’Connor and colleagues (1992) attribute this to “a common genetic liability,” it is clear that criminal behavior and depression share common sociological antecedents and risk factors (e.g., stressful life events, abusive pasts, low social-structural positions; see Hoffman and Su 1998; De Coster and Heimer 2001). This is particularly true of long-term, habitual offenders, the “great pretend- ers” (Shover 1996) who cycle in and out of prisons, jails, and drug treatment facilities (see esp. the psychological portrait in Zamble and Quinsey [1997]).

Comorbidity between depression and long-term offending poses something of a puzzle for neutralization theory, which suggests that offenders should (possibly increasingly) externalize, not internalize, negative events in their lives, like getting arrested. According to the psychological literature, a depressed person faced with conviction might think to herself, “I’ve done it again. I’m just a born idiot.” In contrast, neutralization theory would predict an offender might think, “I was framed.”

This tension can be clarified by utilizing Barriga and colleagues’ (2000) useful discussion of “self-serving” versus “self-debasing” cognitive distortions. Self-serving distortions include neutralization techniques and any cognitions designed to “protect the self from blame or a negative self-concept” (p. 38). Self-debasing distortions involve more of the types of thinking linked to depression, processes like catastrophizing,
overgeneralizing, personalizing, and selective abstraction, which act to diminish self-efficacy and self-esteem (Lefebvre 1981). Whereas self-serving cognitions are thought to be implicated in externalizing behaviors (aggression, conduct disorder, delinquency), self-debasing cognitions are thought to lead to internalizing pathologies (depression, withdrawal) (see esp. Kendall 1991; Dodge 1993). “Comorbid youths are likely to vacillate between self-serving and self-debasing modes of cognitive distortion” (Barriga et al. 2000, p. 53). Ahmed and colleagues (2001) likewise found that, while some of the young people in her sample were “pure” aggressors and some were “pure” victims, there was a sizable population of victim-bullies who were both victimized and victimizers (see also Hazani 2003). This group suffered the “worst of both worlds” in terms of their thinking styles, internalizing shame for their behavior, but also externalizing blame onto others.

B. Locus of Control

The internal-external dimension of explanatory style is related to the concept of “locus of control,” or an individual’s generalized expectancy of being able to affect some future outcome relevant to him or her (Rotter 1966). Writing in this tradition, for instance, DeCharms (1968, pp. 273–74) argues that people can be divided up into “origins” and “pawns” based on their locus of control. Origins perceive their behavior to be determined by their own choices and decisions; pawns perceive their behavior as determined by external forces beyond their control.

Studies of locus of control and crime have produced decidedly mixed and conflicting results (Hollin 1989), sometimes indicating that offenders have unusually external loci of control (e.g., Kelley 1996) and sometimes finding the opposite (e.g., Lefcourt and Ladwig 1965). According to Hollin (1989), these inconsistent findings can be explained by two unfounded assumptions in this research: that locus of control is a unitary concept and that offenders form a homogeneous population.

Research following Rotter’s original formulation has uncovered numerous dimensions of locus of control (e.g., stability, globality) that are not measured in Rotter’s (1966) locus-of-control scale. As explanatory style research has progressed in other areas, the internality-externality dimension central to the locus of control concept has become of less interest to researchers. “It has more inconsistent correlates than do stability or globality, it is less reliably assessed and
there are theoretical grounds for doubting that it has a direct impact on expectations per se” (Peterson 2000, p. 48).

Explanatory style is thought to have differential effects depending on the nature of the event (positive/negative, past/future) being explained (Lefcourt et al. 1979). For instance, persons who use high-internal explanations for positive life events often use low-internal explanations for negative life events. This is the “hedonic bias” identified in countless research studies with nondepressed adults. The locus-of-control score would combine both types of attributions into one dimension, making the final score difficult to interpret (i.e., is an internal locus of control adaptive or counterproductive?) (Peterson, Buchanan, and Seligman 1995).

The locus-of-control score makes no distinction between past and future events. Wortman (1976) and others have found that externalizing negative behaviors in the past can enhance an individual’s feelings of control over his or her environment in the future. Brickman et al. (1982) provide an interesting framework for understanding and modeling this shift in locus of control. Unlike the locus-of-control literature, they divide the concept of personal responsibility into two dimensions: blame and control. In other words, they distinguish between taking responsibility for the origin of a problem and taking responsibility for its solution (see also Weiner et al. 1987).

Instead of dividing personality types into “pawns” and “origins,” this framework allows for four different orientations toward behavior: a moral model, an enlightenment model, a medical model, or a compensatory model. In a moral model, persons hold themselves responsible for their problems and for solutions to those problems. In an enlightenment model, individuals hold themselves responsible for their problems, but not for solutions. In a medical model, people do not hold themselves responsible for their problems or for the solutions. Finally, in the compensatory model, individuals do not blame themselves for their problems, but hold themselves responsible for the solutions. A compensatory model ex-offender might say, “I only got into crime and drugs because of my disadvantaged childhood, but now I am working hard to go straight.”

C. Excuse Theory

Excuses have been the subject of considerable research in social psychology. Unlike neutralization theory, most of this research emphasizes the benefits of excuse making. The philosopher John Langton
Austin once made a “Plea for Excuses” (Austin 1979). He argued that the function and purpose of excuses were both understudied and underappreciated. Schlenker, Pontari, and Christopher suggest that “since the time of Austin’s plea, the pendulum may have swung too far in the other direction.” They write: “There is now such a research emphasis on the advantages of excuses that their disadvantages have been relatively neglected. Indeed, it is now commonplace to encounter, in books and journal articles, the recommendation that people who fail or otherwise encounter difficulties in life should be taught to shift causal responsibility away from core components of the self, thereby making excuses to shield them from the emotional and interpersonal costs” (2001, p. 20).

Posing the question “do excuses work?” Snyder and Higgins (1988) conclude that excuse making is a highly adaptive mechanism for coping with stress, relieving anxiety, and maintaining self-esteem. Individuals who make excuses for their negative actions tend to have better psychological adjustment and even better health than those who assume full responsibility for their shortcomings (see also Snyder, Higgins, and Stucky 1983). Excusing past mistakes can even enhance one’s sense of control over future challenges of the same nature (Wortman 1976). One can think “I will win this time, since the only reason I lost last time was because the wind was against me.”

Excuse making seems to have numerous social benefits for the excuse maker. Excuse making is understood by social psychologists as “a type of aligning action indicating to the audience that the actor is aligned with the social order even though he or she has violated it” (Felson and Ribner 1981, p. 138). Extensive research in psychology demonstrates that the provision of excuses (or mitigating accounts) for one’s harmful actions can reduce conflict (McLaughlin, Cody, and O’Hair 1983), preserve the speaker’s reputation (Crant and Bateman 1993), and reduce negative sanctioning (Blumstein et al. 1974). Indeed, mitigating accounts have been shown to protect individuals against punitiveness even in the case of highly violent crimes (Kleinke, Wallis, and Stalder 1992; Rumgay 1998, p. 204). Often, listeners collaborate in the construction of excuses, helping wrongdoers arrive at acceptable accounts of their actions (Snyder and Higgins 1988).

One setting in which excuses do not seem to be welcomed by listeners is the prison environment, where the notion of excuse making as pathology has taken hold (see Fox 1999a). Attributional research in the prison environment suggests that prison staff routinely make
internal attributions when explaining disruptive behavior among inmates, blaming prisoner personality rather than the structural frustrations of the prison environment (Saulnier and Perlman 1981). Contrary to findings of similar research in the outside world (e.g., Felson and Ribner 1981), prisoners who offer external explanations (e.g., excuses) for their behavior seem to be punished more severely than those who offer no such mitigation (Steinke 1992). Steinke concludes: “In prison, there may be no excuse or justification for violence, but outside the prison there may be many excuses and justifications for violence. . . . Previously learned strategies for mitigating blame may not have applied” (1992, p. 484).

Finally, considerable research has documented that use of excuses in social situations may protect the “victim” of the offense in question, “who might otherwise be insulted or harmed by the conduct” (Schlenker, Pontari, and Christopher 2001, p. 17). Harm doers who communicate to their victims that their actions were unintentional or the product of extenuating circumstances tend to provoke less anger than harm doers who do not offer excuses (Greenberg 1993; Sitkin and Bies 1993). Excuse making seems to convey a level of respect for the victim. “The very fact that the perpetrator thinks that the victim is due an explanation signals respect for the victim and tends to diminish the victim’s anger” (Miller 2001, p. 537; see also Bies 1987).

The research on excuse making is not entirely positive, of course. Schlenker and colleagues (2001) warn that, among other problems, excuse making can rob the individual of his or her sense of self-control and erode one’s sense of self-efficacy. Although excuses are generally thought to preserve self-efficacious beliefs by separating the actor from the negative act, overuse of excuses can imply that the actor lacks the ability to control his or her own actions. Schlenker, Pontari, and Christopher argue that excuses may be socially and personally advantageous under three primary conditions: when they are credible; when they maintain self-engagement in cases of important, recurring tasks; and when they maintain goodwill for the excuse maker and do not give the impression of narcissism (2001, p. 25).

D. Narrative Psychology

Following Sykes and Matza, criminologists have tended to characterize neutralizations as discrete, individual “techniques,” counting how often these apparently free-floating and isolated thoughts are used, or else measuring the subjects’ levels of support for each. As a
result, criminology has not approached the greater understanding of “the internal structure of techniques of neutralization as a system” advocated by Sykes and Matza (1957, p. 670). This would require an analysis that moves from the techniques themselves to an “integrating theory” of sense making that can guide model building and empirical research (Ward, Keenan, and Hudson 2000).

Tony Ward (2000) has made a parallel argument concerning sex offender treatment. Although considerable attention has been paid to documenting the content of specific cognitive distortions in sexual offenders, he argues that there has been “little attempt to develop a theoretical account of the mechanisms generating these distorted attitudes” (p. 493). Much research on sex offender rationalizations, he argues, provides little more than a long list of common “cognitive distortions” used by offenders of various stripes with the “underlying assumption . . . that these beliefs constitute separate, and unrelated, vulnerability factors” (p. 492). Ward (2000) suggests that instead of focusing on the symptoms of distorted thinking, correctional counselors should look to the source or the cognitive schemata underlying these patterns of belief. Likewise, Beech and Mann (2002, p. 268) describe a schema-based treatment program that focuses not just on offense-justifying attitudes but also on underlying self-understandings, motivations, and implicit beliefs.

The suggestion here is that neutralization acceptance—like excuses, locus of control, or explanatory style—is one outward manifestation of a person’s self-identity or self-concept. This makes sense as theorists across numerous disciplines have started to agree that one’s identity takes the form of a personal narrative used to guide and organize human behavior (e.g., McAdams 1985; Sarbin 1986; Giddens 1991).

The narrative identity can be understood as an active information-processing structure, a cognitive schema, or a construct system that is both shaped by and later mediates social interaction. People construct stories to account for what they do and why they do it. These narratives impose order on our actions and explain our behavior with a sequence of events that connect up to explanatory goals, motivations, and feelings. These self-narratives act to shape and guide future behavior, as people act in ways that agree with the stories or myths they have created about themselves (McAdams 1985). According to Bruner, “eventually, the culturally shaped cognitive and linguistic processes that guide the self-telling of life narratives achieve the power to structure perceptual experience, to organize memory, to segment
and purpose-build the very ‘events’ of a life. In the end, we become the autobiographical narratives by which we ‘tell about’ our lives” (1987, p. 15, emphasis in original).

Identity narratives are closely related to the “cognitive scripts” (Abelson 1976) or “cognitive schemata” (Crocker, Fiske, and Taylor 1984) that help us structure, organize, and interpret new information (while also allowing us to ignore substantial quantities of information in our environments). Others have linked self-narratives to an “implicit theory” or a “theory of reality” akin to the theories used to explain, predict, and interpret in the world of science (Epstein and Erskine 1983; Ward 2000). These implicit theories tend to develop in childhood, but can be modified over time. “Just as in science, theory evaluation is a comparative process,” Ward (2000, p. 497) argues, and theories can be replaced by alternatives that better account for the accumulation of data. Unfortunately, like biased scientists, lay people “often fail to critically evaluate evidence” and often their “theories dictate the way evidence is appraised” (Ward 2000, p. 495).

A variety of methods have been proposed for accessing these internalized, identity constructs in qualitative research, but most involve intensive, semistructured interviews (see, e.g., Hollway and Jefferson 2000). The transcribed life story documents produced in such research are not themselves thought to be the self-narratives that guide an individual’s behavior. The stories people tell social scientists about themselves, however, are assumed to “hold the outlines” of their internalized self-narratives in the same way that answers to an attitude survey or a pencil-and-paper personality test represent a person’s attitudes or personality traits (McAdams 1993). A well-established body of research suggests that there are strong links between our public and private “selves,” so that convincing others of one’s self-narrative may be an important step toward convincing oneself (see, e.g., Baumeister 1982; Tice 1992). When done skillfully, in an environment that is at least passably conducive to open discussion, a life story interview can provide rough indicators of the internal self-story that the person lives by. These narrative transcripts can be quantitatively coded and systematically compared for cross-case similarities and differences in theme, tone, style, motivation, and characterization.

Neutralization acceptance is just one dimension in which these narratives can be understood. Importantly, our crimes and misdemeanors take on a central role in all of our life stories, as all of us have done morally questionable or regrettable things in our lives. Baumeister and
Wilson write: “These acts then pose the most direct challenge to the construction of a life story. . . . Whether these are heinous war crimes or petty interpersonal cruelties, the person may devote significant effort to explaining these episodes in a way that reduces or eliminates personal guilt” (1996, p. 323). Neutralization techniques become a central way of maintaining a particular narrative of the self and creating a sense of cohesion out of lives that are “experienced as discontinuous, radically changing and full of shame and guilt and that is felt or feared to be worthless” (Lofland 1969, p. 282).

E. Cognitive Dissonance, Shame, and Self-Esteem

A central, but infrequently discussed, issue is what, exactly, neutralization techniques are intended to do. That is, neutralization theory proposes a relationship between neutralizing and offending (this might be less a causal relationship than a “soft determinism”). Yet, dynamic processes mediate the relationship between these two variables. Neutralizations are variously meant to protect a person from pangs of conscience, cognitive dissonance, shame, guilt, remorse, self-awareness, loss of self-esteem, public labeling, and stigma. All these processes have been the subject of considerable research in criminology and beyond, yet little of this has been incorporated into our understanding of neutralization theory.

An important starting point in refining neutralization theory in this way would be to integrate the theory with its sister theory in psychology (twin sister, perhaps, as it was born in the same year): Festinger’s (1957) notion of “cognitive dissonance.” Dissonance theory predicts that if a person holds cognitions that are psychologically inconsistent (e.g., stealing is wrong, I have stolen something, I am a good person), he or she would experience dissonance and would in turn seek ways of reducing this dissonance. This can be done by changing the behavior (so long as it is not in the past tense) or adjusting one’s beliefs. This hypothesis has been the focus of an enormous amount of research. “In its heyday, the theory generated over a thousand separate experiments, many of which were startling at the time, teaching us hundreds of things about human behavior” (Aronson 1992, p. 304).

In his reformulation, Aronson (1968) argues that cognitive dissonance is especially potent when a person’s self-concept is at risk. Aronson (1968, 1992) argues that individuals strive to preserve a sense of self that is consistent, stable and predictable, competent, and morally good. One chief strategy of preserving these aspects of the
self, he argues, is to distort one of the inconsistent self-images using culturally approved denials (e.g., “it isn’t what it looks like”) to make it fit with preexisting self-concepts.

In an extension of dissonance theory, Wicklund and Brehm (1976) argue that personal responsibility for undesirable consequences is the ultimate cause of dissonance. They use the example of a habitual smoker who learns that smoking is bad for health. Because this knowledge is dissonant with the cognition that he continues to smoke, he will seek to reduce the dissonance either by giving up cigarettes or by changing his cognition about the effect of smoking on health. In other words, cognitive dissonance theory, like neutralization theory, predicts that the individual will seek to neutralize the cognition through a variety of excuses and justifications.14

Stice (1992) has linked the concept of dissonance to the even-better-studied concept of shame, arguing that dissonance and shame are similar in their eliciting conditions, subjective experiences, and consequences. Stice suggests that analogous techniques (from confession to alcohol consumption) seem equally capable of reducing both guilt and the pains of cognitive dissonance. Similarly, Michael Lewis (1992) has linked attributional theory to shame, arguing that the attribution of positive and negative events in our lives is directly related to whether we feel pride, hubris, regret, or shame.

These reformulations accord with the definition of shame offered by Nathan Harris (2001) and his colleagues in the Australian reintegrative shaming experiments (Ahmed et al. 2001) in their exhaustive literature review and study of the experience of shame. Harris uses the “ethical identity” concept of shame developed by moral philosopher Bernard Williams (1993). In this framework, shame can be understood as a threat to the individual’s identity or a betrayal of one’s image of self.

Feelings of shame are associated with numerous painful experiences, including social isolation, depression, hopelessness, and social withdrawal (Lewis 1992; Tangney 1995). Although underdiscussed in neutralization research, shame has also become an important area of research in criminology in general (e.g., Grasmick and Bursik 1990).15

---

14 For other discussions of the links between cognitive dissonance and neutralization theory, see Dunford and Kunz (1973) and Hazani (19916).

15 Remorse is another, highly underresearched possibility for an intermediate variable between neutralizations and offending. Horne (1999, p. 31) argues that “though the experience of remorse may deter in certain special cases, it is most unlikely that (remorse) has much effect in the general run of cases.” Little empirical work has tested whether recidivism is higher among the remorseless than the remorseful, all other factors
In particular, John Braithwaite and colleagues (Braithwaite 1989; Ahmed et al. 2001) have argued that successful reintegration of ex-offenders into society requires a process of “shame management.” They argue that shame is a crucial but dangerous construct. When individuals can acknowledge their shame for personal wrongdoing, apologize, and be forgiven, shame can be reintegrative. When shame is experienced as stigmatizing, or when shame is unresolved or “displaced” (i.e., projected onto others), shame is likely to have criminogenic effects, according to this framework.

In some understandings of denial and other defense mechanisms (starting with Fenichel 1945), self-esteem is the key intermediate variable. Some researchers have argued that youths who neutralize delinquent behavior should have higher levels of self-esteem than youths who do not (e.g., Rathus and Siegel 1973; Costello 2000).

However, the relationship between self-esteem and crime is anything but clear. Social scientists examining the relationship between self-esteem and deviance have found decidedly mixed and often contradictory results (much of this research has been cross-sectional and has not untangled the chronological sequences; see, e.g., Wells and Rankin 1983; McCarthy and Hoge 1984; Jang and Thornberry 1998).

Baumeister (2000) and others have argued that aggressive and violent individuals tend to have higher than average self-esteem (which tends to be unstable and under some sort of threat). Yet, considerable research suggests that individuals with high self-esteem are more resilient in the face of setbacks, work harder, and bounce back from failures better than those with low self-esteem (Shrauger and Sorman 1977). In developing his “self-enhancement theory,” Kaplan (1975, 1976, 1978) used a three-wave panel study of adolescents to discern the dynamic relationship between self-esteem and delinquency. He concludes that low self-esteem leads to delinquent behavior, which in turn raises self-esteem. The reciprocal nature of self-esteem and delinquency is also supported by Rosenberg, Schooler, and Schoenbach (1989).

Rather than self-esteem, neutralizations are probably best understood as “insulation from labeling” (Covington 1984, p. 621), in which “stigma” (Braithwaite’s nonreintegrative form of shame) is the key
intermediate variable between neutralization and offending. The rationalizations employed by offenders, therefore, may emerge out of a “human need to protect the self” (V. Braithwaite and J. Braithwaite 2001, p. 327). The need to protect oneself against stigma does seem fairly universal (see Rogers and Buffalo’s [1974a] nine modes of adaptation to a deviant label). If this is the case, and stigma is associated with offending, it is difficult to see why neutralizations would increase offending rather than decrease it. Preserving one’s sense of self as essentially noncriminal (through neutralizations or any other technique) may be necessary for offenders to desist from crime (see Minor 1981, p. 331; Costello 2000, p. 324). Meisenhelder argues, “The plan to exit from crime is in large part founded on the sense of the self as noncriminal” (1982, p. 140).

As predicted by cognitive dissonance theory, the primary motivation behind neutralizations is the establishment of a sense of internal consistency. Our identities as individuals are largely premised on our abilities to “keep a particular narrative going” (Giddens 1991, p. 54). Disruptions to this sense of predictability in our lives require some revisionary effort, of which neutralization techniques are just one version.

III. Research on Neutralization Theory

Research on neutralization theory has generally taken two forms. In the first, investigators use inductive and qualitative methods to illustrate how neutralizations are used by deviant actors. This research provides rich insight into the ways that deviants understand and make sense out of their lives, and offers a rare peek “inside the mind” of the deviant actor. It has informed us about the diverse motives and excuses that deviants invoke to justify their actions and has uncovered many new neutralization techniques. The second form of neutralization research consists of empirical assessments. These studies typically use survey designs to test core assumptions by locating a sample of known offenders and a control group sample of “innocents,” then asking respondents in both groups to agree or disagree with a list of neutralizations (often in relation to hypothesized scenarios). When a relationship is found, the correlation between acceptance of neutralizations and criminality seems to be small to moderate at best (see Ball 1966; Minor 1981; Thurman 1984; Agnew and Peters 1986; Mitchell, Dodder, and Norris 1990). These mixed results are generally blamed on methodological problems inherent in the theory (Akers 2000).
A. Illustrative Research

Qualitative methods, such as participant observation (e.g., Friedman 1974; Hong and Duff 1977), content analysis (e.g., De Young 1988; Durkin and Bryant 1999; Dabney and Vaughn 2000), and semistructured interviewing (e.g., Forsyth and Evans 1998; Erez and Laster 1999) are well suited for investigating neutralizations. These methods comprise a large portion of research on neutralizations. Such studies appear almost every year in journals, confirming that people situationally invoke neutralizations when accounting for social transgressions as diverse as deer poaching (Eliason and Dodder 1999), becoming a hit man (Levi 1981), committing hate crimes against the Amish (Byers, Crider, and Biggers 1999; Byers and Crider 2002), assaulting prostitutes (Miller and Schwartz 1995), stealing office supplies (Hollinger 1991), contributing to genocide (Stewart and Byrne 2000), and snitching on peers (Pershing 2003). The cross-study consistency in the types of accounts used to explain these disparate acts suggests that neutralization theory provides a highly robust framework for explaining how deviants allay their feelings of guilt.

Scully and Marolla’s (1984) research on convicted rapists’ accounts exemplifies this type of research. They conducted in-depth interviews (lasting between three and seven hours) with 114 incarcerated rapists. Interviews revealed that a large number denied that what they did was wrong. These deniers used two forms of justifications, both of which “ultimately denied the existence of the victim” (p. 542). The first stems from the “cultural view of men as sexually masterful and women as coy but seductive” (p. 542). Rapists who denied they sexually assaulted anyone argue that women often say no, but in truth they really mean yes. One rapist expressed it this way: “When you take a woman out, woo her, then she says ‘no, I’m a nice girl,’ you have to use force. All men do this. She said ‘no’ but it was a societal no, she wanted to be coaxed. All women say ‘no’ when they mean ‘yes’ but it’s a societal no, so they won’t have to feel responsible later” (p. 535).

In the second form of victim denial, “the victim was portrayed as the type of woman who ‘got what she deserved’” (Scully and Marolla 1984, p. 542). Those who deny that they are rapists often claim that “nice girls don’t get raped” and point to the victim as being “loose” or overly seductive. If the victim had not dressed or acted the way she did, then the sexual assault would not have occurred. Many rapists hold this belief even though they committed the rape with a weapon or during a burglary. Perhaps the most important finding is that rapists rely on
cultural stereotypes to justify their actions. Scully and Marolla conclude that “convicted rapists have learned the attitudes and actions consistent with sexual aggression against women” by interacting with others (p. 530).

Another exemplary study that used in-depth interviewing is Benson’s (1985) analysis of the accounts of white-collar offenders. Benson interviewed thirty individuals convicted of white-collar crimes, including antitrust violators, tax violators, and fraudsters. Each offender type presented neutralizations unique to the crime they committed. For instance, antitrust violators maintained that they were simply following standard business procedures that were necessary for the company to survive. They often characterized their wrongdoings as benign, especially when compared to those of street criminals. Some stated that the only reason they were prosecuted was because of personal motives of the prosecutors. By contrast, fraudsters denied committing any crime at all. They claim that they were either set up by associates or were duped by others. Thus, the accounts appear to be structured by “the nature of the offense, its organizational format and history, and by the requirement that they undermine the conditions of successful degradation ceremonies” (Benson 1985, p. 602).

B. Problems with Interview-Based Research

Research of this type has provided highly valuable insights into the motives and mindsets of deviants, and how these are learned through interaction with subcultural others. The findings from interview-based studies (which admittedly do not endeavor to “test” or evaluate Sykes and Matza’s theory) should be interpreted cautiously and not be seen as strong evidence in favor of neutralization theory. This research has done little to develop neutralization theory beyond identification of new techniques for neutralizing guilt.

In addition to the lingering, unresolved problems with sample selection and generalizability that always plague qualitative research, these designs are problematic because of the situational demands of the interview situation. The ubiquity of neutralizations in the qualitative literature might be a product of the circumstances under which these neutralizations are elicited. The nature of the interviewer/interviewee relationship often makes it clear who is the deviant. Respondents may believe that they are forced into a position where they must defend themselves and others like them. Thus the neutralizations they offer may be artificially created for the interviewer (Hindelang 1970). This is...
especially true of interview subjects who are located in prison or some other correctional setting after being convicted of the crime they are asked to account for. These accounts might not be representative of other accounts given in other circumstances.

Presser (2004) has drawn attention to the interactional nature of the narrative process and the influence of situation and setting on the content of offender stories. There are obvious and important differences in the social demands of settings such as “a court room, an everyday conversation, a confrontation with a jealous lover [or] a psychotherapy session” (Cohen 2001, p. 50). Goffman (1956) discussed the problem of “audience segregation,” where individuals give different accounts of themselves in different social circumstances—portraying one identity to one’s family, another to one’s football team, and another in the office place (see also Emler and Reicher 1995). Preliminary research suggests that justifications of violence and deviance are most commonly used among peers, while exculpatory excuses are used when presenting one’s story to strangers or outsiders (Harvey, Weber, and Orbuch 1990; Toch 1993). One’s peers better understand justifications like “the bastard had it coming” or “I did it for us” than the outsider would. This “radically social” character of account makings, understood at least since Mills (1940), calls into question the reliability of interview methods for testing neutralization theory.

More important, these qualitative studies of neutralizations cannot be considered a test of the central neutralization premise because they almost never include a comparison group. It is impossible to know whether acceptance of these neutralizations is in some way unique to deviants or if the beliefs are widely accepted throughout society. Frequently, the implication in such studies is that the rationalizations used by offenders are widely shared. For instance, the accounts of convicted rapists in Scully and Marolla (1984) are thought probably to reflect a patriarchal culture’s rape myths. If rapists believe the same myths as nonrapists, something more than the myths must account for the offending behavior.16

16 This seems to be the case in empirical tests of the “cognitive distortions” of sex offenders. Reviewing a vast array of research on the cognitive and psychological profiles of sex offenders, Tony Ward and his colleagues (Ward, Keenan, and Hudson 2000, p. 47) conclude that sex offenders tend to have “traditional and conservative views” that are “indiscriminable from men in the general population.” Similarly, Beech and Mann (2002, p. 265) write: “There have been no consistent findings to date of distorted attitudes related to rape among rapists as compared to non-offenders.”
C. Measuring Neutralization

As a consequence of these difficulties with qualitative designs, there have been numerous attempts formally to operationalize the neutralization concept in scales of survey items. The most widely used neutralization scale was developed by Ball (1966) for use in his dissertation research. Ball administered his questionnaire to fifteen-to eighteen-year-old youths at a state school for delinquents and at an inner-city high school in Ohio (see Ball 1966). The neutralization scale consisted of four behavioral situations: gang fight, shoplifting, robbery with a knife, and gang fight with weapons. A scenario for each behavior was provided. For instance, the scenario for gang fighting: "Jack gets a club and goes with his friends to look for another group of boys. They find them in the park, and a fight starts. During the fight Jack hits another boy with the club, and almost kills him."

For this and similar scenarios, respondents were presented with ten different sets of "realistic excuses" (Reckless 1973, p. 26) and were asked to what degree they agreed with the excuse. These excuses were designed to measure each of Sykes and Matza's five original neutralizing techniques. Excuses that corresponded to the gang fight scenario include "People should not blame Jack this time if he was trying to protect himself" and "People should not blame Jack if he had been drinking." Responses to forty excuses were summed to determine the level of neutralization acceptance (high scores equal high neutralization acceptance).

Ball's scale has since been used several times in its original form and has been modified to fit particular types of deviance and deviants. Versions have been used to measure the neutralizing beliefs of Singaporean adolescents (Khoo and Oakes 2000), college students (Haines et al. 1986), juvenile offenders (Mannle and Lewis 1979), and adult offenders (McCarthy and Stewart 1998).

The measure is not without limitations. McCarthy and Stewart (1998, p. 281) describe Ball's and similar scales as providing a "crude measure of offenders' basic moral evaluations of offenses." The wording of vignettes and questions in the scale certainly leave much room for confusion and misinterpretation. Shields and Whitehall (1994, p. 227) further argue that Ball's scale is "too long (in that many young offenders have short attention spans), too verbally sophisticated (given their limited vocabularies), and beyond the reading skills of many."

17 This may be too critical. The scale, reprinted in Reckless (1973), appears appropriate for an adolescent reading level by most standards (e.g., Cutts 1995).
Others were prompted to develop their own neutralization scales (see, e.g., Shields and Whitehall 1994). Norris and Dodder (1979) developed a scale based on the idea that neutralizations were one form of moral relativism (see also Dodder and Hughes 1993). They argued that individuals fell along a continuum ranging from goodness to rebellion.

The continuum consisted of four ideal types of behavior: moral absolutes, situational ethics, neutralizations, and rebellious absolutes. They created statements to reflect the amount of moral relativism individuals applied to sixteen behaviors, ranging from minor forms of delinquency (disobeying parents, speeding, vandalism) to serious criminal offenses (auto theft, check forgery, murder). For each behavior, respondents were asked if they agreed with four statements measuring moral relativism. For example, concerning speeding, respondents were asked if “I believe it is wrong to exceed the speed limit,” “I believe it is OK to drive past the speed limit if I am taking someone to the emergency room,” “I believe it is OK to drive faster than the speed limit if it is too slow for the time of day I am driving,” and “I believe it is right for me to drive as fast as I want whenever I want.” To be classified as a moral absolutist the respondent would agree only with the first statement. To be counted in the situational ethic category he had to agree with only the first two statements. Those in the neutralization category agreed to the first three, and those in the rebellious category agreed only to the last.

Rogers and Buffalo (1974b) devised a scale to measure the neutralizing beliefs individuals held for the specific offense for which they were convicted. The boys were presented with nine neutralizing statements, and they responded using a five-response Guttman scale. The statements included, “What I did was not so bad, no one was really hurt,” “I got into trouble because I got in with the wrong boys,” and “The judge and the court were against me from the start.” Responses were then scaled to create a scale measuring neutralization acceptance.

Ideally, neutralization measures are created specifically to measure neutralizations (e.g., Shields and Whitehall 1994); however, many studies rely on questions designed for other purposes (e.g., Hollinger 1991; Agnew 1994).18 In these studies, researchers fit existing measures into neutralizing concepts. For instance, in the National Youth Survey,

18 Numerous other scales have been developed by criminological psychologists to measure the related concept of attribution of blame in offender populations, including the Blame Attribution Inventory (Gudjonsson 1984), the Attribution of Blame Scale (Loza and Clements 1991), the Measure of Automatic Thinking Errors (Garvin 1990),
respondents are asked to what extent they agree with such statements as “It’s alright to beat up people if they started it” and “It’s alright to physically beat people up who call you names.” Agnew (1994) used these and similar questions as measures of the denial of victim.

D. Theory Testing: Does Neutralizing Predict Deviance?

Neutralization theory states that since delinquents and nondelinquents are similarly committed to conventional values, it is only because delinquents are successfully able to neutralize that they are able to engage in delinquency. Therefore, we would expect a correlation between findings on neutralization scales and delinquency. Research directed at measuring the effect of neutralizations on deviance has typically used one of two research strategies, both of which tend to find weak but positive relationships between acceptance of neutralizations and participation in delinquency (Ball 1966; Minor 1980; Mitchell and Dodder 1983; Thurman 1984; Haines et al. 1986; Hollinger 1991; Agnew 1994).

One strategy is to compare a sample of known delinquents with a sample of nondelinquents to determine if delinquents are more accepting of neutralizations than nondelinquents. Ball’s (1966, 1977) research exemplifies this methodology. In his first study, Ball found that delinquent boys, regardless of whether delinquency was measured with self-report or official reports, scored significantly higher than nondelinquent boys. Mitchell and Dodder (1983) found similar results when comparing high school boys with institutionalized adolescent offenders, which was interpreted as support for the theory.

The second way this question has been addressed is by using measures of neutralization acceptance to predict self-reported delinquency in a single sample. This design has been used to examine the correlation of neutralization scores and relatively minor deviant acts such as college cheating (Haines et al. 1986; LaBeff et al. 1990), workplace deviance (Hollinger 1991), drinking behaviors (Dodder and Hughes 1993), shoplifting (Agnew and Peters 1986), and minor delinquency (Hirschi 1969; Mannle and Lewis 1979). Overall, this research has found positive but weak effects of neutralizations on deviance. Unfortunately, the bulk of this research has used cross-sectional designs, which cannot

---

44 Shadd Maruna and Heith Copes

Mechanisms of Moral Disengagement (Bandura et al. 1996), and the HIT (How I Think) Questionnaire (Barriaga and Gibbs 1996). More general scales measure explanatory style, attributions, and locus of control among the general population.
disentangle the sequential relationship of neutralizations and deviance (e.g., Hamlin 1988; Agnew 1994). Without longitudinal designs, there is no way to determine whether neutralizations precede criminal behavior or are merely after-the-fact rationalizations.

Longitudinal designs are much better equipped to answer this fundamental question. Unfortunately, only a handful of longitudinal studies have been conducted. In the first such study, Minor used a two-wave panel survey of undergraduates at the University of Maryland. At the beginning of the semester, students completed a survey asking them about moral evaluations, acceptance of excuses, and self-reported participation in minor forms of deviance (e.g., marijuana use, cocaine use, fighting, being drunk, cheating on exams, and shoplifting). Students were then given a similar survey at the end of the semester. Minor concluded that “For several forms of minor deviance, excuse acceptance is found to be related to subsequent behavior in the manner predicted by the theory” (1981, p. 295). Minor employed a longitudinal design, but the work has been criticized not only for the convenience sample used but also for the short period between the two surveys (three months).

Using a more relevant sample of known offenders, Shields and Whitehall (1994) administered their neutralization survey to youths incarcerated at a juvenile facility. They checked the official status of a sample of these youths one year later to determine how many were reconvicted and resentenced to closed-custody incarceration. Initial neutralization scores were higher for youths who recidivated than for those who did not. The results were weak, however; other scales on their survey were better predictors of delinquency that was the neutralization scale.

Perhaps the soundest longitudinal study of neutralizations was conducted by Agnew (1994). Using the National Youth Survey, Agnew explored the relation between neutralization and future violent behaviors using a national sample of youths. The survey consisted of three waves (spaced a year apart) beginning in 1977 (Elliot, Huizinga, and Ageton 1985). Only waves 2 and 3 were analyzed because they “contain a fuller set of violence neutralizations than the first wave” (p. 564). Agnew used a neutralization scale created from responses in wave 2 to predict self-reported violence in wave 3.19 Respondents were asked to

19 One potential problem with this design is that delinquency and violence committed during wave 1 could affect responses to the neutralization scale at waves 2 and 3. Agnew did not include controls for wave 1 delinquency or wave 1 violence.
state the degree to which they concurred with three neutralizing statements: “It’s alright to beat up people if they started it”; “If people do something to make you really mad, they deserve to be beaten up”; and “It is sometimes necessary to get into a fight to uphold your honor or ‘put someone in their place.’” Response categories were strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree. Responses were averaged to create the neutralization scale. Violent behavior at time 3 was then regressed on neutralization at time 2, while controlling for approval of violence, delinquent peers, and previous violence. Agnew (p. 572) concluded, “Taken as a whole, the longitudinal data suggest that neutralization may be a relatively important cause of subsequent violence.”

E. Problems with Survey-Based Neutralization Research

Despite this mostly underwhelming empirical support, criminologists have not given up hope on neutralization theory. Researchers have offered a variety of explanations for the mixed findings of most quantitative evaluations of the theory (e.g., Agnew 1994) and argue that “all the studies conducted to date fell short of any real standard of proof either supportive or nonsupportive of neutralization theory” (Hamlin 1988, p. 428). Although survey research has provided a great deal of information regarding the use of neutralizations, it suffers from several, seemingly insurmountable problems.

First, Sykes and Matza’s fairly uncomplicated theory is often misrepresented. For instance, except for the few longitudinal designs (e.g., Minor 1981; Agnew 1994), survey research has been unable to determine accurately if neutralizations precede criminal behavior. Many tests of neutralization theory are actually testing whether people who have been convicted of a crime tend to score higher on neutralization-like measures than young people who have not. This is not a test of the rationalization process.

Likewise, neutralization research often fails to distinguish between beliefs that neutralize conventional bonds and beliefs that simply show unconventional commitment (Austin 1977, p. 124). In typical neutralization measurements, including Ball’s survey and those similar to it, respondents are asked if they agree with statements such as “People should not blame Jack [for shoplifting] if this was the normal thing to do where Jack lived” (Ball 1973, p. 28). Acceptance of this statement is subject to multiple interpretations. It could mean that respondents thought this was a good excuse, or they could have thought shoplifting
was morally acceptable regardless of whether an excuse was used. Even Austin’s measure of neutralization (e.g., “Suckers deserve to be taken advantage of”) can be interpreted as acceptance of unconventional values rather than neutralizations.

Agniew argued that the “mixed results as well as the small-to-moderate effect of neutralization found in many studies, may be due to a fundamental flaw in virtually all research on neutralization” (1994, p. 560). Agnew suggests that most tests presume a causal relationship between neutralizations and offending that misrepresents Matza’s (1964) clear explanation that neutralizations only “allow for” delinquency. Neutralizations are likely to lead to delinquency only among those who are in situations in which the neutralizations are applicable, who encounter opportunities for delinquency, and who have a “strong need or desire to commit the offense” (Minor 1981, p. 301). For example, college students who think it is acceptable to cheat on exams if other people around them are cheating must believe that people around them are actually cheating, which is what Haines et al. (1986) found in a survey of college students. Agnew and Peters (1986) argue that this reformulation of neutralization theory can explain the contradictory findings of other researchers. For example, Ball (1977) found that sixth graders accepted just as many neutralizations as institutionalized delinquents but did not engage in as much delinquency. It is possible that the sixth graders encountered fewer situations where the neutralizations were acceptable. This same explanation can be used to explain the findings that females accept the same number of neutralizations as males, if not more, but commit far less crime and delinquency.

Second, researchers frequently rely on inappropriate samples. Several of the most frequently cited tests of neutralization theory use all-too-convenient samples of university students enrolled in criminology or sociology courses (e.g., Mitchell and Dodder 1983; Minor 1984; Agnew and Peters 1986; Dodder and Hughes 1993). Generalizing from such samples to the population of typical interest to criminologists (e.g., street offenders) is problematic. Discussing one such study, Hamlin (1988, p. 428) observed, “The sample was so inadequate that even [the] limited support must be seriously questioned.”

Neutralization research also relies heavily on incarcerated samples, with a person’s incarcerated status being used as evidence of deviance (e.g., Slaby and Guerra 1988; Barriga et al. 2000). The many problems with using incarcerated samples (see Polsky 1969) are magnified in
cognitive research, where familiar findings of low self-efficacy, weak locus of control, and overall levels of frustration and hostility are magnified or distorted by deprivations associated with incarceration (see Blatier 2000). Such prison-based cognitions may have no relevance to the same person’s thinking patterns outside a total institution (Foglia 2000). Countless situational demands inherent in the prison setting can magnify the possibility of response bias. In an early attempt at operationalizing the “thinking errors” found in clinical work, for instance, Garvin (1990) developed a sixty-four-item pencil-and-paper questionnaire, the Measure of Automatic Thinking Errors (cited in Barriga and Gibbs 1996). She tested the instrument on a group of incarcerated juveniles and a control sample of nonincarcerated young people. She was surprised to find that the control group scored higher on their acceptance of thinking errors and speculated that “the incarcerated youth anticipated that prosocial responding to positively valenced items…could yield rewards in the form of better staff treatment or early release from the institution” (Barriga and Gibbs 1996, p. 325).

Finally and most important, survey research suffers from a fundamental artificiality problem. As opposed to the exploratory studies that have uncovered neutralizations in spontaneous explanations of deviant behavior, survey-based studies measure neutralizations almost exclusively in the abstract. Typical neutralization items on a survey include questions such as “It’s alright to physically beat up people who call you names.” Questioning a respondent’s approval or disapproval of criminal behavior—even in select, hypothetical situations like this—treats neutralizations as generalized beliefs rather than personal reconstructions of events from a person’s own life.

Yet, Sykes and Matza argued that neutralizations matter because these cognitive beliefs protect an offender from “serious damage to his self-image” (1957, p. 667). They are techniques for preserving a noncriminal self-concept, despite the commission of criminal acts. If an act has never been committed, and is therefore not a threat to the person’s identity, it requires no neutralization. As Hirschi (1969, p. 208) remarks, “It is in fact in many cases difficult to imagine how the boy could subscribe to the belief without having engaged in delinquent acts.” This argument has been empirically demonstrated in several studies (e.g., Minor 1981; Wortley 1986; McCarthy and Stewart 1998) that indicate that offenders tend to subscribe primarily to neutralizations relating to offenses they had committed.
Pencil-and-paper questionnaires regarding abstract neutralizations may miss the real cognitive insight of neutralization theory: the way people reconstruct and schematize their own past lives can have an important impact on their future behavior (McAdams 1993). Causal schemata like explanations and accounts are highly personalized phenomena based in salient episodes in a person’s life experience (Dodge 1993, p. 565). Cognitive psychologists argue that our causal beliefs are “storied” (Bruner 1990)—that is, they take the form of narratives and depend upon a person’s lived context and perspective. People use rationalizations to provide their often chaotic lives with a sense of meaning, control, and predictability (Taylor 1989). Abstract questionnaire items may not be able to tap into this aspect of a person’s identity in a meaningful way.

F. A Way Forward?

It may be necessary to rethink future research designs. Neutralizations are dynamic cognitive processes and are specifically theorized as autobiographical accounts used to protect a person’s self-concept from a deep-seated sense of personal shame. Qualitative methods (especially life history interviews) seem ideally suited to understanding them. Unfortunately, the vast majority of qualitative research using neutralization theory has been illustrative instead of theory testing (Hazani 1991a).

Future research should use a mixed nomothetic-idiographic strategy that allows for the development of quantitative indexes derived from qualitative information. Thematic content analysis involves “scoring” verbal material for manifest (rather than latent) content or style, rendering the data comparable across individual cases or between groups (Smith 1992). Conveniently enough, examples of such designs can be found in our own work (Maruna 2001, forthcoming; Copes 2003) discussed in detail in the next section.

The primary advantage of this approach is that it allows for systematic analysis of neutralization content within the spontaneous, everyday language and actual life stories of individuals (Harvey, Weber, and, Orbuch 1990). Peterson (1992, p. 379) describes this sort of thematic

20 This deductive analysis can be supplemented by a more traditional, “grounded” approach to the narrative data. This analysis will add the “flesh” to the reductionist assessment of narrative coherence and help determine whether consistent patterns exist in the overall structure of the life narratives (or “gestalt” narratives) used to “make sense” out of these coded life events.
content analysis as a “particularly good assessment strategy, not a 'second-best' procedure” for assessing attributions and cognitive style. The primary disadvantage of using content analysis as opposed to survey research is that the former is far more time consuming and difficult to analyze.

The use of content analysis techniques also introduces numerous reliability problems stemming from the multiple interpretations of subjective materials. Significant measures need to be taken to protect against bias in the coding process. For instance, independent coders can be trained in the use of published coding manuals, rater-training procedures, decision rules, and practice materials provided by authors of the various systems. Fortunately, several empirically derived, refined, and validated coding schemes are already available for measuring neutralization content of accounts (Schönbach 1990) and an individual’s wider explanatory style (Peterson et al. 1992). To protect against bias, coders need to be blind to each other's scoring, the hypotheses being tested, and the identifying characteristics of the speaker. Scores can also be corrected for correlations with verbal fluency (total word counts) or the number of accounts being described.

By allowing for systematic comparison of patterns of neutralization use across different groups of individuals, this strategy combines the rigor and theory-testing element of quantitative designs with the richness and theoretical validity of qualitative methodologies. This methodology is also able to neutralize the most common criticisms of interview-based and survey-based neutralization research. For instance, the argument that neutralizations are “merely” performances for the benefit of the interviewer (or are otherwise contaminated by the research situation) would carry less weight because this criticism would apply equally across the different individuals and groups in the study. Even if neutralizations are purely self-presentational techniques, it would still be theoretically interesting if some groups of individuals consistently seek to present themselves in certain ways when those in some comparison sample do not. This might even be stronger evidence in favor of the theory if there really is an inherent tendency to neutralize in the research interview situation. Likewise, this design avoids the artificiality and awkwardness of survey-based designs (i.e., why would those who have never done or contemplated an act need to neutralize it?). As long as everyone in the group has committed comparable acts in need of explanation, their use of neutralization techniques in this telling can be measured and compared.
IV. Recasting Neutralizations as a Theory of Desistance

The most significant stumbling point for neutralization theory has been sequencing. Sykes and Matza are clear that in their view neutralizations precede delinquency; otherwise delinquents could not free themselves of the potential harm to their self-concept. Critics like Michael Hindelang (1970) argued that the neutralizations that researchers constantly identify are just “after-the-fact rationalizations” meant to justify wrongdoing. This debate has continued ever since. Although Sykes and Matza’s chronological ordering was supported by Agnew’s (1994) longitudinal study of neutralizations and violent behavior, Cromwell and Thurman do not exaggerate much when they write: “No one … has yet been able to empirically verify the existence of pre-event [as opposed to post-event] neutralizations” (2003, p. 547). Observers like Hamlin (1988) have sided with Hindelang, arguing that neutralizations only really make sense as ex post facto explanations.

This lingering “chicken-or-the-egg” debate could have been settled long ago. Hirschi in 1969 suggested a truce and a way out. He argued that both sides could be right: neutralizations might start life as after-the-fact rationalizations but become the rationale or moral release mechanisms facilitating future offending (see also Cromwell and Thurman 2003). Hirschi (1969, p. 208) describes the acceptance of neutralizations as part of a “hardening process.” Akers reaches much the same conclusion: “Initial acts may and do occur in the absence of definitions favorable to them; rather the definitions get applied retroactively to excuse or redefine the initial deviant acts. To the extent that they successfully mitigate others’ or self-punishment, they become discriminative for repetition of the deviant acts and, hence, precede the future commission of the acts” (1985, p. 60).

Neutralization theory, then, is best understood as an explanation of persistence or desistance rather than of onset of offending. Several authors (e.g., Minor 1981; McCarthy and Stewart 1998) have suggested that neutralization use predicts a process of graduated desensitization or “hardening.” Minor (1984, p. 1017) argues that “we should note that ‘neutralization as a hardening process’ is only a partial theory. Its logical complement might be ‘intolerance (of excuses) as maturational reform.’” In other words, if the acceptance of neutralizations is important for maintaining criminal involvement, then rejection of these

---

21 In the therapeutic literature, cognitive distortions are considered to play a maintenance role rather than a causal function in offending (Murphy 1990).
neutralizations should be associated with desistance from crime. This is certainly the theory behind countless cognitive correctional interventions: take away the neutralizations, and end the behavior!

Alternatively, Maruna (2001, p. 144) argues that “instead of a hardening process . . . the acceptance of neutralizations might even be the first step in a softening process.” If neutralization techniques indicate an acceptance of shared moral values, then their invocation might signal a weak attachment to crime and a willingness to change. Neutralizations might be especially associated with juvenile delinquents because such persons are often on the verge of terminating their typically brief flirtation with criminal behavior. This is a central tenet of Matza’s (1964) notion of drift and Lemert’s (1951, p. 75) formulation of primary and secondary deviation, which suggests that deviance that can be rationalized is likely to be sporadic and temporary, unlike deviance that becomes entwined with a person’s identity.

These hypotheses are fairly easily tested. Longitudinal research could test whether reductions in the use of neutralizations over time predict a reduction in criminal activity (Shields and Whitehall 1994, p. 234). Cross-sectional studies could compare neutralization acceptance in the life history narratives of desisting and active offenders (e.g., Maruna 2001). Here the researcher is not in the awkward position of comparing the neutralization acceptance of a group of “innocents” who have nothing to neutralize. Rather, both groups have plenty of crimes and misdemeanors to account for and explain. If neutralizing beliefs are implicated in persistent offending, then desisting ex-offenders either should not manifest neutralization techniques at all or should do so less than active offenders.

A. Research on Neutralizations and Desistance from Crime

Unfortunately, little systematic evidence to date can either confirm or refute the role of neutralization acceptance in the reform process. Indirect evidence can be gleaned from evaluations of cognitive-based therapy in correctional and drug addiction treatment settings (Allen, MacKenzie, and Hickman 2001). However, the multifaceted nature of these cognitive-based programs makes it difficult to ascertain the precise contribution of changes in social cognitions to successful outcomes. It would be fallacious to assume that, just because an intervention is cognitively based, any resulting behavioral changes can be attributed to shifts in personal cognitions.
Some evidence might be found in reconviction studies that include a cognitive component. Roger Hood and his colleagues (2002) provide some support for the “softening” idea in their important study of 250 persons convicted of sex offenses released from British prisons. Offenders deemed to be “in denial” by the parole board (about one-third of the sample) were much more likely to be rated as “high risk” than sex offenders who admitted responsibility for their offense. However, only one of these “high-risk deniers” was subsequently reconvicted of a sexual offense (compared to seventeen of the ninety-seven nondeniers; a statistically significant difference). This strikingly high rate of false positives suggests that “being in denial” is given more weight as a risk factor by the British parole board than is justified (see also Marshall and Barbaree 1990). Two recent meta-analyses of sex offender recidivism studies (Hanson and Bussiere 1998; Lund 2000) also conclude that official measures of “denial” play a negligible role in predicting recidivism.

Hood and colleagues explain this finding by suggesting that a few of the “deniers” might have actually been innocent of the crimes of which they were charged. Just as important, however, they write: “Some ‘deniers,’ when faced with the stigma of conviction and punishment may not accept their deviant sexual acts as a reflection of their ‘real self.’ Nor may they wish to associate with those they regard, unlike themselves, as ‘real’ sex offenders. It is possible that such persons may be less likely to become ‘secondary deviants,’ that is, persons who accept and seek to justify their sexual deviance” (Hood et al. 2002, p. 387).

Lemert was clear on this point: “The deviations remain primary deviations or symptomatic and situational as long as they are rationalized” (Lemert 1951, p. 75, emphasis added). According to Lemert, an individual does not move into secondary deviation until she or he undergoes “a process of identification” through which the deviant acts are “incorporated as part of the ‘me’ of the individual” (p. 75). In other words, those who cling most tightly to excuses may be the least likely

22 The word “denial” is sprinkled throughout Sykes and Matza’s formulation of neutralization theory (denial of injury, denial of victim, etc.), and denial and rationalization are sometimes equated in the literature (e.g., Cramer 1991). Still, there are differences between neutralizations and outright denial. Barbaree (1991, p. 2), for instance, writes: “Denial and minimization are the results of a psychological process involving distortion, mistaken attribution, rationalization, and selective attention and memory…. Denial and minimization are both products of the same self-serving cognitive processes, but they . . . represent different degrees of the process.”
to persist in criminality, because this would counter their images of themselves as innocents.

Additional evidence might be found in qualitative research on desistance from crime and recovery from substance abuse (see esp. Maruna forthcoming). Considerable research suggests that successfully desisting from crime requires changes at the psychological level as well as changes in a person’s social circumstances (see esp. Shover 1985, 1996; Laub and Sampson 2001; Giordano, Cernkovich, and Rudolph 2002). Desistance is a product both of fortunate social conditions (e.g., opportunities for employment or marriage) and adaptive interpretations and understandings of these processes (see Burnett 1992). Few of these analyses mention a specific role for neutralization techniques. Research by Mischkowitz (1994) is an exception. In his analysis of findings from the Tübingen Comparative Study of Young Offenders in Germany, Mischkowitz (1994, p. 319) concludes, “Using neutralization techniques also enables [the desisting ex-offender] to reconstruct his own biography and modify his past in a manner that is conducive for his present self-concept.” He describes in detail how these “techniques of self-presentation” are usefully employed to find a suitable marriage partner—one of the best known correlates of desistance.

B. Neutralizations and Changes in Commitment

One way to interpret these findings is that the deeper one is immersed in criminal behavior and a criminal lifestyle, the less need one has to neutralize criminal acts. Minor (1981), for instance, argues that neutralization theory and subcultural theories might be made compatible if neutralizations are viewed as a “facilitating element in the gradual (or not so gradual) process of becoming committed to unconventional norms” (Minor 1981, p. 301). Thus, in early stages of delinquency, youths may need to use neutralizations to relieve the cognitive dissonance that occurs when their actions are not in line with their values. By using these neutralizations, delinquents’ commitment to those conventional values are eventually weakened to the point that there is no longer a need to neutralize (Hirschi 1969, p. 208). As Minor (1984, p. 1018) notes, “over time, either the desire or the moral disapproval should dissipate, leading one to either conformity or guilt-free deviance.”

If this is correct, neutralization use should be most commonly associated with individuals in a state of drift who are partly committed
to criminal lifestyles and partly to mainstream values. An absence of neutralization should be associated with those who are strongly committed either to the mainstream or to criminal activities. This curvilinear relationship between criminality and neutralizations might account for some of the mixed empirical support for the theory (Topalli 2003).

The little evidence on this hypothesis is mixed. Using the Richmond Youth Survey, Costello (2000) examined the relationship between parental attachments and use of neutralizations; delinquent youths who were strongly attached to their parents were less likely to use neutralizations than those with weaker parental attachments. Mitchell, Dodder, and Norris (1990) surveyed 694 university students and found that church attendance (a measure of socialization) was negatively related to the use of neutralizations. According to Thurman (1984), use of neutralizations had the greatest effect on deviance among the “least morally committed.” When violating the norms of society, those who are strongly committed to the values of conventional society may experience levels of guilt that are too strong to overcome by simple neutralization techniques.

However, a substantial body of research seems to show that many offenders are mostly committed to their misdeeds (e.g., Schwendinger and Schwendinger 1967; Hindelang 1970, 1974; Regoli and Poole 1978; Sheley 1980). Even Matza (1964) agrees that a small proportion of offenders become unconditionally committed to delinquent values. Since these committed offenders are unattached to conventional values there may be nothing for them to neutralize. Dodder and Hughes (1993) found that college students who did not consider underage drinking to be wrong were significantly less likely to use techniques of neutralization than college students who did.

Agnew (1994) has argued that many of these studies improperly measure the level of general approval for deviance. Research tends to ask whether young people “approve” of deviant acts (Hindelang 1970; Regoli and Poole 1978) or whether they feel the acts are good or bad (Calhoun 1974), “wrong” (Massey and Krohn 1986), “acceptable” (Johnson 1979), “OK,” or “alright” (Rankin 1976). Citing script theory (Abelson 1981), Agnew argues that a survey respondent is likely to answer such questions in the context of his or her own social world. When asked if they approve of fighting, their response will reflect their views about “fighting as they know it” (Agnew 1994, p. 559). This strategy is likely to confound individuals who approve of fighting more
generally with those who approve of fighting only conditionally and hence may not be a real test of neutralization theory.

Attachment to mainstream values is probably better represented using measures uncontaminated by the mention of the crimes themselves. An example can be found in Copes (2003). Copes explored the nature and frequency of neutralization among auto thieves who were on probation or parole when interviewed. The guiding question was whether commitment to social norms predicted the use of neutralization techniques. Semistructured interviews focused on the offenders' accounts of their crimes, various aspects of criminal decision making (e.g., target selection, risks of crime, etc.), and current lifestyles. Measures of societal attachment included marriage, stable employment, and high school or equivalent education. Copes divided his sample of forty-two individuals on parole for auto theft into high-attached and low-attached groups. He then systematically coded the qualitative interview transcripts of members of both groups for their use of neutralization techniques. Offenders in the high-attached group were more likely to use neutralizations when accounting for their offending than were the low-attached offenders. The two groups relied on different techniques to account for their crimes. High-attached offenders most often relied on appeal to higher loyalties, while low-attached offenders used denial of victim. When the two groups used the same techniques, they did so in different ways. When denying the victim, high-attached offenders focus on victims being careless or foolish, while low-attached offenders focus on victims as being deserving because of their actions directed toward the offenders (Copes 2003, p. 121).

There is also evidence that high-rate offenders use neutralizations less frequently than low-rate offenders (Minor 1981, 1984). McCarthy and Stewart (1998) found that the number of excuses an individual accepted decreased with subjects’ rate of offending. Low-involvement property offenders reported a higher use of neutralizations than high-involvement property offenders. Copes (2003) found that high-frequency and low-frequency auto thieves were equally likely to elicit neutralizations; two-thirds of each group used at least one neutralization. The two groups, however, differed in the types of neutralizations they used.

Age is a potentially important factor in the evolution of neutralization use. There is evidence that young and older people neutralize their offenses differently. Minor’s (1981) reformulation of neutralizations as a hardening process suggests that older, more experienced offenders
might rely on neutralizations less often than younger offenders. Hollinger (1991) examined the relation between neutralization acceptance and employee deviance among young (twenty-five years old and younger) and old employees (over twenty-five). He found that employees over twenty-five years old relied more heavily on denial of injury than did employees twenty-five and younger. This led him to conclude that “neutralizations may work best among older adults, not among the younger individuals with which [the theory] was tested in the past” (Hollinger 1991, p. 196).

Finally, the mediating variable between excuse making and criminal behavior—say, shame, self-esteem, mood, or affect—might have differential effects for juveniles and adults. Gudjonsson (e.g., Gudjonsson and Bownes 1991) has consistently shown that older offenders score higher on measures of guilt than younger ones. He attributes this to both social desirability and age-related increases in introversion. Alternatively, juveniles with high self-esteem may be more likely to participate in criminal behaviors, which are developmentally normative for that age group (Moffitt 1993), while adults with high self-esteem may be more likely to participate in the legitimate worlds of work and family. If so, neutralizations may promote delinquency (by increasing hubris) but have no great impact on adult criminal behavior. Some evidence for this was found in research by McIvor, Murray, and Jamieson (2004), who found that self-esteem levels were not useful in distinguishing between persisters and desisters at younger ages but might predict which offenders would persist at older ages.

C. A Study of Desistance and Explanatory Style

The Liverpool Desistance Study (LDS) looked at the role of neutralizations (Maruna 1998, 2001) and explanatory style (Maruna forthcoming) in desistance. The original purpose was to understand how exconvicts “make sense” of their lives, not to predict behavioral outcomes. Still, the LDS sample provides unique, but certainly not ideal, data for empirically exploring the relationship between explanatory style and desistance. Although not a prospective study like Burnett (1992), the Liverpool study contains considerable evidence about how reformed, former offenders think (not just how we imagine

23 Some research has attempted to examine whether the use of neutralization techniques is related to one’s stage in Kohlberg’s (1964) theory of moral development (Radoevich and Krohn 1981; Lanz-Kaduce, Radoevich, and Krohn 1983).
they might think). The study also contains a comparison sample of active offenders, allowing us better to isolate what thinking patterns are uniquely associated with the process of self-transformation or desistance from crime, and what patterns are common to all former offenders.

The LDS involves ongoing collection of life stories of a snowball sampling of ex-convicts with extensive criminal records. More than 100 ex-convicts have agreed to share their personal autobiographies with researchers associated with the University of Liverpool (see Canter et al. 2001). Approximately fifty-five have been classified as desisting from crime. These were individuals who were once long-term, habitual offenders, but who at the time of the interview reported having been crime free and drug free for over a year, and, importantly, reported no plans for future involvement in criminal behavior (for a more extensive discussion of sampling, see Maruna [2001]). On the other side, thirty-four have been classified as persisting or active in their criminal careers. They reported recent criminal activity and admitted to plans to continue selling drugs, robbing convenience stores, and so forth.

Following admonitions of Polsky (1969) and Wright and Decker (1994), all of the participants were out of “captivity” and free in the community. The “active group” consisted of individuals who are actively involved in criminal behavior and not drawn from a captive population. This sample is far more likely to provide insights into “criminal thinking” than is a group of prisoners who might not have committed a crime in years. The desisting and the persisting samples mirrored one another as closely as possible on static variables such as year of birth, gender, types and number of crimes committed, age of criminal onset, and high school completion.

Life story interviews (two hours on average) with both groups were tape-recorded and transcribed, preserving the original language of the interviewees. These transcripts were then content analyzed using Peterson, Schulman, Castellon, and Seligman’s (1992) CAVE (Content Analysis of Verbatim Explanations) system. The CAVE system is an innovative and well-established method for measuring cross-event consistency in the explanations that individuals provide for positive and negative events in their lives. The CAVE method has been used in studies of depression, precursors of mental illness, and the success of presidential candidates. This previous research provides strong support for the construct validity of the CAVE technique, and coders trained
by the system’s authors have achieved interrater reliability levels exceeding .90 (Peterson et al. 1992, p. 386).

Two graduate students, blind to the hypotheses of this research, were trained in Peterson’s method for extracting attributions from transcribed interview material (see Peterson et al. 1992, pp. 383–86). At least four negative life events and a minimum of three positive life events were identified in every full-text transcript using the conservative and specific criteria for extractions specified by Peterson and colleagues. Once these passages were extracted, two additional graduate students, also blind to the hypotheses, coded the explanations using Peterson’s content coding scheme. After a two-week training process using pilot interview transcripts, the coders were provided with long lists of unidentified attributions, extracted from the larger context of the life story interview and randomized within and between subjects. Because of the various precautions taken, raters were not biased by previous ratings for the same subject and would have no way of easily connecting any series of passages. Most important, coders had no way of knowing whether the speaker of any particular passage was an active offender or a desisting ex-offender. Any passage that mentioned desistance or persistence in crime and in the present tense (e.g., “That is what has kept me straight these last few years” or “which is why I am still selling drugs today”) was excluded from the coding or modified in such a way as to remove the reference to desistance or persistence.

Coders rated each extracted attribution on three dimensions (internal, stable, and global) using a scale of one to seven, with a seven representing the highest score. Over 1,250 separate attributions, slightly more than fourteen on average in each of the eighty-nine interviews, were extracted and then double coded by separate raters on all six key dimensions of explanatory style. This was painstaking and highly labor-intensive. The two independent scorers achieved a correlation of .79 in their coding of these extractions.

As hypothesized, internal, global, and stable explanations for negative life events were negatively associated with desistance, and internal, global and stable explanations for positive events were all positively associated with reform. Although all six measures had modest, zero-order correlations with criminal reform in the expected direction, the logistic regression indicated that only the three dimensions of positive explanations and the negative-internal dimension had unique and significant contributions ($p < .05$). The odds that a participant in this sample is in the desisting group more than double for each unit
increase in their positive-internal, positive-stable, or positive-global explanations. The implication is that the more individuals are able to attribute positive life events to broad, long-lasting personal qualities (e.g., “Because I am a worthy individual”), the greater the odds that they will stay crime-free.

Likewise, and more controversially, the acceptance of highly internal attributions for negative behavior was significantly associated with persistence in crime. The odds of being in the desisting group diminish by a factor of 1/0.45 (or 2.22) for each unit increase in negative-internal explanations. Consistent with labeling theory (and a “softening hypothesis”), this suggests that someone might be less prone to desist to the extent that negative events are seen as originating from internal sources (“This is just the way I am”)—especially when these are stable (“I’ve always been this way”) and global (“I fail at everything I do, no matter where I go”) characteristics.

Maruna (2001) argues that the self-narratives of persistent offenders closely resemble those of depressed persons. Utilizing a “condemnation script,” long-term, active offenders portray themselves as “doomed to deviance.” The mindset consistent with persistent offending, therefore, may be something like what Abramson, Seligman, and Teasdale (1978) refer to as “learned helplessness” or a sort of hopeless, “poor me” justification for making no effort to change self-defeating behaviors.

Those who were able to turn their lives around and desist from crime and addiction were better able to “distort” the often grim realities of their past lives. One aspect of this sort of distortion was a tendency to externalize the blame for their past acts and minimize their own internal deviance. These accounts ranged from “act adjustment” (e.g., “We never really hurt anyone”) to “actor adjustment” (“I wasn’t ever really as bad as they say”). The most common story took the shape of what Cohen (2001, p. 62) calls the most “radical” actor adjustment: the narrator denied responsibility by “attributing the action to another part of the self that has been disengaged from the ‘real’ me.”

On the surface, this appears to contradict a central component of neutralization theory: that denial of responsibility should be related to offending. Although all LDS sample members used a litany of neutralizations to account for their past behaviors, desisting ex-convicts did not seem to use any fewer (and judging by their scores on the internal-external dimension, might even have used marginally more) than active or persisting offenders. This was confirmed in a separate analysis (see
Maruna 1998, pp. 135–55) that utilized a modified version of Schönbach (1990) to code for excuses and justifications (as well as concessions, denials, and objective reports). That analysis, using a similar methodology as the CAVE analysis, also failed to uncover any differences between desisting or persisting participants in use of excuses or justifications per 1,000 words of dialogue.

This finding needs to be interpreted cautiously. The LDS sample has a variety of important limitations that confine the generalizability of these findings. Although similar demographically to other samples of offenders (see Maruna 2001, pp. 57–71), the LDS sample has one major difference. With a median age of thirty (mode of twenty-nine), the Liverpool sample is older than the “average” group of offenders. Neutralization theory was a theory of juvenile delinquency, and some research suggests that very young offenders may have quite different cognitive patterns than the persistent, long-term offenders (many are struggling with long-term addictions) in the LDS sample. It is possible that adolescents (and in particular “adolescence-limited offenders”) have very different ways of rationalizing criminal behavior than do long-term adult offenders (see Dodge 1993; Moffitt 1993; Gibbs, Potter, and Goldstein 1995; Barriga et al. 2000).

Finally, there is no evidence in Maruna’s cross-sectional study that suggests a causal relationship between styles of thinking and continuing criminal behavior. Although there seems to be a correlation here, it may be that a change in criminal activity led to the apparent change in thinking rather than the reverse. Perhaps persistent offending causes depression, and this level of dysphoria accounts for differences in thinking patterns. (The LDS protocol contained no independent measure of mood or mental health.)

A better research design would follow a single set of individuals involved in crime or deviance over time and measure how changes in their thinking patterns correlate with changes in their offending. Such a design may or may not disentangle the magic question of which comes first: changes in thinking or changes in behavior? (see Laub and Sampson 2001). Such a question might not even be answerable if, in fact, the relationship between neutralizations and behavior is more interactive and mutually influential (see Rumgay 1998). It would, however, allow for a much better test of the hardening/softening hypothesis than is possible from the LDS and other existing data sets.
V. What We Already Should Know
There is a great deal to discover and learn about the functions and
effects of neutralization techniques in deviance. On numerous issues,
there is ample evidence and reasonable consensus. On many issues,
however, neutralization theory in the research literature is vexed and
confused. If our review of the literature is accurate, however, neutral-
ization theory is frequently misused in the applied world of correc-
tional treatment. Below, we try to allay some of this confusion by
outlining aspects of neutralization theory that might have been
controversial fifty years ago, but by now should largely go without
saying.

A. Neutralization Theory Is a Theory of Social Cognition
Today’s neutralization theorists can primarily be found in depart-
ments of psychology studying “human aggression” and issues such as
“information processing” (Dodge 1993), “moral disengagement”
(Bandura et al. 1996), and the “cognitive mediators” of violence (Slaby
and Guerra 1988; Gibbs, Potter, and Goldstein 1995). No one would
know this from reading the criminological literature.

The basic assumptions behind neutralization theory have been long
understood to be “much more psychological than sociological”
(Hamlin 1988, p. 427). This research needs to be better tied into
the psychological literature to remain viable. It should go without
saying that research on offenders’ accounts and explanations for their
behavior relate to wider research on the accounts and explanations
that all humans offer. Research subjects who have sent threat letters
through the mail or downloaded snuff videos from the Internet are
presumably subject to the same cognitive processes (e.g., dissonance,
shame, ego defense) as the rest of us. Literature reviews for neutral-
ization theory, therefore, need to expand beyond demonstrations of
neutralization techniques in convenience samples and instead con-
sider the well-researched psychological processes underlying these
techniques.

Neutralization theory should be seen as a theory of narrative sense
making and hence part of the process of identity construction
(McAdams 1993). Aronson (1992, p. 304) makes a similar point in
arguing that cognitive dissonance theory is “essentially a theory about
sense-making—how people try to make sense out of their environ-
ment and their behavior—and, thus, try to lead lives that are (at least
in their own minds) sensible and meaningful.” Understanding how
humans construct this sort of meaning out of their lives is at the heart of the cognitive revolution (Bruner 1990) and has opened innumerable directions for new research and improved understanding of human behavior. If sense-making research in the interpretivist tradition is largely disregarded or even stigmatized in criminology (as argued forcefully by Athens [1997]), it is so much the worse for our discipline.

B. Neutralization Theory Is Not Just a Theory of Social Cognition

The roots of neutralization theory are not in cognitive psychology. They can be traced to an interesting mix of symbolic interactionism and Freudian ego psychology. If this history is lost, and neutralizations are treated simply as “processing errors,” “faulty reasoning,” or “cognitive deficits,” much of the richness of the theory is lost.

The origins of the theory stress the social nature of neutralization and the origins of these techniques in the wider culture. Changing such cognitions in individual offenders may be less a psychological or clinical matter than a sociological one. The “internal soliloquies” of offenders are drawn from “a repertoire of culturally acceptable legitimations” (Murphy 1999, p. 205) and “do not materialize out of thin air at the individual’s discretion” (Hamlin 1988, p. 431). Cohen (2001, p. 59) writes, “An account is adopted because of its public acceptability. Socialization teaches us which motives are acceptable for which actions.” Neutralizations, then, may say more about a culture than about an individual (Mills 1940).

Neutralization theory, unlike some of its counterparts in cognitive theory, is a theory of both motive and motivation, not simply information processing. According to Cohen (2001, p. 42), “The cognitive revolution of the last thirty years has removed all traces of Freudian and other motivational theories. If you distort the external world, this means that your faculties of information processing and rational decision making are faulty.” Neutralizations are not “mistakes.” The theory explicitly states that they are cunning, unconscious mechanisms needed to ward off threats to one’s ego. Any attempt to “correct” these well-honed ego-defense skills and strategies in the name of therapy has to keep this function in mind.

—-24—- Bruner (1990, p. 2), considered by many the “father” of the cognitive revolution, makes much the same point about the narrowness of much contemporary research on cognition.
C. Neutralization Theory Has to Be about More than a List of Five Techniques

Sykes and Matza (1957) originally named five different neutralization techniques. This list served them well, making the theory readily accessible and easy to remember. Yet, researchers need to stop reciting these five techniques as if they were gospel. There is absolutely nothing magical about the techniques Sykes and Matza listed.

One could make a strong case that the original techniques are not conceptually distinct enough from one another to be considered different categories. For instance, denial of victim and denial of injury are very similar and go hand in hand (Landsheer, t’Hart, and Kox 1994). When offenders deny people victim status, they deny the amount of harm done. Some auto thieves steal only from “rich” people because they believe these people can afford the loss, which means they are also able to deny the amount of damage done because the victim can easily replace the stolen items (Copes 2003). Thus evidence exists to support the contention that the individual techniques are in need of further refinement. Relying on broad, unclear categories presents problems for future researchers attempting to operationalize the techniques and may be one reason for the inconsistent results of empirical evaluations.

Subsequent researchers have identified dozens (even hundreds, depending how finely one cuts them) of techniques that seem to serve the same function as neutralization techniques. What is interesting about neutralization theory is this function (what the neutralizations do), not the flavors it comes in. If research requires a taxonomy of neutralization “types,” Scott and Lyman (1968) or Schönbach (1990) provide categories that are much more theoretically precise. Preferably, however, the individual use of specific neutralizations should be understood within the wider context of sense making that is the self-narrative process.

D. Neutralizing Negative Behaviors, in Itself, Is Not Interesting

Journal editors should be wary of publishing articles reporting the unremarkable finding that people who do disreputable things use neutralizations to account for them. This is not news (nor is it, alone, evidence in favor of neutralization theory). A study that demonstrates that incompetent surgeons make neutralizations for their shortcomings may contribute to understanding of the subjective experience of medical malpractice, but they contribute little to neutralization theory. Plenty of research shows with some certainty that providing ex post facto
excuses is normative, socially rewarded behavior (Snyder and Higgins 1988). As Cohen (2001, p. 249) concludes, “Instead of agonizing about why denial occurs, we should take this state for granted. The theoretical problem is not ‘why do we shut out?’ but ‘why do we ever not shut out?’”

At the same time, simply finding that some offenders do not make excuses (those who say, “Nobody forced me, I did it for the money”) is not, in itself, terribly interesting, either. Everyone makes excuses, true, but we do not make excuses in every situation. Future research could more usefully investigate the nature of neutralization use in contrasting situations, circumstances, contexts, and cultures.

E. Neutralizing Negative Behaviors, in Itself, Is Not Pathological

Excuses and justifications enjoy the awkward position of being “universally condemned while being universally used” (Schlenker, Pontari, and Christopher 2001, p. 15). Correctional research and various offender therapy programs (often designed around the mantra of “no more excuses!”) have started classifying use of neutralization techniques as a form of “criminal thinking.” At the same time, the psychological literature on excuse making is so overwhelmingly positive that it has led to something of a backlash pointing out the potential downside of this form of psychic defense. What should be unquestioned, however, is that neutralization techniques are as common as breathing. Stan Cohen, for instance, argues that to deny painful truths about our own behavior is “part of being human” (Cohen 2001, p. 37) and that “every personal life and every society is built on denial” (Cohen 2001, p. 295). Indeed, we probably couldn’t manage in a world without neutralizations (Rumgay 1998, p. 207).

The “normality of neutralizations” was always explicit in the early writing. The techniques themselves are “widely available” (Matza 1964, p. 50), “conventional” (Matza 1964, p. 91), and even patterned after the explanatory accounts of social workers, defense attorneys, and positivist social scientists (Matza 1964, p. 61; see also Matza and Sykes 1961). Redl and Wineman (1951, p. 145) emphasize that use of their “tax evasion” techniques is anything but limited to delinquents. “On the contrary, the defense of impulsivity…is a task which, up to a certain degree, every ego is summoned to perform at times.” Scully and Marolla (1984), in their study of rapists’ neutralizations, take this point farther, implicating the tendency to pathologize neutralizations in the process of deviance generation itself. They argue that “cultural
perspectives” (or ideas widely popularized in the popular culture) “and not an idiosyncratic illness” seemed to motivate the behavior of the sex offenders in their sample. Further, they argue that “the psychiatric perspective” of pathologizing sex offenders “has contributed to the vocabulary of motive that rapists use to excuse and justify their behavior” (p. 542).

Describing neutralization techniques as evidence of a “criminal personality” can only be explained as a misreading of Sykes and Matza’s work and the social cognition research that followed it. Central to the new notion of “criminal thinking” or the “criminal personality” is that “criminals do not think like law-abiding prosocial people” (Sharp 2000, p. 2).

Yet, the idea that “good people” do not make excuses is simply unsubstantiated. In his review of thirty-eight studies, Zuckerman (1979) found substantial confirmation for the idea that all of us make predominantly external attributions for our failures and predominantly internal attributions for our successes. This is perfectly normal behavior; considerable research has shown the personal and interpersonal benefits of this process of blame deflection (Snyder and Higgins 1988; Seligman 1991).

Habitually abusing a child might be pathological behavior. Accounting for this behavior through an appeal to excuses and justifications (“I was drunk,” “My parents did the same to me”) is, on the surface at least, not necessarily pathological. A case could be made that persons who abuse their children and provide no neutralization for the behavior (“I don’t know why I do it” or “I abuse children because I want to and I enjoy it”) are equally if not more in need of psychological attention than parents who do provide justifications.

The sociological literature contains many examples of nonoffenders who use neutralization techniques routinely in their self-accounts. One interesting example is victims of crime. Like offenders, victims seem to rely on neutralizations to protect their self-identity from the threat of shame (see Ahmed et al. 2001). Ferraro and Johnson (1983, p. 328), for instance, extended Sykes and Matza’s concepts by providing six neutralizations that battered women use: appeal to the salvation ethic, denial of the victimizer, denial of injury, denial of victimization, denial of options, and appeal to higher loyalties. Higginson (1999) likewise found that teenage mothers used a number of neutralization techniques in accounting for their experience of statutory rape.
Another group well-known for using high levels of rationalizations and justifications are social and political leaders. As Stanley Cohen (1985) demonstrated in his catalog of “controltalk,” the administrators of criminal justice systems worldwide have no problem employing a wide array of neutralizations in rationalizing and justifying their own budget expansions, human rights violations, and day-to-day business. Similarly, as Cohen chronicles in his later work, *States of Denial* (2001), politicians, military commanders, and captains of industry might be said to owe their careers to their ability to neutralize. These examples certainly do not prove neutralizations are a “good thing,” but they call into question the notion that neutralizations are a sign of a “criminal personality” (Yochelson and Samenow 1976) and the “spuriously exact ‘scientific’ classification and a fetishized ‘medical’ diagnosis” that accompanies this notion (Cohen 2001, p. 35).

Pathologizing offender neutralizations as “cognitive errors” often seems to verge on a type of “fundamental attribution error” writ large. The fundamental attribution error is the tendency to attribute our own behavior to circumstances, but others’ behavior to personal traits (Jones and Harris 1967). If I am driving recklessly, it is because my daughter is distracting me, I’m under a lot of stress, and so on. Yet, if I see a car in front of me driving recklessly, I immediately assume the driver is some kind of depraved nut.

The same process happens with offenders and the “rest of us.” “We” are allowed to use explanations to account for why we do bad things, but “they” have “no excuse” (see also Saulnier and Perlman 1981). Their excuses are simply more evidence of the depth of their pathology (Fox 1999a). Pathologizing excuse making and trying to prohibit the use of neutralizations in correctional programming, then, seems an iatrogenic strategy for the creation of widespread personality “sickness.” If the only criterion for the diagnosis is an external locus of control in regard to wrongdoing, then all of us suffer from “criminal thinking” and “criminal personalities” (see also Beech and Mann 2002, p. 265).

The criminal justice community’s seemingly deep-seated desire to make offenders take “full responsibility” for their behavior may result from cognitive dissonance involved in criminal justice work. If we are to punish (or arrest, convict, study, classify, etc.) a person as an “offender,” the individual needs to be responsible for the offense. In the face of a body of social science work that exculpates offending behavior by shifting blame to parents, schools, communities, and capitalism (among other forces), there is no small comfort in having the individual him or
herself take full responsibility for the crime. The idea that “offenders” are only a stand-in or scapegoat population—symptoms of an unequal and unfair society that we have partly created and fully benefited from—may be too uncomfortable for most of us, especially those of us working in the system (Tidmarsh 1999, p. 50). This may explain why forcing confessions out of wrongdoers has been a part of justice systems since long before the discovery of neutralization theory or cognitive therapy (Foucault 1988, p. 42).

F. Neutralizations Do Not Necessarily Have to “Cause” Delinquency to Be Important

Although we think that neutralization theory has value as an explanation of persistence and desistance, there is a chance that future research will falsify this theory and demonstrate that neutralizations are not consistently related to future offending. But, predicting future behavior is not the be-all and end-all of criminological research. Criminologists have many reasons to examine the nature and dynamics of neutralization. As Hamlin (1988, p. 429) points out: “Why delinquent behavior has to be tied to neutralization in a causal sense can only be attributed to the positivism inherent in U.S. sociology and, perhaps, the perceived importance of instrumental reason in modern industrial societies.”

Following Mills (1940), Hamlin argues that the value of studying neutralization techniques is not in the social-psychological aspects of the theory, but in what the particularities of these rationalizations tell us about “the historical-structural context in which the motives are elicited and manifest themselves” (Hamlin 1988, p. 434; see also Laub and Sampson 2003). Even if it is not the Holy Grail for the explanation of criminal etiology, or even persistence, exploring neutralizations has helped us understand how the human mind works (as has long been apparent outside of criminology in work as diverse as Aronson [1968], Hazani [1991b], Bandura et al. [1996], and Orbuch [1997]). This is no small thing. Even if neutralization theory is consistently falsified as a predictive theory, neutralization techniques themselves should hardly be abandoned by those seeking to understand criminality.

Yet, this much need not be conceded. Neutralization theory is not sufficient alone as an all-encompassing explanation for criminal behavior and has been understood, almost from its origins (see esp. Matza 1964), to be only one part of a wider theory. Neutralizations were never
meant to “cause” offending, only to allow for it (see esp. Matza 1964). One important contribution of Matza’s (1964) work is the notion of “soft determinism” (see, e.g., Brezina and Piquero 2001), in which cause and outcome are not conceived as discrete entities, but are “interrelated and overlapping, such that some part of cause is constituted by some part of the event produced in part by it and vice versa; but all of the event is not all of the cause (and vice versa)” (Henry and Milovanovic 1996, p. 126).

The chicken-or-egg question that haunts neutralization theory might not be such a problem after all. Criminality and cognition are ongoing processes with no fixed chronological “beginnings” or “endings.” Subtle changes in cognition (as “cause”) and changes in desistance (as “effect”) are likely to overlap in time, with cause affecting the effect and the effect affecting the cause—calling the utility of both constructs into question. For instance, it is not always clear if pessimistic thinking increases the risk of depression or if depression increases the risk of pessimistic thinking (see, e.g., Whisman and Pinto 1997). Weiner and Graham (1999, p. 605) write: “Answers to a question such as ‘Why have I failed?’ surely can affect self-esteem (consider the consequences of the answer ‘I am stupid’). In addition, self-esteem is likely to influence the answer to that question.” In personality psychology, this mixing of the chicken and the egg—what Volkan Topalli (2003) calls the “quiche” model—is not viewed as terribly problematic. Dan McAdams writes: “The truth probably lies with a little bit of both possibilities—cognition as both a cause and a result of the personality constellation we call depression, personality and cognition influencing each other, linked in an intricate web of mutual causation” (1994, p. 511).

Establishing a strong correlation between such thought patterns and behavior may be the best this research can hope to accomplish. This seems a perfectly reasonable role for neutralizations to play, and should not be seen as detracting from the power of cognition. Just the opposite: the proximity between behavior and cognition should make offender thinking patterns an obviously crucial area of research.

VI. What We Still Need to Find Out
There are many questions left to ask about neutralizations and their role in criminality. We conclude by listing some of the most important for the next generation of neutralization research.
A. “Good” Neutralizations and “Bad” Neutralizations

Currently, the working assumption in neutralization theory research is that all neutralizations are created equal: essentially any neutralization is a bad neutralization. This blanket treatment is unfortunate and might be one reason why neutralization theory has failed to gain a great deal of empirical support. It would make sense for future research to try to identify the elements that make some neutralizations adaptive and others nonadaptive (Schlenker, Pontari, and Christopher 2001, p. 25). A priority might be identifying which neutralizations are the “most toxic” and separating these out from the more neutral or even benign examples. Some contenders for the worst of the worst might be “dehumanizing one’s victims,” “seeing the world as hostile,” and “labeling one’s self as ‘naturally’ deviant,” yet all have to be empirically tested.

Considerable research suggests that the dehumanizing and demeaning of one’s victims promotes and allows for further offending. This can be seen in both laboratory research on aggression (e.g., Bandura, Underwood, and Fromson 1975) and in systematic studies of mass violence (e.g., Kelman 1973). Kelman and Hamilton (1971, p. 338) cite Lieutenant William Calley’s description of the massacre in My Lai: “I did not sit down and think in terms of men, women, and children. They were all classified the same . . . just as enemy soldiers.” Likewise, Ahmed and colleagues (2001) found that angry neutralizations that blame others for one’s problems, creating scapegoats—the process they call “shame displacement” or converting shame into “humiliated fury”—were correlated with hostile and aggressive behavior among young people. Matza (1964, p. 102) writes, “The cry of injustice is among the most fateful utterances of which man is capable.” Moreover, there is considerable support for the relationship between aggression and a “hostile attribution bias” (Dodge 1993) by which aggressive school children misinterpret social cues, perceiving external threat and aggressive antagonism under ambiguous conditions. It is an open question, however, whether such perceptions are more or less criminogenic than other forms of “denying the victim” or than other neutralizations (see also Barriga et al. 2000).

Labeling theory would predict that neutralizations that appeal to some self-label (“I had to steal the drugs because I’m an addict” or “I’m just a kleptomaniac, it wasn’t my fault”) might be the most criminogenic accounts. This assessment is not limited to criminologists of a labeling-theory persuasion. Schlenker, Pontari, and Christopher (2001,
p. 28) conclude: “Excuses that attach a label to the excuse-maker . . . are potentially disastrous in their long-term, negative repercussions for the excuse-maker” (see also Higgins and Snyder [1989] on “self-defeating excuses”). Although they disagree on several key points, shame theorists Tangney (1995) and Braithwaite (1989) agree that the internalization of self-stigma is the most damaging form of shame one can experience.

Even this neutralization can be found among nondeviants, indicating that the technique might not always be a bad thing. The most obvious example is the tendency toward “deviance avowal” (Rumgay 1998, p. 82) among persons recovering from alcoholism and other addiction problems. Alcoholics Anonymous members are encouraged to accept that they are “alcoholics” and to incorporate this stigmatized label into the core of their self-identity (i.e., “My name is Bob, and I’m an alcoholic”). Accepting a so-called sick role and attributing one’s problems to being an alcoholic, mentally ill, or suffering from problems like attention deficit disorder can protect an individual from the stigma of being morally corrupt or weak willed. The stigma of having a disease is, after all, far less than the stigma associated with character flaws such as selfishness, laziness, or sloth. Nusbaum (1983, p. 229) writes: “Ultimately, the illness model allows for greater reality negotiation on the part of deviant drinkers. They may choose to negotiate reality by openly avowing their deviance through self-labeling and adopting a repentant-deviant role.”

Some critics of the “disease model” of addiction (e.g., Peele and Brodsky 1991) argue that this protection against stigma comes at too high a cost. By accepting medical explanations for their behavior, deviants are essentially given free range to relapse back into negative behavior patterns (“It can’t be helped, I am sick after all”). Roman and Trice (1968, p. 245) write, “The expectations surrounding these sick roles serve to further develop, legitimize, and in some cases even perpetuate the abnormal use of alcohol.” The consequences of self-labeling for different individuals in different circumstances is an open question and one that deserves considerable research.

**B. Looking Beyond Locus of Control**

Future neutralization research needs to transcend the overly simplistic and long passé (see Weiner and Graham 1999) notion that an internal locus of control is always better than an external locus of control and develop a more complex understanding of neutralizations.
One suggestion from the psychological literature is to look beyond issues of internality-externality altogether and consider other dimensions of attributions, such as stability, globality, intentionality, and controllability. These other dimensions have been found to be more predictive (see, e.g., Wilson and Linville 1985; Peterson 2000). Ahmed writes: “In the social-developmental literature, attributions of stability and intentionality in the context of wrongdoing have been associated with maladaptive outcomes (e.g., shame, anxiety, despair). In contrast, attributions that connote expectations of change for the better in the wrongdoer . . . and avoid labeling individuals as possessing global personality deficits have been associated with positive outcomes (e.g., pro-social behavior, empathy, self-esteem)” (Ahmed et al. 2001, p. 258).

The stability dimension is a potent area for future neutralization research. Regardless of where one locates the source of his or her offending (e.g., in society or deep inside one’s own soul), a feeling that the cause is permanent (i.e., “and that is the way it is always going to be”) does not bode well for the ability to change. Following Abramson, Metalsky, and Alloy (1989) and others, Maruna (2001) proposed a subtype of persistent offending based on hopelessness, with criminality perpetuated as a consequence of pessimistic (stable and global) attributions. The idea of being “doomed” to deviance refers to the belief that one is somehow stuck in a deviant life with no hope of escape.

C. An Optimal Balance of Control

Another implication following from accepting the “normality of denial” is the idea that there may be an optimal balance of control (see Tittle 1995), in terms of offending, between accepting and denying one’s personal responsibility. As Cohen writes, the academic understanding of denial is rather pragmatic at the moment: “If the patient gets better, we say that denial has been healthy; if worse, then denial was pathological” (Cohen 2001, p. 31). We do not know, however, whether there is a “right” amount of denial and responsibility. Similarly, Baumeister, Dale, and Sommer (1998, p. 1114) argue that self-denial might have both healthy and pathological versions: “Mild forms of defense may bolster self-esteem, minimize emotional distress, and thus facilitate mental health and adjustment, while stronger forms could have the opposite effect.” Mental health would thus be bimodally distributed with an “optimal margin of illusion.” Identifying this
optimal level of responsibility and irresponsibility is a priority for the next generation of research.

The challenge each of us faces is to integrate negative life events into our self-narratives without making the self out to be a bad person (see also Lofland 1969, p. 282). Baumeister and Wilson (1996, p. 324) suggest that “one solution” is “to construct temporal boundaries that separate immoral actions from one’s present self, whereas one may acknowledge guilt and blame for them in the past.” Drawing lines between an old, negative self and a new, positive self is a process frequently identified in research with ex-offenders. Scully and Marolla (1984, p. 542) found that the two dominant excuses offered by sex offenders were drug or alcohol intoxication and emotional problems. They suggest that ex-offenders used these excuses “to negotiate a moral identity for themselves by viewing rape as idiosyncratic rather than typical behavior.” This allowed them to “reconceptualize themselves as recovered or ‘ex-rapists,’ who had made a serious mistake which did not represent their ‘true’ self” (Scully and Marolla 1984, p. 542).

Another interesting possibility for understanding the optimal balance of responsibility can be found in the addiction recovery literature’s well-known paradox of “empowerment through surrender.” Central to many theories of recovery is the paradoxical notion that to gain control over one’s drinking or drug use, an individual needs to first admit his own powerlessness over the substance (e.g., Tiebout 1949; Kurtz 1979). Kurtz has labeled this theme as “not-God” as in “the fundamental and first message of Alcoholics Anonymous to its members is that they are not infinite, not absolute, not God” (1979, p. 3). Recovery narratives, therefore, frequently recognize (and draw strength from) the fact that the individual is fundamentally imperfect and “other than omnipotent or absolutely autonomous” (Kurtz 1979, p. 196). Whereas the addict’s or deviant’s story is often that they are “real men” (and women) and would never need to capitulate to something as weak and soft as a weepy support group, recovering persons are able to acknowledge their shared vulnerability and need for mutual aid. According to O’Reilly, the “surrender” of control involved in the recovery process signifies “less a relinquishment of ‘power’ than a clarification of personal power’s finiteness. . . . (Surrender is) a marshaling of what is available rather than a wholesale abnegation of control or initiative” (1997, pp. 23–24).

An alternative possibility for striking an optimal balance between denial and responsibility can be found in Brickman’s framework for
understanding models of coping with personal problems. A small number of recent studies suggests that Brickman’s “compensatory model” (in which individuals do not blame themselves for their problems, but hold themselves responsible for the solution to the problems) might provide an adaptive framework for addressing one’s shortcomings while avoiding deep-seated feelings of stigma and shame. Brickman et al. (1982, p. 372) quote the Reverend Jesse Jackson’s various slogans as being representative of this model of responsibility (e.g., “You are not responsible for being down, but you are responsible for getting up,” and “Both tears and sweat are wet and salty, but they render a different result. Tears will get you sympathy, but sweat will get you change”).

Maruna’s (2001) Liverpool data suggest that this compensatory model characterizes the self-narratives of successfully desisting ex-convicts. He concludes that, although it may be therapeutic for a person to locate the roots of one’s problems in the social environment (disadvantage, inequality, victimization), to desist one might need to internalize responsibility for overcoming these obstacles (see also Mischkowitz 1994).

Hanninen and Koski-Jannes identified this pattern in their study of recovering alcoholics. In what they call a “growth story” narrative of recovery, the person says that he or she has been transformed “from a victim or a puppet to a consciously acting independent subject”: “In the moral sense the growth story releases the protagonist from guilt by seeing oppressive relations as the cause of problems. The responsibility of one’s life required for staying sober is seen to emerge as part of the personal growth process” (1999, p. 1843).

Finally, Valerie Braithwaite and John Braithwaite contend that this “compensatory” story is a central resolution strategy in restorative justice conferencing:

In the all-too-common cases of children in poverty who have been physically or sexually abused, they do frequently feel that they are not responsible, that their life circumstances have condemned them to regular encounters with the criminal justice system. Criminal lawyers see moral peril in allowing the law to accept poverty as an excuse. An attraction of restorative justice is that it creates a space where it can be accepted as just for such victimized offenders to believe: “I am the real victim in this room. While I am not responsible for the abused life that led me into a life of crime on the
streets, I am responsible for getting out of it and I am also responsible for helping this victim who has been hurt by my act." [In other words] a refusal to take responsibility for being down while accepting responsibility for getting up. (V. Braithwaite and J. Braithwaite 2001, p. 319)

Which of these possible balanced narratives is best suited to promote desistance among ex-offenders is a crucial area for future neutralization research.

D. Neutralizations and Offense Type

In their original article, Sykes and Matza called for research into the relationship between neutralization techniques and various types of delinquent behavior. They wrote, “Certain techniques of neutralization would appear to be better adapted to particular deviant acts than to others” (1957, p. 670). This statement has several important implications.

First, it implies that offenders may not have a general acceptance of neutralizations that they carry around with them. Instead, neutralizations may be used in specific situations and contexts, depending on the type of crime that is being or has been committed (Mills 1940; Agnew and Peters 1986). Researchers would be well advised to “focus on beliefs regarding particular types of deviance, rather than beliefs regarding deviance in general” (Agnew 1994, p. 557). Offenders are thought to accept neutralizations only for wrongdoing that they themselves engage in and not to accept neutralizations for all types of crime. Thus, telemarketing fraudsters typically accept only neutralizations for fraud and not those for armed robbery or burglary (Shover, Coffey, and Hobbs 2003). In a comparison of three groups of offenders (robbery, murder, and theft) and a nonoffender group, Wortley (1986) found that offenders who committed a particular offense were more likely to agree with neutralizations for that offense than for offenses they did not commit.

A second implication is that neutralization theory may be better suited to explain certain types of crime than others. Some researchers argue that neutralization theory can explain participation only in minor forms of crime and delinquency and is ineffective at explaining participation in more serious criminal behavior (Mitchell and Dodder 1980, 1983). Mitchell and Dodder (1980) show that as the seriousness of crime increases, the reliance on neutralizations decreases. Delinquents are more likely to use neutralizations for minor forms of delinquency
than for predatory or aggressive delinquency. These effects exist even when controlling for delinquent peer association (Mitchell and Dodder 1983). Minor (1981) surveyed 478 college students regarding acceptance of neutralizations and their participation in a variety of illegal and deviant acts. Excuse acceptance had a significant effect on drug use, a minor effect on aggressive offending, and a negative effect on sexual misbehavior.

A final implication of Sykes and Matza’s statement is that there may be a relation between the type of deviant act and the specific neutralization techniques used to neutralize guilt from it, that is, neutralizations may be crime specific. Benson (1985) argues that the nature of the offense largely determines the technique that offenders use. It is assumed that offenders only use techniques they believe will be accepted by others, which is largely a function of the offense in question. The neutralizing audience is unlikely to accept “metaphor of the ledger” neutralizations for serious street crime, but they will accept it for certain forms of work-related indiscretions. In a survey of employees, Hollinger (1991) found that the techniques “condemnation of the condemner” and the “metaphor of the ledger” successfully predicted production deviance (i.e., long lunch breaks, slow or sloppy work, and work under the influence of drugs or alcohol) but not property theft (i.e., misuse of discount privilege, taking store merchandise, and underringing a purchase). Minor (1981) did not find any differences between killers’ and assailters’ use of two neutralization techniques, denial of responsibility and denial of the victim. Shields and Whitehall (1994) examined earlier neutralizations for a group of recidivists. They found that shoplifters and non-shoplifters were equally likely to endorse the denial of victim neutralization.

The crime specificity of neutralizations has been largely ignored by researchers. To answer adequately the question of whether the techniques are crime specific, it would be necessary to examine a group of individuals who commit multiple types of crime and deviance, then question them about their neutralizing beliefs for each deviant act. This would allow researchers to determine if individuals use different neutralizations for different types of crime.

E. Neutralizations and Offender Background

Sykes and Matza recognized that their theory of neutralizations needed further refinements. They suggested a few lines of investigation that they saw as critical, such as “there is a need for more knowledge
concerning the differential distribution of techniques of neutralization, as operative patterns of thought, by age, sex, social class, ethnic group, etc.” (Sykes and Matza 1957, p. 669). Neutralizations are learned patterns of thought that are prevalent in society and do not appear spontaneously to individuals. We would expect that different types of people may call upon different types of neutralizations based on their position in the social structure.

Because of differing positions in the social structure, one might expect that different racial and ethnic groups would use different neutralizations. Differences may appear in the number of neutralizations that are required and in the use of specific techniques. Several studies have investigated the differential use of neutralizations by race and ethnicity. Overall, the results suggest that race has little effect on the frequency of neutralization acceptance (Ball 1966; Rogers and Buffalo 1974b; Mannle and Lewis 1979). In regard to ethnicity, Mitchell and Dodder (1990) examined differences between Hispanic and white males. They administered a self-report survey to college students in Texas and Oklahoma and found only small differences in the importance of neutralizations for explaining delinquency between white and Hispanic students.

One of the few studies to examine the role of ethnic culture and choice of neutralization was conducted on juvenile car thieves in Tel Aviv, Israel. Hazani (1991a) examined the neutralizations used by two groups of Israeli boys who engaged in car theft followed by reckless driving. The two groups differed from one another in social class, ethnic origin, and culture. The two groups of boys were embedded in different social and cultural structures. Therefore, they drew upon different “symbols banks” when they accounted for their crimes. This implies that aligning vocabularies are drawn from one’s sociocultural milieu.

F. Neutralization and Cultural Differences

Vocabularies of motives “vary in content and character with historical epochs and societal structures” (Mills 1940, p. 913). Culture itself can be understood as “a set of cognitive constraints” (Stokes and Hewitt 1976, p. 837). More comparative research is needed contrasting the patterned use of neutralizations in different cultures.

Some research suggests that certain cultures are more prone to neutralizations than others. For instance, violent offenders in Northern Ireland scored significantly higher on measures of external attribution
and lower on measures of guilt when compared to violent offenders in England (Gudjonsson and Bownes 1991). Braithwaite (1989) argues persuasively that excuses and justifications might be a particularly Western phenomenon, and that cultures such as Japan are characterized by a more frequent use of apology and confession (see also Wagatsuma and Rosett 1986). Some neutralization techniques might be particularly favored in different cultural contexts.

G. Interactions with Other Variables

Neutralizations are meant only to allow for, not “cause” delinquency; an additional “push” is needed once someone has been freed from the moral binds that prevent offending (Matza 1964). More research is needed on how neutralization acceptance interacts with social and structural processes. Many theories of criminal persistence and desistance focus on the role of social factors like employment and marriage in creating turning points in the lives of people involved with crime (Sampson and Laub 1993). Sociocognitive theory would suggest that factors like employment and marriage have differential effects on criminal behavior, depending on a person’s explanatory style. In his desistance study, Maruna (forthcoming) argues that ex-offenders who attribute these sorts of positive life events to internal, stable, and global characteristics might be more likely to desist from drug use and criminal behavior after finding a job or getting married. Alternatively, those who credit such life events to luck or random chance will be less likely to respond positively to these life changes.

H. How Do Neutralizations Change?

Another crucial question is how individual accounts themselves change over the life course. How and why do individuals change their stories? From all indications, neutralization techniques seem to be difficult to sustain over time. Cressey’s research on white-collar offenders found that his subjects “at first looked upon themselves not as criminals but as borrowers,” utilizing a variety of neutralizations. Eventually, however, they recognized they were “in too deep” and were “forced to recognize that their reasoning in regard to borrowing had been ‘phony’ or that they have been ‘kidding themselves’ about repaying the money” (1953, p. 120). This same phrase—“kidding” one’s self—appeared with remarkable regularity in Maruna’s
(2001) research on ex-convicts in the United Kingdom fifty years later. Maruna argues that when the essential inconsistency between feeling that one is a good person but is doing bad things is thrust to the foreground by “disorienting episodes” (Lofland 1969), deviant identities may get deconstructed. These speculations need much further investigation, perhaps using sample sizes in the range of $N = 1$ to allow for in-depth, longitudinal analyses of complex individuals over a life course.

I. Neutralizations and the Criminal Justice System

Finally, how this deconstruction process might be used to inform offender treatment is a pressing question. The potential role for a refined neutralization theory is vast. In areas in which cognitive theory is better developed, as in the study of depression (e.g., Abramson et al. 1978), cognitive interventions routinely outperform pharmacological and other attempts at symptom reduction in controlled trials (see Schwartz and Schwartz 1993; McGuire 2000). However, although neutralization theory flags social cognition as an important target of correctional practice, the theory has little to say about the best method for changing cognitions. Cognitive therapy might not be any more likely to impact offenders’ self-narratives than, for instance, job training, victim-offender mediation, or community service work. The best target for such change remains an empirical question.

Some correctional programs employ direct confrontation techniques to deal with neutralizations. When offenders minimize or rationalize their previous offending, they are challenged instead to take responsibility for their behavior (see Fox 1999a). Other treatment programs explicitly avoid challenging minimizations as a means of gaining trust and securing cooperation (see esp. Marshall et al. 2001). Failing to honor an account offered by an individual might disrupt the trust between client and counselor. One such alternative to the direct confrontation of neutralizations is the “telling of stories” (Zehr 1990), central to restorative justice conferencing, through which all participants in an offense have the opportunity to share their interpretation of the event and collectively decide upon a way of resolving the matter (see esp. J. Braithwaite and V. Braithwaite 2001, p. 45; Sullivan and Tifft 2001, p. 46). Other alternatives include motivational interviewing (Mann and Rollnick 1996) and an approach that Jenkins
(1990) calls “an invitation to responsibility,” involving a combination of group and individual therapy.

Finally, neutralization theory has implications for criminal justice practice outside of offender treatment. Judges often make reference to verbal indications of “remorse” and “denial” in sentencing; lawyers point to such factors in pleas and in mitigation; and parole boards are often reluctant to grant parole if an offender refuses to take full responsibility for his or her past actions (Hood and Shute 1995; Horne 1999; Tidmarsh 1999). Indeed, most courts “accept without question” the long-standing assumption that before people can change, they have to admit they have a problem (Kaden 1999). This assumption can have serious consequences. The acceptance of personal responsibility for one’s crime is commonly a prerequisite for admission into treatment programs and correctional alternatives. Treatment clients who try to neutralize and minimize past offenses in treatment can suffer consequences in the form of negative reports by treatment providers to the courts and parole authorities. Individuals who are unwilling to accept full responsibility for an offense can find themselves legally terminated from treatment and punished with probation revocation or extended stays of imprisonment. These consequences can be especially severe in the case of sex offenders mandated to treatment (see Kaden 1999).

Because it is understood to be a necessary part of the recovery process, even self-incrimination is thought to be in the individual’s best interest.25 These “real world” implications make future research on offender neutralizations more than just an intriguing academic exercise.

25 In a watershed five-to-four decision (McKane v. Lile, no. 00–1187), the U.S. Supreme Court ruled that prison rehabilitation programs that require inmates to reveal even previously undisclosed crimes do not violate the constitutional right to avoid compelled self-incrimination—even in cases where inmates lose privileges for refusing to participate (see Greenhouse 2002).


*Criminology* 4:22–32.
*American Psychologist* 44:1175–84.
*Psychological Bulletin* 91:3–26.


Neutralizations


Hunter, Christopher H. 1984. “Aligning Actions: Types and Social
Aldine de Gruyter.
Peers, and Delinquency: A Test of the Self-Enhancement Thesis.” American
Jenkins, Alan. 1990. Invitations to Responsibility: The Therapeutic Engagement of
Men Who Are Violent and Abusive. Adelaide, South Australia: Dulwich Centre.
Jesilow, Paul, Henry N. Pontell, and Gilbert Geis. 1993. Prescription for Profit:
Inferences: From Acts to Dispositions.” In Advances in Experimental Social
Kaden, J. 1999. “Therapy for Convicted Sex Offenders: Pursuing Rehabil-
itation without Incrimination.” Journal of Criminal Law and Criminology
Kaplan, Howard B. 1975. “Increases in Self-Rejection as an Antecedent of
801.
Research: A New Approach to Old Problems.” In Developments in Crime
and Crime Control Research: German Studies on Victims, Offenders, and the
Chicago: University of Chicago Press.
Kelley, Thomas. 1996. “At-Risk Youth and Locus of Control: Do They Really
the Dehumanization of Victims and Victimizers.” Journal of Social Issues
Social Psychology of Authority and Responsibility. New Haven, Conn.: Yale
University Press.
Women’s Prisons: A Critical Analysis of Therapeutic Assumptions and


Neutralizations


THE UNIVERSITY OF CHICAGO PRESS

CRIME AND JUSTICE  2004-2005

OFFPRINT ORDER FORM

Please return this form with your proofs even if no extra offprints are ordered. You may also send in your order directly to Jane Seiberling, The University of Chicago Press, 1427 East 60th Street, Chicago, Illinois 60637, fax number: (773) 753-3383. Prices for offprints are listed on next page. Please note: Fifty free offprints with covers are offered for each article in Crime and Justice. This form is to order additional offprints. Note that the free offprints are per article, not per author. If you have co-authors, you may wish to order additional offprints for them.

Author(s): _____________________________________________
Essay title: _____________________________________________
Length of essay: ________ pages

Order only in multiples of 50. Please send me the following:

( ) ________ (#) additional offprints with covers
( ) No extra offprints desired

Ship to: _____________________________________________  Bill to: _____________________________________________
________________________________________  _____________________________________________
________________________________________  _____________________________________________
________________________________________  _____________________________________________
________________________________________  _____________________________________________
________________________________________  _____________________________________________

Author signature(s) _____________________________________________

Send foreign shipment via ( ) PARCEL POST INSURED or ( ) AIR MAIL.
Domestic orders are sent via UPS.

INSTITUTIONAL ORDERS: Please list as "Vendor" The University of Chicago Press, 11030 South Langley Avenue, Chicago, Illinois 60628, federal tax ID # 362177139. If purchase order is sent separately from this order form, it must be marked "Confirming." All purchase orders must include the name of the publication, author's name, and the number of offprints desired.

Official purchase order number: ________________ ( ) attached ( ) will follow
## OFFPRINT PRICES

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50 offprints:</td>
<td>$168.00</td>
<td>$178.00</td>
<td>$187.00</td>
<td>$192.00</td>
<td>$204.00</td>
</tr>
<tr>
<td>100 offprints:</td>
<td>$196.00</td>
<td>$209.00</td>
<td>$226.00</td>
<td>$242.00</td>
<td>$255.00</td>
</tr>
<tr>
<td>150 offprints:</td>
<td>$224.00</td>
<td>$243.00</td>
<td>$270.00</td>
<td>$288.00</td>
<td>$306.00</td>
</tr>
<tr>
<td>Add'l 50's:</td>
<td>$28.00</td>
<td>$34.00</td>
<td>$41.00</td>
<td>$47.00</td>
<td>$53.00</td>
</tr>
</tbody>
</table>

1 For longer essays, please combine rates, e.g. the rate for 75 pages = rates for 40 + 35

### FORMAT:
Offprints are printed exactly as essays appear in the book. They are trimmed on all sides and stapled in the upper left corner. Covers are printed on white stock and include essay title, author's name, and copyright information.

### SHIPPING:
Offprints are shipped three weeks after publication of the book. Offprint prices are F.O.B. shipping point. **Postage charges are entered as a separate item on the invoice;** prepaid orders, however, are shipped postage free. Foreign shipments will be sent via insured parcel post unless otherwise specified on the reverse side of this sheet.

### LATE ORDERS:
The prices listed above apply only to orders placed by the author(s) and returned promptly as requested. After the book is published, estimates for reprints in multiples of 100 may be obtained from Journals Reprint Coordinator, The University of Chicago Press, 1427 East 60th Street, Chicago, Illinois 60637, fax: (773) 753-3383.

### PREPAYMENT:
May be made by institutional purchase order (please see 1st page), personal check, or

( ) Visa  ( ) Mastercard  

# ____________________________

Exp. date ___________  Signature ____________________________