

Culpable Control and Causal Deviance

*[Forthcoming in Journal of Personality and Social Psychology Compass]*

Mark D. Alicke

Ohio University

David Rose

Rutgers University

## Abstract

Actions that are intended to produce harmful consequences can fail to achieve their desired effects in numerous ways. We refer to action sequences in which harmful intentions are thwarted as deviant causal chains. The culpable control model of blame (CCM) is a useful tool for predicting and explaining the attributions that observers make of the actors whose harmful intentions go awry. In this paper, we describe six types of deviant causal chains; those in which: an actor's attempt is obviated by the intervention of another person or the environment; the intended effects could not have been produced regardless of the actor's behavior; other causes diminish the actor's causal role; the actor brings about foreseen but undesired consequences as a result of pursuing his or her focal goal; the focal action produces a chain of increasingly remote causal events; and the actor derives unforeseen benefits from his or her nefarious actions. A basic assumption of the CCM in these cases is that attributions for the participants' actions will depend on positive and negative evaluations of their intentions and behaviors. We describe empirical findings that are consistent with this assumption, and predict other findings for causal deviance phenomena that have not yet been investigated empirically.

*Culpable Control and Causal Deviance*

Most intentional actions produce the consequences that we anticipate and desire. I pour the milk on my Fruit Loops, pick up my spoon, and most of the time, devour my breakfast and satisfy my hunger (and perhaps my sweet tooth). Of course, it doesn't always happen this way. I might discover that the milk is sour, that my son finished the Fruit Loops, or a maniac could break into the house and force the Fruit Loops down my throat. Life, and action sequences, can be complex. People have many desires that they never act on, they act on desires without achieving the expected consequences, and they achieve the expected consequences in unexpected ways. To complicate matters further, behaviors frequently accrue additional consequences that were neither desired nor foreseen, that were desired but not foreseen, or that were foreseen but not desired. In short, even the simplest actions can ramify into a labyrinth of effects.

Behaviors and their sometimes tortuous consequences are the grist for evaluating social conduct. As the foregoing example suggests, even mundane, innocuous events can go awry. Of course, many of our daily actions are more important than eating Fruit Loops: We discipline our children, compliment or criticize our co-workers, act in selfish or generous ways toward our friends and relationship partners, satisfy or dissatisfy our employers—and the consequences of these actions can produce various intended and unintended outcomes. Because of the complex relationships that obtain among desires, beliefs, actions and consequences, attributions for social actions are not easily explained by simple normative models. Theories of social conduct have been largely informed by the norms and prescriptions of Anglo-American jurisprudence (e.g., Fincham & Jaspars; Shaver, 1985; Shultz, Schleifer, & Altman, 1981; Tetlock, 2002). In the criminal law, *mens rea* and *actus reus* are the two overarching principles that determine whether behaviors that cause harm, or that have the potential to do so, are accessible to criminal

sanctions. *Mens rea* corresponds roughly to a guilty mind, whereas *actus reus* refers to culpable actions as defined by legal statutes and case law. Simply put, in the criminal law, people must intend harm, act on these intentions, and cause or potentially cause harm to be eligible for blame and legal responsibility.

But as the foregoing example suggests, these criteria are too general and imprecise to cover the intricate relations that affix the various elements in an action sequence. In this paper, we employ the Culpable Control Model (CCM) of blame to conceptualize how and when people attribute blame to agents who set out to achieve a desired effect and either fail to achieve it, or achieve other effects that were undesired and/or unforeseen. We refer to events in which an agent's actions either fail to achieve a desired effect or bring about undesired and/or unforeseen effects as deviant causal chains. Our analysis is confined primarily to harmful or socially undesirable intentions and actions since these are the main focus of theories of blame, responsibility, and moral judgment, or what we refer to as "conduct evaluation."

We focus in particular on six species of causal deviance. The first is the very common experience of acting on a desire but failing to achieve the desired effect. In the law, this entails the broad category of attempted and inchoate offenses. In the context of evaluating ordinary social behavior, it includes trying, but failing, to cause someone psychological pain, material loss, or a diminution in status or position.

The second form of causal deviance is one that has been investigated in counterfactual reasoning studies, namely, outcome mutability. In some instances, the outcomes that actors foresee are foreclosed in the sense that only one consequence can occur, regardless of the action. In these cases, the outcomes of an agent's behavior are immutable: nothing the agent does can alter them. The important question in this regard is when and whether people's actions are

excused when outcomes are immutable in contrast to conditions in which their actions could achieve the desired consequences.

The third type of problem is one of causal “overdetermination” in which unforeseen, contributing causes lessen the impact of the agent’s contribution to the outcome, or render it wholly superfluous. Person A might, for example, spike Person B’s Mountain Dew with a drug that is intended to put him into a coma without realizing that Person C had already spiked it with a drug sufficient to kill him. The question in this case is how much observers will mitigate Person A’s blame as a result of Person C’s superseding causal influence.

The fourth type of causal deviance involves undesired but foreseen offshoots (or “side-effects”) of a desired action. Decision-makers frequently confront the possibility, and sometimes the certainty, that the actions they pursue will produce undesired effects. In military operations, for example, the phrase “collateral damage” refers to undesired but foreseen civilian deaths that result from military operations. In more mundane social affairs, people’s focal goals also bring about undesired peripheral effects. Buying a large SUV to transport a family and provide protection in bad weather, for example, has the foreseeable but undesired side-effect of harming the environment. Many recent studies by both psychologists and experimental philosophers on the “Knobe effect” (Knobe, 2003) pertain exactly to this sort of problem. The main question in these studies concerns the degree to which observers view as intentional the peripheral, undesired side-effects that are produced as the result of pursuing and achieving a focal goal.

The fifth source of causal deviance that we address involves unforeseen and extended outcomes of an intentional action. Suppose, for example, that Sally decides to tell Martha that she saw her boyfriend kissing his old girlfriend in a bar. This has the desired effect of informing Martha that her boyfriend is a jerk. However, it also causes Martha to become depressed, which

causes her to neglect her schoolwork, which causes her to be put on academic probation, which causes her to take drugs, which makes her forget to feed her goldfish, and so on. The question we consider in this case concerns the extent to which observers link extended consequences to an initial action and use them as a basis for ascribing intent, causation, and blame.

The sixth and final source of causal deviance that we consider involves an unusual type of behavioral constraint, namely, being forced to do something that reaps unexpected benefits. How do observers judge, for example, a person who is coerced into an illegal action from which she receives substantial reward? Or, to take an example from ordinary social life, suppose that Ivan blackmails his friend Boris into escorting his unpopular sister, Natasha, to the high school prom, perhaps by threatening to make public very embarrassing information about Boris. As it turns out, Boris and Natasha fall in love immediately and live happily ever after. The question of interest in this case concerns perceptions of constraint and its influence on judgments of blame and responsibility. That is, to what extent will observers perceive Boris's dating decision to have been constrained by Ivan in comparison to a condition in which the decision has negative consequences (i.e., in which he has an agonizingly dreadful date with Natasha) and how much, if at all, will these perceptions of constraint influence judgments of blame and responsibility?

### *The Culpable Control Model*

We consider each type of causal deviance from the perspective of the Culpable Control Model (CCM) of blame. The CCM assumes that when observers assess behavioral events, they attempt to ascertain how much control an actor (or actors) exerted over a harmful or potentially harmful outcome. The three main aspects of control that are considered are behavior control (did the actor behave purposively and knowingly?), causal control (to what extent did the actor uniquely cause the outcomes that occurred?), and outcome control (did the actor foresee the

outcomes and the process by which they eventuated?). However, while observers consider the control evidence, they simultaneously make evaluative judgments (i.e., assessments of "goodness-badness") of the actor's character, motives, actions, or the consequences that ensue (Alicke, 1992, Alicke, 2000, Alicke, 2008, Alicke et al., 2008; Alicke, Rose and Bloom, in press). Thus, when people estimate how much control an actor exerted over the event's outcomes, they automatically conflate these assessments with their evaluations, especially when their evaluations are extreme. Spontaneous, negative evaluations induce observers to process information about an event in a "blame-validation" mode. Blame validation is akin to a confirmatory test strategy (Wason, 1960). In essence, the desire to blame actors who arouse observers' disapprobation leads them to construe the available evidence in a way that supports the attribution they wish to make. Instead of personal control assessments providing an independent basis for ascribing blame, the desire to blame the actor partly determines personal control judgments. The phrase "culpable control" reflects the fact that blame is elided into the criteria by which it is supposed to be independently evaluated.

The relationship between spontaneous evaluations and control evidence is presumed to be compensatory. When the control evidence is unambiguous, and observers are fairly certain of the actor's causal and intentional involvement in the outcome, negative reactions to the actor, his or her actions, or the outcomes that occur, are unlikely to wield much influence over blame attributions. Spontaneous evaluations or reactions exert their strongest effect when the control evidence is ambiguous. In this respect, such evaluations may influence blame indirectly by altering perceptions of the control evidence, or directly, in which case control estimations are adjusted post-hoc to validate the attribution that has already been made. We do not distinguish between these two possibilities, however, in the following examples.

We use the phrase “causal deviance” in a deviant way. Pizarro, Uhlmann, and Bloom (2003) use causal deviance to refer to cases (adapted from Chisholm, 1966) in which an intended outcome is effectuated in an unanticipated manner. We employ causal deviance more generally to refer to causal sequences that deviate from the actor’s expectations and/or desires. In the terminology of the CCM, this may indicate a diminution of causal control, in which the actor’s causal influence on the outcome is diminished or negated, a reduction in outcome control, in which the outcome that occurs was unforeseen by the actor or occurs via an unanticipated causal process, or both.

By emphasizing the role of evaluations in estimations of control and blame, the CCM differs from traditional theories of blame and responsibility that have posited a forward connection between intent and/or causation and blame (e.g., Darley & Shultz, 1990; Shaver, 1985; Shultz & Schleifer, 1983). In essence, the CCM argues that the desire to blame an actor based on disapproval of his or her motives and values, actions, or the outcomes that eventuate, leads observers to work backward to alter their perceptions of behavioral, causal or outcome control.

The predictions from the CCM with respect to the different facets of causal deviance are straightforward. To the extent that behaviors are perceived to have issued from bad as opposed to good intentions or motives: a) attempts that miss their mark should be perceived to have come closer to achieving the desired effect; b) people should be blamed more when unfortunate consequences of their actions were alterable versus unalterable; c) people should be blamed more when more potent causal factors override or supplant their own causal contribution to a harmful outcome; d) undesired but foreseen side-effects of people’s actions should be seen as more intentional and thus more blameworthy; e) people should be blamed more for causally remote



consequences of their actions; and f) people should also be blamed more when they receive unexpected benefits from their coerced actions.

*Causal Deviance 1: Thwarted Attempts*

With all other facets of an event being equal, people who intend to cause a harmful outcome tend to be blamed more when the outcome is achieved than when the same action inadvertently fails to accomplish its purpose (Cushman, 2008). Legal criteria for failed attempts are often indeterminate. The most important criterion is that the actor comes close to achieving his or her goals, but this criterion fails to specify just *how* close one has to come to be held responsible. Suppose, for example, that a spouse who suspects his wife of infidelity decides to shoot her presumed lover. He goes to the man's house, peers into his bedroom, sticks his gun in the open window, and fires at the object under the covers. Unfortunately for the jealous husband, the object that he shoots is a large pillow. As it turns out, his rival was at that moment in the next room composing a romantic e-mail to the jealous husband's wife. Is this attempted murder?

Because he was so far from accomplishing his nefarious goal, it is unlikely that the jealous husband would be prosecuted for attempted murder. What he did accomplish—demolishing a pillow—is not even a misdemeanor. And yet, the law in practice is sharply at odds with the ideals of jurisprudence and ordinary common sense. The gun-toting, hapless husband is a lawless and dangerous character, just as dangerous as one who has the better fortune (at least from his vantage) of eliminating a rival rather than a pillow.

In predicting how ordinary observers will view this case, the CCM focuses on their evaluations of the participants. In this example, the rival is enamored of the jealous husband's wife, which may be bad, but which is nonetheless trumped in malevolence by the husband's

murderous intentions. Contrast this with a case in which the husband was far more justified in his actions, perhaps one in which the person he was pursuing had threatened or previously harmed his family. In assessing whether the husband came close enough to be tried for attempted murder, the CCM predicts that the negative evaluations of the jealous husband's motives would be elided into the focal judgment of closeness. This would not occur for the husband whose actions were more justified and who would, therefore, be less negatively evaluated. In CCM terms, these inflated judgments of closeness can be viewed as exaggerated estimations of potential causal control (potential because the intended outcome did not occur). Accordingly, the jealous husband would be seen to have come closer in his attempt, and would be blamed more for his actions, than he would have if his motives or actions had been less negatively evaluated.

The same type of reasoning applies to inchoate offenses such as conspiracy. Conspiracy is a unique and controversial legal category because it enables criminal prosecution with no *actus reus*. That is, in charges of conspiracy, the mere concoction of a plan suffices for criminal sanctions apart from the activities that the plan specifies. To our knowledge, such cases have not been applied to ordinary social conflicts, although the possibilities are legion. Suppose, for example, that a group of students conspire to give another, named Brad—whom they dislike—erroneous information so that he studies the wrong chapters and fails a test. In one case, the students are jealous of Brad because he studies harder and receives the best grades in the class. In the other case, Brad frequently gets good grades by cheating. As it turns out, Brad finds out about the students' plan in time and reads the correct chapters. The question concerns the extent to which the students will be considered to be guilty of conspiring against Brad and blamed for their actions. Conspiracy is a complex control element, including aspects of behavior control

(did they know exactly what they were doing?), causal control (did their plans bring about the harmful outcome?), and outcome control (did they anticipate the manner in which the outcome eventuated?). The culpable control model predicts that conspiracy will be seen as a less apt description of the students' actions when the person they plot against is described negatively versus positively.

*Causal Deviance 2: Outcome Inevitability*

The impact of an intentional action is potentially lessened when other causes contribute to an effect. This assumption has a long history in psychology, particularly in theories of discounting (Kelley, 1972; McClure, 1992). The “discounting principle” in attribution theories simply states that the force of any single cause is diminished when other factors contribute to an event's outcome. While this assumption has been corroborated in many different contexts, it fails to account for the multitudinous ways in which competing causal conditions can alter an actor's impact on the ultimate outcome.

One causal condition that has been thoroughly investigated in counterfactual reasoning studies is the mutability of an outcome. In an early study by Miller and McFarland (1986, Study 1), for example, participants learned that a man was shot and injured while visiting his usual convenience store or a different one. The primary response measure was the amount of money participants awarded to the victim. Results showed that more compensation was recommended when the man was shot in the unusual than in the usual store, presumably because it is easier to mutate or undo the harmful event when the victim departs from his normal routine. This would be represented in counterfactual ruminations such as: “If only the actor had gone to his usual store, he wouldn't have been shot.” Counterfactual thoughts presumably lead observers to sympathize with the victim and to award him more compensation.

Counterfactual reasoning theories, however, do not unambiguously predict Miller and McFarland's results. Counterfactual theories stipulate that observers will respond emotionally to unfortunate events and that they will seek to explain such events with reference to their mutable features. But counterfactual theories do not explain fully the link between mutability and blame. In particular, the assumption that the unusual store scenario will lead observers to sympathize more with the actor's plight is not a unique derivation. One could as easily assume that observers would blame the actor more for his willful decision to go to the unusual store and therefore award him less compensation. In this case, observers might reason counterfactually: "If only the victim had not made the stupid decision to go to the unusual store, he wouldn't have been harmed."

The main limitation, therefore, of counterfactual reasoning theories as applied to conduct evaluation is that they fail to supply a bridge from the affective responses that are heightened by the knowledge that a harmful outcome was avoidable to judgments of causation and blame. That is, counterfactual reasoning theories predict people's affective responses to events but not the judgmental consequences. By contrast, the CCM predicts that negative affective reactions in response to an agent's judgments and actions will lead observers to emphasize the agent's causal and/or intentional influence. As a result, judgments of causation (i.e., causal control in the CCM) and blame will be greater when the agent elicits negative as opposed to positive reactions.

In a representative study, we presented participants with a story in which a man was traveling to the hospital to see his dying mother (Alicke et al., 2008, Study 3). In one condition, he stopped along the way to help the victims of a car accident, whereas in the other, he stopped to sell a friend some drugs. Each detour detained him for the same amount of time. In one version of the story, the man got to the hospital just after his mother died, whereas in the other,

his mother died just as he was setting out in his car, so he would not have been in time regardless of the detour. As predicted, when the actor's delay was characterized positively (i.e., he stopped to help accident victims), the mutability of the outcome had no effect on ratings of his causal role or blameworthiness. On the other hand, the actor whose delay was characterized negatively (i.e., he stopped to make a drug deal) was blamed more, and seen as more causal, when he would otherwise have been on time to visit his dying mother. Findings such as these suggest that judgments about immutable outcomes depend, in part, on evaluations. When actors behave favorably or responsibly, the fact that they could have averted harmful consequences by selecting alternative behavioral routes has no effect on attributions of causation or blame. Conversely, actors who behave unfavorably or irresponsibly are seen as more causal, and blamed more, when the unfortunate consequences could have been avoided.

*Scenario 3: Causal Overdetermination*

Miller and McFarland's scenarios are perhaps the prototype of counterfactual reasoning studies, involving a situation in which a minute change in events would produce a monumentally different outcome. Causal deviance in such situations is represented by a counternormative event that leads to an unintended (and usually unwanted) outcome. A different causal deviance problem arises when an individual's causal contribution is negated by a preexisting or subsequent intervention. This situation represents a type of "causal overdetermination" in which another causal factor diminishes or eradicates an actor's otherwise causally impactful contribution.

To study this problem from the vantage of the CCM, we wanted first to establish a clear basis for positive or negative evaluations of the main character in the story. The story that we created (Alicke, Rose, & Bloom, Study 3, in press) was one in which the victim of a shooting,

Edward Poole, was described as either a dangerous ex-convict who broke into a home (negative characterization), or a physician who entered a home at a neighbor's request to feed her cat (positive characterization). In both instances, Poole was shot by the woman's husband, Turnbull, who came home unexpectedly and encountered Poole walking up the stairs. A basic CCM prediction is that Turnbull would be blamed more, and seen as more of the cause of the victim's death, when the victim was characterized positively as opposed to negatively.

We assumed, however, that these positive versus negative characterization effects would be cancelled when another factor superseded Turnbull's causal involvement. As noted previously, the CCM assumes a compensatory relationship between spontaneous evaluations and control estimations. In this instance, we assume that when Turnbull's causal role is clearly negated, his positive or negative character would have no effect on causation or blame judgments. To test this, we created a causal overdetermination condition in which an intervening circumstance negated the actor's causal influence. In this condition, an autopsy revealed that Poole suffered a brain aneurysm at virtually the same moment that he was shot by Turnbull. Under these circumstances, we assumed that even those who have strong reactions to an innocent victim's death will be reluctant to ascribe more causal influence to Turnbull than when a perpetrator was killed because his behavior was unnecessary to produce Poole's death in the immediate situation. Accordingly, we expected to obtain uniformly low causation and blame ratings in this condition regardless of whether Poole was characterized negatively or positively.

We included two other contrast conditions. One version of the story made no mention of competing causal influences (control condition), while in the other version, participants were told that the autopsy indicated that Poole was seriously ill and would have died from a brain tumor within a few weeks. In comparison to the previous condition, we assumed that the constraint in

this circumstance—that Poole would have died in a few weeks from a brain tumor—would allow for positive and negative evaluation effects. Thus, higher causal and blame ratings were predicted in the condition in which Poole was characterized positively than in the condition where he was characterized negatively. In sum, we wanted to show that only severe constraints (i.e., causal overdetermination), and not moderate ones (i.e., the victim would have died in the near future), mitigate effects of positive and negative characterizations on causal assignment and blame. Results supported this assumption in the form of the anticipated interaction: Positive versus negative characterization effects were obtained in the control condition and when the victim had an illness that would have killed him in the near future, but they were negated in the causal overdetermination condition in which his death was imminent regardless of the homeowner's actions.

*Causal Deviance 4: Undesired Side Effects*

Another causal deviance problem derives from the fact that virtually all behavioral decisions have both focal and peripheral consequences. Focal consequences pertain to actors' primary goals, whereas peripheral consequences are foreseeable side-effects that actors are willing to tolerate to achieve their primary purposes. This was a foundational observation in Jones and Davis's (1965) theory of non-common effects, which emphasized that intent attributions are based on the effects that uniquely distinguish decision alternatives. For example, if a student were deciding whether to go to the University of Chicago or Stanford for graduate school, the effects of being expensive, prestigious, and high quality are common to both and are therefore unlikely to be cited as the reason for the student's decision. If the student picked Chicago, and the main source of differentiation was that it was close to home, observers would infer that the student's focal goal was the desire to remain near her family.

The theory of non-common effects, however, had limited applicability to the general issue of peripheral consequences because it dealt with a very specific situation; that is, one in which there were two clearly specified alternatives, each with consequences that were known in advance. Everyday decisions are a bit messier. People choose from among various behavioral alternatives, each of which may have many side-effects that vary in their foreseeability.

A particular manifestation of the side-effect problem was introduced by Joshua Knobe and has stimulated a large number of studies by both psychologists and a burgeoning group of experimental philosophers. Knobe's oft-cited situation (Knobe, 2003) is one in which the CEO of a company is starting a program which will increase company profits (focal goal) but which is known to have either helpful or harmful effects on the environment (side-effect). The CEO doesn't care about these side-effects; his only goal is to make money for the company. As anticipated, the environment is either helped or harmed. The finding that has stimulated researchers' interest is that the CEO is seen to have harmed the environment intentionally far more than he is seen to have helped the environment intentionally.

The CCM offers a fairly straightforward explanation of these findings (Alicke, 2007). The CEO who doesn't care about helping the environment is no hero, but he is no villain either, and because his actions help the environment, evaluations of his actions should be somewhere in the neutral range, resulting in little effect on intentionality ratings. On the other hand, the CEO who doesn't care about harming the environment, and who winds up doing so, is a deplorable character, and this negative evaluation is elided into intentionality judgments.

The CEO's character is only one of many possible bases for evaluative reactions in the Knobe paradigm. A second prominent source of evaluation is whether the side effects actually turned out to be beneficial or harmful. As research on outcome bias has shown (Alicke & Davis,



1989; Alicke, Davis & Pezzo, 1994; Alicke & Mazzocco, 2004; Baron & Hershey, 1988), people whose actions bring about negative outcomes are blamed more than those who effectuate positive ones, even with all other facets of an event equated. According to the CCM, this effect is due, at least in part, to the difference in evaluative reactions that occur in response to positive versus negative outcomes. Thus, the CCM predicts a general effect of outcome such that the CEO who brings about negative side effects will be blamed more than one whose actions lead to positive side effects.

To test these assumptions, we constructed a scenario in which a company was creating a new learning system to promote academic development in elementary schools (Alicke, Rose & Bloom, 2012). The CEO's motivation was either to enhance the students' learning experience (altruistic motive) or to increase profits (profit motive). A five-year follow-up study showed that the side effects were positive (i.e., the program fostered abstract learning) or negative (the program hindered abstract learning). The primary response measures of concern in this study were ratings of blame-praise and the degree to which participants thought that the CEO intended for the program to be helpful or harmful to the students' ability for abstract learning. In the terminology of the CCM, this aspect of intention (intention of outcome) is defined as outcome control.

Consistent with CCM reasoning, the CEO was deemed more blameworthy when his motive was to increase profits than to promote learning, and when the side effects were negative rather than positive. And, as the CCM predicts, when the outcome of the side effect was positive, the CEO with the education motive was seen to have intended benefitting the students to a greater extent than the CEO with the profit motive. Conversely, when the outcome of the side effect was negative, the CEO with the profit motive was seen to have intended to harm the

students to a greater extent than the CEO with the education motive. We believe that this study provides initial support for the idea that the Knobe effect can be explained with reference to the numerous aspects of the two scenarios (helping versus harming the environment) that produce negative versus positive evaluative reactions.

*Causal Deviance 5: The “Accordion” Effect*

Actions have infinite consequences, only some of which are connected to a focal intention. For example, suppose that a homeowner burns leaves in his yard without anticipating the strong wind that suddenly arises and burns down his house. This conflagration causes the house next store to burn down, and then the next, and then eventually, all of Chicago. For which of these consequences is he liable?

The guiding principle for determining the scope of liability in tort law is causal proximity (along with a measure of practicality). In turn, the guideline for deciding proximity is reasonable foreseeability. Determinations of foreseeability, however, can be quite ambiguous. In the previous example, there is no clear answer as to whether the homeowner should have foreseen the wind, the burning of his house, the subsequent burning of the next house, and so on.

One of the basic assumptions of the CCM is that blame validation processing is most likely to occur when the objective evidence is ambiguous. Ambiguous causal chains pervade ordinary social life. In fact, almost all important life decisions have consequences that extend far into the future, some of which are highly foreseeable and others that are completely surprising. A basic prediction of the CCM in this regard is that censorious as opposed to laudatory actions or motives will lead observers to ascribe more causal control to distant events in a causal chain. The reason for this is that negative reactions to the actor’s initial behavior or anti-social motives

either lead observers to perceive greater causal influence in subsequent events, or simply to ascribe greater causal influence in order to justify the blame attributions they favor.

In one supportive study (Alicke, 1992, Study 3) participants were provided with stories in which an actor, whose initial actions were culpable or inculpable, set into motion an extended causal chain. In one story, for example, John and Melissa had been dating for nine months when John found out that he had a sexually-transmitted disease. In the culpable version, John caught the disease from a woman he picked up in a bar while Melissa was out of town, whereas in the inculpable condition, he contacted it from his old girlfriend and told Melissa about it, after which she decided to continue to have safe sex with him. Soon after this, Melissa became ill with severe symptoms, which caused her to need extended medical treatment, which caused rumors in her company, which caused her to lose her job, after which she had trouble finding another job, which forced her to accept a job with lower pay. As the CCM predicts, and as the results confirmed, John was perceived to be more the cause of remote events in this chain (such as Melissa getting a job with lower pay) when his initial action was culpable versus inculpable.

### *Causal Deviance 6*

Behavior can be constrained by both personal incapacities (e.g., blindness, paralysis, mental illness, fear) and environmental events and circumstances (e.g., third-party intervention, bad weather, physical compulsion). In the law, harmful behavior is potentially excused or mitigated by a perplexing array of constraints as represented by legal defenses such as diminished capacity, insanity, duress and entrapment.

One of the difficult aspects of determining the legitimacy of such defenses is that the behaviors they entail are typically intentional. For example, both a vengeful man who is provoked into a violent act and an addict who takes drugs usually act with desire for and

foreknowledge of the various consequences. Decisions about whether to excuse people for their actions and its consequences, therefore, are not based on the absence of intent, but rather, on perceptions of the degree to which their actions were constrained. Again, however, constraint decisions can be highly ambiguous. Did the provoked person have to commit a violent act; could the sex addict have refrained from multiple affairs; and was the kleptomaniac really compelled to steal an object from a jewelry store?

The original formulation of the CCM devoted at least as much attention to diminutions of control as to factors that establish or augment it. The CCM model assumes that decisions about constraints are strongly influenced by observers' evaluations of the offensive behavior and its outcomes. One factor that may influence such evaluations is whether actors are perceived to have reaped benefits or costs from their actions, or more technically, whether they are spurred by positive or negative reinforcement contingencies. Many legal claims of the diminished capacity variety may be vanquished or vitiated by the rewards the actor reaps: The kleptomaniac acquires valuable objects, the drug addict gets high, and the sex addict has multiple reasons to rejoice.

An important aspect of observers' behavioral evaluations in these circumstances is their perception of the relationship between the actor's first- and second-order desires (Frankfurt, 1986). A first-order desire refers to the actor's immediate goal in the situation, whereas a second-order desire refers to whether the actor wants to be motivated by that desire. For example, the first- and second-order desires of a contented drug addict are perfectly aligned because the addict likes being driven by the need for the drug. On the other hand, those who are tortured by their compulsions and addictions have discordant first- and second-order desires. The CCM model predicts that the perceived legitimacy of constraint-based excuses will depend

on whether the actor seems to be deriving benefits (concordant first- and second-order desires) or misery (discordant first- and second-order desires) from his or her actions.

Previous research by Pizzaro, Uhlmann, & Salovey (2003) has shown that second-order desires can negate the moral benefits that accrue from positive actions or cancel the moral demerits that stem from negative ones. For example, participants in one study (Study 2) read about an actor who committed a positive or negative act deliberately or impulsively. In general, actors ascribed as much praise for impulsive positive acts as for deliberative ones, but ascribed less blame for negative, impulsive acts. When participants learned that the actor's second-order desire was to disavow his positive impulse (i.e., he wishes he hadn't been driven to do a good deed), his praiseworthiness was mitigated. When participants learned that the actor had a second-order desire to avoid his harmful, impulsive action (i.e., he wishes he hadn't been driven to do a bad deed), his blameworthiness was mitigated. The CCM assumes that second-order desires influence evaluations of the actor's behavior: The actor who wishes he hadn't behaved positively is evaluated less positively and, therefore, praised less; whereas the actor who wishes he hadn't behaved negatively is evaluated less negatively and blamed commensurately less.

A particularly interesting form of the second-order desire problem involves instances in which a person is forced to do something that he or she would have preferred to do anyway. Woolfolk, Doris, & Darley (2006) have studied this type of problem in a series of studies based on Frankfurt's (1988) notion of identifying with or embracing a behavioral outcome. High identification with an outcome indicates that the actor would be happy to see that outcome occur. Woolfolk, Doris and Darley were interested in whether high identification with a negative outcome would increase perceived responsibility for that outcome even if the actor behaved under strong constraint (e.g., a gun pointed at the actor's head) or virtually complete constraint

(e.g., a mind control drug). In one version, Actor A learned that Actor B had been having an affair with his wife and wished to kill him for it (high identification), whereas in the other, Actor A decided that he did not wish to stand in the way of his wife and her lover (low identification). In both cases, Actor A is forced at gunpoint to shoot Actor B. Results showed that Actor A was seen as more responsible for Actor B's death when he desired it (high identification) than when he didn't (low identification), even under high constraint. Analogous findings were obtained even when Actor A was completely constrained by a mind-controlling drug. As the authors note, these findings are quite compatible with the CCM. The CCM assumes that strong negative evaluations of the murderous husband override control estimations, leading to heightened responsibility ascriptions.

### *Summary*

We considered six ways in which events can diverge from an actor's intentional vantage, and viewed the types of attributions that observers make about such events from the perspective of the culpable control model of blame. Although the Anglo-American legal system and writings in jurisprudence provide some guidelines for judging problems such as failed attempts, immutability, overdetermination, causal chains, and behavioral constraints, these issues are among the most perplexing that legal decision-makers confront. We also examined analogues of these issues in ordinary social life, where they also pose fascinating judgment puzzles for lay observers.

It is in circumstances such as these, in which rational judgment guidelines are absent or ambiguous, that the CCM is especially valuable. The CCM accords a prominent role to observers' evaluations of the actor's intentions, actions, and to the outcomes of the actor's behavior. While evaluation rarely provides a complete explanation of causal deviance

phenomena, it often provides the link to understanding otherwise enigmatic attributional tendencies.

## References

- Alicke, M.D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63, 368-378.
- Alicke, M.D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126, 556-574.
- Alicke, M.D. (2008). Blaming badly. *Journal of Cognition and Culture*, 8, Special Issue—*On Folk Conceptions of Mind, Agency, and Morality*, 179-186.
- Alicke, M.D. & Davis, T.L. (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, 25, 362-377.
- Alicke, M.D., Davis, T.L., Buckingham, J.T., & Zell, E. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, 34, 1371-1381.
- Alicke, M.D., & Rose, D. (2010). Culpable control or moral concepts? *Behavioral and Brain Sciences*, 33, 330-331.
- Alicke, M.D., Rose, D., & Bloom, D. (in press). Causation, norm violation, and culpable control. *Journal of Philosophy*.
- Alicke, M.D. & Davis, T.L. (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, 25, 362-377.
- Bargh, J.A. and Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychology* 54, 462-479.



- Baron, J., & Hershey, J.C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569-579.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analysis in moral judgment. *Cognition*, *108*, 353-380.
- Darley, J.M., & Shultz, T.R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525-556.
- Fincham, F.D., & Jaspars, J.M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 81-138).
- Fincham, F.D., & Shultz, T.R. (1981). Intervening causation and the mitigation of responsibility for harm. *British Journal of Social and Clinical Psychology*, *20*, 113-120.
- Frankfurt, H.G. (1986). Freedom of the will and the concept of a person. In J.M. Fischer, (Ed.), *Moral responsibility* (pp. 65-80). Ithaca, Cornell Press.
- Jones, E.E., & Davis, K.E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220-266). New York: Academic Press.
- Karlovac, M., & Darley, J.M. (1988). Attribution of responsibility for accidents: A negligence law analogy. *Social Cognition*, *6*, 287-318.
- Kelley, H.H. Causal schemata and the attribution process. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.
- Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. *Analysis*, *63*, 190-193.

- Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis*, 64, 181-187.
- Lepper, M.R., & Greene, D. (Eds.) (1979). *The hidden costs of reward*. Hillsdale, N.J.: Erlbaum.
- Mazzocco, P., J., & Alicke, M.D. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, 26, 131-146.
- McCann, H. 2005: Intentional action and intending: recent empirical studies. *Philosophical Psychology*, 18, 737-748.
- 7-20.
- McClure, J.M.
- Pizarro, D.A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39, 653-660.
- Pizarro, D.A., Uhlmann, E., & Bloom, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, 14, 267-272.
- Schlenker, B.R., Britt, T.W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review*, 101, 632-652.
- Shaver, K. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.
- Shultz, T.R. Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science*, 13, 238-253.
- Tetlock, P.E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109, 451-471.
- Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wells, G.L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, *56*, 161-169.

Woolfolk, R.L., Doris, J.M., & Darley, J.M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, *100*, 283-301.