

4-1-1998

Testing Competing Theories of Justification

Paul H. Robinson

John M. Darley

Follow this and additional works at: <http://scholarship.law.unc.edu/nclr>Part of the [Law Commons](#)

Recommended Citation

Paul H. Robinson & John M. Darley, *Testing Competing Theories of Justification*, 76 N.C. L. REV. 1095 (1998).Available at: <http://scholarship.law.unc.edu/nclr/vol76/iss4/2>

This Article is brought to you for free and open access by Carolina Law Scholarship Repository. It has been accepted for inclusion in North Carolina Law Review by an authorized administrator of Carolina Law Scholarship Repository. For more information, please contact law_repository@unc.edu.

TESTING COMPETING THEORIES OF JUSTIFICATION

PAUL H. ROBINSON & JOHN M. DARLEY*

Nearly every jurisdiction's criminal law recognizes justification defenses, defenses that exculpate a person whose conduct would otherwise be criminal because the conduct is accepted or encouraged due to special circumstance. While justification defenses themselves are nearly universal, there is much disagreement over whether the defense is given because a person's act avoids a greater harm, the so-called deeds theory, or because she acts for the right reason, the so-called reasons theory. At least part of this debate focuses upon which theory best reflects community intuitions. In this Article, Professors Robinson and Darley report the results of an empirical study measuring community intuitions regarding justification defenses. They conclude that the deeds theory of justification better accords with community views than the reasons theory. The study's results suggest, and the authors discuss, a reformulation of many aspects of offense definitions, reforms to justification defenses, mitigations for mistake as to a justification, and reforms of jury acquittal verdicts. Such conclusions illustrate the potential usefulness of social science research for resolving issues disputed among criminal law theorists, as well as for providing valuable information to the drafters of criminal codes. Finally, the authors argue that reforms arising from such empirical studies increase the law's moral credibility, which in turn increases its long-term effectiveness in crime control.

I. COMPETING THEORIES OF JUSTIFICATION	1098
A. <i>The Unknowingly Justified Actor</i>	1101
B. <i>Mistake as to a Justification</i>	1101
C. <i>The Importance of Community Views in the Formulation of Criminal Law</i>	1105
II. TOOLS FOR TESTING CRIMINAL LAW THEORIES: SOCIAL PSYCHOLOGY SCENARIO RESEARCH	1108

* Paul H. Robinson is Professor of Law at Northwestern University; John M. Darley is Warren Professor of Psychology at Princeton University.

A. <i>Scenarios and Measures</i>	1108
B. <i>The Sample of Subjects</i>	1114
III. LIABILITY PREDICTIONS.....	1115
A. <i>The Contrast Cases</i>	1116
Scenario 1. Intentional (Unjustified) Burning	1116
Scenario 2. Attempted (Unjustified) Burning.....	1116
Scenario 3. Created Risk of (Unjustified) Burning, Realized—Reckless Commission	1116
Scenario 4. Created Risk of (Unjustified) Burning, Unrealized—Endangerment	1117
Scenario 5. Attempted Risk Creation—Attempted Endangerment.....	1118
Scenario 6. Intentional Justified Burning	1118
B. <i>The Test Cases</i>	1119
Scenario 7. Unknowingly Justified Burning	1119
Scenario 8. Knowingly Justified Burning but with Bad Motive	1119
Scenario 9. Mistake as to Justification, Reasonable.....	1121
Scenario 10. Mistake as to Justification, Negligent	1122
Scenario 11. Mistake as to Justification, Reckless.....	1122
IV. LIABILITY RESULTS.....	1124
A. <i>The Contrast Cases</i>	1124
B. <i>The Cases for Which the Theories of Justification Have Different Predictions</i>	1127
V. IMPLICATIONS FOR CRIMINAL LAW REFORM.....	1129
A. <i>The Formulation and Grading of Offenses</i>	1129
B. <i>The Formulation of Defenses</i>	1131
C. <i>Reform of Acquittal Verdict Forms</i>	1133
VI. CONCLUSIONS.....	1135
APPENDIX: STIMULUS STORIES.....	1137

Justification defenses, such as self-defense and law enforcement authority, are common in every jurisdiction. They share the characteristic of exculpating a person whose conduct otherwise would constitute a criminal offense, because the conduct is accepted or encouraged given the presence of special justifying circumstances.¹ For example, a police officer's conduct in making an arrest may satisfy the requirements of assault, but she is free from liability if that conduct also satisfies the requirements of the law enforcement

1. See generally PAUL H. ROBINSON, CRIMINAL LAW 401-69 (1997) (discussing justification defenses in detail).

justification for the use of force. A lost camper who takes food from another's cabin may have committed an act that satisfies the elements of the crime of theft, but he is exculpated under a lesser evils justification defense if the taking is necessary to prevent his starving to death.

Justification defenses are distinguishable from excuse defenses in a fundamental way. Both exculpate, but for different reasons. An actor pleading justification claims to have acted properly, that she did the right thing. An actor pleading excuse, such as insanity, duress, or involuntary conduct, admits that what she did was wrong, but claims that some characteristic or her condition leaves her blameless for the offense.

Despite the universal recognition of justification defenses, there is disagreement over the underlying theory of the justificatory principle, and thus the proper legal formulation of such defenses. At the core of the debate about the principle is the following question: Are justification defenses given because the actor's deed avoids a greater harm, or because she acted for the right reason?

The *deeds theory* of justification justifies conduct that avoids a greater harm, and thus it is conduct that we would be happy to tolerate under similar circumstances in the future. Justified conduct, under this theory, occurs when the actor has done the right deed, hence, the "deeds" theory of justification. The *reasons theory* looks not to the deed but to the reason for the deed. The reasons theory, then, gives a defense when a person acts for the right reason, generally *trying* to avoid a greater harm. The issue between the two theories concerns the focus of justification. Is the focus of justification the nature of the *deed*, or the actor's *reason* for acting?

The debate to date relies in large part upon legal and philosophical arguments.² But frequently a third source of authority is brought into play. Each side buttresses its arguments with claims that its theory better tracks community intuitions,³ a common claim in criminal law arguments. In this Article we test those claims about community intuitions, using policy-capturing social science research techniques designed for such inquiries. In the process, we learn about community views on the proper theory and formulation of justification defenses, as well as other criminal law doctrines, and

2. See, e.g., Paul H. Robinson, *Competing Theories of Justification: Deeds v. Reasons*, in HARM AND CULPABILITY 45, 47-49 (A.P. Simester & A.T.H. Smith eds., 1996) (citing various authorities).

3. See, e.g., GEORGE P. FLETCHER, *RETHINKING CRIMINAL LAW* 556-57 (1978); 2 PAUL H. ROBINSON, *CRIMINAL LAW DEFENSES* § 122(e), at 23-27 (1984).

about the value of social science research to criminal law formulation.

I. COMPETING THEORIES OF JUSTIFICATION

In many cases, an actor's conduct will be both objectively and subjectively justified. The actor believes his conduct avoids a greater harm and he acts to avoid that greater harm, thereby satisfying the reasons theory, and his conduct does in fact avoid a greater harm, thereby satisfying the deeds theory. But in many other cases, the deeds and reasons theories clash and give different results, and it is these cases of conflict that are the focus of this study.

Where the reasons theory of justification (sometimes called the "subjective" theory) is adopted, the standard justification formulation provides that "an actor is justified if he *believes* that the conduct is necessary"⁴ to defend against unlawful aggression, to make an arrest, to maintain order on the vehicle, and so on.⁵ Under this reasons theory, a person will get a justification defense as long as she believes that the justifying circumstances exist. Whether they actually exist or not is irrelevant.

Under a deeds theory (sometimes termed an "objective" theory), the rationale for justification is whether or not the conduct is something that we are content to have the actor perform due to the justifying circumstances and to have others perform under similar circumstances in the future. The test for justification is whether, on balance, the conduct in fact avoids a net societal harm (in the broadest sense). An actor's reasons may be relevant to liability under other criminal law doctrines: A mistaken reasonable belief that the conduct is justified may exculpate under an excuse defense; a mistaken belief that the conduct is not justified may inculpate as an impossible attempt offense. But an actor's reasons are not relevant under the deeds theory in determining whether a justification defense is available.

This, then, is the point of dispute in the theory of justification: Is the objectively justified nature of the *deed* central, as the deeds theory would have it, or irrelevant, as the reasons theory suggests?

Most commentators have signed on in support of the reasons theory and in opposition to the deeds theory,⁶ some suggesting that

4. MODEL PENAL CODE § 3.01 (1985) (emphasis added).

5. See, e.g., *id.* §§ 3.01-3.11 (setting forth justification formulations).

6. See, e.g., WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., *SUBSTANTIVE CRIMINAL LAW* 685 (1986) (claiming that in order to have the benefit of justification one

the latter is "absurd,"⁷ unfair,⁸ or unduly burdensome.⁹ (It is worth noting that the first two of these objections are based on the moral intuitions of the writers, coupled with their certainty that others share their moral intuitions.) Taking the minority side, one of us has argued that a deeds theory of justification is better for a variety of reasons, including that it generates liability results that are more just and that better match our collective intuitions of what is just.¹⁰

Most, but not all, states appear to follow the reasons theory in their criminal law,¹¹ although there is often some ambiguity as to which theory of justification they actually adopt, despite the apparent clarity of first appearances.¹² The Model Penal Code formulation is quoted above: An actor is justified if she *believes* that her conduct is necessary for defense.¹³

Current English law also appears to adopt the reasons theory. Professors Smith and Hogan, for example, conclude that the law "is

must act for that particular purpose); J.C. SMITH & BRIAN HOGAN, *CRIMINAL LAW* 37 (8th ed. 1996) (requiring state of mind as well as state of fact for justification to be reasonable); Michael Corrado, *Notes on the Structure of a Theory of Excuses*, 82 J. CRIM. L. & CRIMINOLOGY 465, 489 (1991) (arguing that state of mind is a necessary component of justification and that Robinson's proposed externalist perspective is impossible to accept); Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1897 (1984) (recognizing that most modern statutes require a subjective belief in justification and that Robinson's fully objective approach is an exception).

7. See Brian Hogan, *The Dadson Principle*, 1989 CRIM. L. REV. 679, 680 ("It seems to me absurd to say that I may *justify* or *excuse* my conduct, however callous it was in the circumstances known to me at the time, by showing that there existed other circumstances which, had I but known of them, would have justified or excused my conduct.").

8. See Arnold H. Loewy, *Culpability, Dangerousness, and Harm: Balancing the Factors on Which Our Criminal Law Is Predicated*, 66 N.C. L. REV. 283, 289 (1988) (arguing that, as a matter of fairness, the issue ought to be one solely of culpability rather than result).

9. See Kevin McMunigal, *Disclosure and Accuracy in the Guilty Plea Process*, 40 HASTINGS L.J. 957, 978 (1989) ("[A purely objective view of self-defense] is a more difficult factual question for the defendant to resolve than the question of her own subjective belief since calculation of the harm threatened involves a number of variables [that] are beyond the defendant's ability to perceive . . .").

10. See Robinson, *supra* note 2.

11. See 2 ROBINSON, *supra* note 3, § 184(b), at 399-403 (listing state criminal code justification sections that include a requirement that the actor "believe" his or her conduct is justified).

12. See Robinson, *supra* note 2, at 51-54 (noting that the Model Penal Code ultimately recognizes the importance of the distinction between a "belief" in a justification and actual, objective justification, when it creates the concepts of "privileged" and unprivileged justification).

13. See, e.g., MODEL PENAL CODE §§ 3.02(1), 3.03(3)(a), 3.04(1), 3.05(1)(b), 3.06(1), 3.07(1) (1985) (emphasis added); see also *supra* text accompanying note 4 (quoting the Model Penal Code formulation).

stated exclusively in terms of the defendant's belief,"¹⁴ citing the cases of *Gladstone Williams*, *Dadson*, and *Thain*.¹⁵ Section 24 of England's Police and Criminal Evidence Act of 1984 appears to be an exception to the general rule, for it justifies an arrest even if the officer did not at the time know of or believe in the justifying circumstances,¹⁶ reflecting a deeds theory of justification. Clauses 44 and 185 of the proposed Criminal Code for England and Wales apparently would broaden this exception to make it the general rule.¹⁷ That is, they adopt a deeds theory as their general approach. They provide a justification defense if the actor "uses such force as, in the circumstances *which exist*," is immediately necessary and reasonable for defense.¹⁸ Interestingly, the drafters claim that the provision codifies the common law of self-defense and defense of another.¹⁹ They concede that it modifies the common law of defense of property but argue that this modification is necessary to avoid an irrational inconsistency between the rules for the different defensive force defenses.²⁰

The contrasts between the two theories are illuminated when we consider how the deeds theory and the reasons theory suggest different results at each of the two conflict points: (1) when the actor mistakenly believes her conduct is not justified (the unknowingly justified actor); and (2) when the actor mistakenly believes her

14. SMITH & HOGAN, *supra* note 6, at 265.

15. *Gladstone Williams*, 78 Crim. App. 276 (1983); *Regina v. Dadson*, 4 Cox C.C. 358 (Crim. App. 1850); *Regina v. Thain*, 1985 N. Ir. 457 (C.A.).

16. Police and Criminal Evidence Act, 1984, ch. 24, § 4(a) (Eng.) (providing that an actor may arrest without a warrant "anyone who is in the act of committing an arrestable offence"); *id.* § 5(a) (providing that an actor may arrest without a warrant "anyone who is guilty of the offence"); *id.* § 7(a) (providing that an actor may arrest without a warrant "anyone who is about to commit an arrestable offence").

17. See 1 A CRIMINAL CODE FOR ENGLAND AND WALES cls. 44, 185 (Law Commission, Report and Draft Criminal Code Bill 1989).

18. 1 *id.* (emphasis added). These same proposed code provisions also allow a defense if the actor uses such force as, "in the circumstances . . . *which he believes to exist*," is immediately necessary and reasonable for defense. 1 *id.* (emphasis added) The use of this "believes" language does not make the provision one based upon a reasons theory of justification, for it still allows a justification without requiring proof of a belief in the justifying circumstances. The effect of such language is to allow a defense either upon actual or believed justifying circumstances. Nothing in the deeds theory prohibits a defense for mistake as to a justification. On the contrary, it assumes that such a defense will be provided but will be understood to be an excuse. Note that the provision of the proposed code does not identify either defense as a justification or an excuse.

19. The drafters explain: "[I]f his defence is that he was defending his person, or that of another, the test at common law is whether what he did *was* reasonable." 2 *id.* ¶ 12.25, at 231 (emphasis added).

20. See 2 *id.*

conduct is justified (mistake as to a justification).

A. *The Unknowingly Justified Actor*

Assume a person's conduct is objectively justified but he does not realize this; he mistakenly believes it is unjustified. For example, the person mugs a jogger, only to find out that the victim was a club-wielding attacker. Whether the beating of the attacker-thought-to-be-a-jogger is justified depends on whether it is the quality of the deed or the actor's reasons for it that provide the rationale for justification defenses.

Under the deeds theory, when the person's conduct in fact avoids a greater societal harm but the person is unaware of this, the conduct is justified despite the actor's ignorance. However, the person's belief that the conduct is not justified will give rise to attempt liability (assuming the jurisdiction punishes legally impossible attempts, as most do²¹). Thus, the use of force against the attacker-thought-to-be-a-jogger is justified and the actor will have a defense to the substantive offense for assault, but the actor will be liable for an attempted assault. Attempt is an offense that exists to punish just such manifested intention to commit an offense, when the harm or evil of the offense does not in fact occur, and presents a situation analogous to that of the unknowingly justified actor, who has manifested an intention to act unjustifiably, but in fact no net societal harm has occurred.

Under the reasons theory, if the justifying circumstances exist but the actor is unaware of them and acts for a different purpose, a justification defense is denied. If what matters is the reason for the deed, not the deed itself, the force used against the attacker-thought-to-be-a-jogger is not justified. While it might have been the right deed, necessary for self-defense, it was for the wrong reason.

B. *Mistake as to a Justification*

More common is the reverse case: A person's conduct is objectively unjustified but the person subjectively, mistakenly believes that it is justified. In such cases of mistaken justification, the actor believes that her conduct avoids a greater harm, when in fact it does not. The club-wielding attacker, when successfully overcome and dragged to the street light, turns out to be a jogger carrying a flashlight whose bulb is out. Whether beating the jogger-mistaken-

21. See 1 ROBINSON, *supra* note 3, § 85(c), at 428 n.28 (listing and updating authorities).

for-an-attacker is justified depends again on whether the justification defense is given (1) because the conduct in fact is justified, or (2) because the person acts for a justified reason.

Under a reasons theory, the force used against the jogger-mistaken-for-an-attacker is justified because it is used for the purpose of self-defense. The actor's reason is right even if the conduct is wrong. Under the deeds theory, a person who mistakenly believes that the conduct is justified is not justified, although the person may gain an excuse defense if the mistake is reasonable or perhaps a mitigation even if it is not.²²

Note that in this case the end result under the two theories seems to be the same. The difference is primarily a labeling matter: The person who reasonably but mistakenly believes that her conduct is justified is "justified" under the reasons theory but only "excused" under the deeds theory. This difference may have practical implications for third parties. For example, a jurisdiction might criminalize resistance to justified force, like a lawful arrest, yet allow resistance to excused force, like that of the psychotic aggressor.²³ This common approach creates problems for use of the reasons theory. Presumably we want the victim to be able lawfully to resist the actor who is mistaken as to a justification. The jogger mistaken for a mugger ought to be able lawfully to resist the misguided attack. But, if the attacker mistakenly believes she is justified and, as the reasons theory would have it, that makes her "justified," then the criminal code must do some fancy dancing to reach the proper result. It must create a special rule that allows defenders to defend against these special kinds of justified attacks but not other kinds of justified attacks.

But, putting aside this third-party complication for the reasons theory, the result of the two theories is the same for the actor at hand. Both the deeds theory's excuse defense and the reasons theory's justification defense exculpate the actor who mistakenly believes she is justified. Yet, the different views of justification may show themselves in another aspect of defense formulations that has been the subject of much disagreement: the proper treatment of mistake as to a justification. All agree that a reasonable mistake as to a justification ought to exculpate fully. What is the proper

22. See, e.g., MODEL PENAL CODE § 3.09(2) (1985) (providing reduced liability for unreasonable mistakes as to a justification).

23. See, e.g., *id.* § 3.11(1) (defining "unlawful force," which triggers a right of justified defensive force, as including the attack of the psychotic aggressor but excluding privileged (objectively justified) force, such as that used to make a lawful arrest).

treatment of an unreasonable mistake as to a justification?

A majority of jurisdictions permit a mistake-as-to-a-justification defense only if the actor's mistake is reasonable. An unreasonable mistake, reckless or negligent, gives no defense and hence generates full liability for the substantive offense.²⁴ A minority of jurisdictions give a complete defense for reasonable mistake but also allow a mitigation for an honest but unreasonable mistake.²⁵ The level of liability, that is, the extent of the mitigation given, typically is tied to the level of culpability of the mistake: A negligent mistake, being less culpable than a reckless mistake, gives a greater mitigation than does a reckless mistake.²⁶

The few jurisdictions that take the deeds approach in formulating their justification defenses objectively (for example, North Dakota and the Proposed Federal Criminal Code) all give mitigations for unreasonable mistakes as to a justification.²⁷ In contrast, a majority of the jurisdictions that take a reasons approach and formulate their justification defenses subjectively take the all-or-nothing approach, giving no mitigation or defense for unreasonable mistakes as to a justification.²⁸ This pattern suggests a connection between the deeds-reasons dispute and the dispute over the proper treatment of unreasonable mistakes as to a justification. But the fact is there is no logical reason why the reasons theory should demand an all-or-nothing approach.

One could speculate about the source of the apparent correlation between the reasons theory and the all-or-nothing approach. If one views mistaken justification as a justification, it would be easy to conclude that an all-or-nothing approach is needed. After all, all-or-nothing is the way objective justification does operate. Either the actor's conduct avoids a greater harm and is to be encouraged or at least tolerated in the future, or it does not avoid a greater harm and is to be discouraged in the future. When the subjective reasons theory of justification combines objective justification and mistaken subjective justification under the same

24. See 2 ROBINSON, *supra* note 3, § 184(a), at 395 n.1 (listing and updating authorities).

25. See 2 *id.*

26. See, e.g., MODEL PENAL CODE § 3.09(2).

27. See N.D. CENT. CODE §§ 12.1-05-01 to -09 (1985); NATIONAL COMMISSION ON REFORM OF FEDERAL CRIMINAL LAW, A PROPOSED NEW FEDERAL CRIMINAL CODE ch.6 (1971). For a list of jurisdictions that have at least one objective justification statute, see 2 ROBINSON, *supra* note 3, § 122(e), at 22 n.19.

28. See 2 ROBINSON, *supra* note 3, § 184(a), at 395 n.1 (listing and updating authorities).

label, "justified," it should be no surprise that such labeling creates the tendency to treat a mistaken justification as if it were a true objective justification. It should be no surprise to see mistake as to a justification treated, like all other justifications, as an all-or-nothing issue.

This same possibility for confusion does not exist under the deeds approach to formulating justifications. The deeds theory distinguishes true objective justifications from mistakes as to a justification, and it treats the latter as excuses. Objective justifications are properly all-or-nothing matters. Mistakes as to a justification, like other excuses, just as clearly are not all-or-nothing matters. Excuses function as part of law's adjudication of an actor's blameworthiness for a violation. Blameworthiness exists on a continuum, as is evident by the doctrines that contribute to this function.

Among the different functions of the criminal law, objective justifications serve the *ex ante* rule articulation function, telling people the rules for future conduct. Conversely, excuses, including mistaken justification, perform an *ex post* adjudication function, assessing the degree of liability and punishment for a violation of the rules of conduct.²⁹

In performing the adjudication function, doctrines commonly express degrees of liability and punishment. For example, criminal codes typically recognize levels of culpability: purpose, knowledge, recklessness, and negligence. The law also recognizes mitigations for partial excuses in both its definition of offenses (such as the extreme mental or emotional disturbance mitigation in homicide³⁰) and its sentencing rules (such as the federal sentencing guideline authorization for sentence reduction below the guidelines for offenders influenced by coercion, duress, or diminished capacity³¹). Under the deeds theory, mistakes as to a justification are seen as excuses, and like other doctrines for the adjudication of an actor's blameworthiness, the resulting liability may reflect a continuum of liability. Reasonable mistakes may excuse entirely, while the culpability inherent in unreasonable mistakes may suggest something less—a mitigation rather than a defense.

29. See Paul H. Robinson, *A Functional Analysis of Criminal Law*, 88 NW. U. L. REV. 857, 889-96 (1994); Paul H. Robinson, *Rules of Conduct and Principles of Adjudication*, 57 U. CHI. L. REV. 729, 729-71 (1990).

30. See, e.g., MODEL PENAL CODE § 210.3(1)(b).

31. See, e.g., U.S. SENTENCING GUIDELINES MANUAL §§ 5K2.12, 5K2.13 (1997).

C. *The Importance of Community Views in the Formulation of Criminal Law*

Before we launch into an exposition of the process of determining community moral intuitions in the reasons versus deeds controversy, we ought to say just why we think community views are relevant to the debate. First, those debating the issue have conceded the relevance of community views, when they make statements pointing out that the deeds theory of justification is "absurd."³² What this conclusion turns out to mean is that the author of that statement feels that the deeds theory violates his moral intuitions, and it implies that his moral intuitions are those that all community members would also hold. This is an empirical proposition; it may turn out that the community holds views that resemble a deeds theory of justification rather than a reasons theory or, as that writer asserts, vice versa. In any event, those writers appealing to moral intuitions to support their theory have conceded, at a minimum, that the community intuitions deserve a place in the debate.³³ Thus one task is to determine community moral intuitions, the degree to which these intuitions support any of the relevant justificatory theories, and the consensus with which community members hold their views.

A second reason for seeking to discover community views is that it is at least necessary to know when legal codes, for whatever reason, conflict with or override the moral intuitions of the governed community. For when they do, it is useful for the code drafters to educate the community as to why the code formulation is preferable, either morally or otherwise. Without this education, conflicts that the community discovers between legal codes and moral intuitions are likely to engender socially destructive sentiments and actions on the part of the governed. (Think of the consequences of prohibition in the United States.³⁴)

Of course, it is possible that citizens are unaware of the conflict between codes and their moral intuitions—which brings us to the third reason to discover community intuitions in various areas governed by codes. Absent knowledge of the true provisions of the code, citizens are likely to believe that the code conforms with their own moral intuitions.³⁵ If the community moral intuitions are in fact

32. See Hogan, *supra* note 7, at 680.

33. See *supra* note 3 and accompanying text.

34. See generally Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453, 487 (1997).

35. In a recent study, New Jersey citizens reported that attempt was criminalized by the state code in ways that were tightly coupled with their own moral intuitions, but very

quite deviant from the actual content of the code, the code is failing its *ex ante* function, failing to provide known clear guidelines that people can use to govern their conduct.

Finally, we have argued elsewhere that the criminal code ought to be in general agreement with the moral principles of those the code governs.³⁶ Here is a brief summary of the argument: The real power to enforce compliance with society's rules of prescribed conduct lies not in the threat or reality of official criminal sanction, but in the power of the intertwined forces of social and individual moral control. The networks of interpersonal relationships in which people find themselves, the social norms and prohibitions shared among those relationships and transmitted through those social networks, and the internalized representations of those norms and moral precepts cause people to obey the law.³⁷

The law is not irrelevant to these social and personal forces. Criminal law, in particular, plays a central role in creating and maintaining the social consensus necessary for sustaining moral norms. In fact, in a society as diverse as ours, the criminal law may be the only society-wide mechanism that transcends cultural and ethnic differences. Thus, the criminal law's most important real world effect is arguably its ability to assist in the building, shaping, and maintaining of these norms and moral principles. It contributes to and harnesses the compliance-producing power of interpersonal relationships and personal morality.³⁸

The criminal law can have a second effect in gaining compliance with its commands. If it earns a reputation as a reliable statement of what the community, given sufficient information and time to reflect, perceives as condemnable, people are more likely to defer to its commands as morally authoritative and as appropriate to follow in those borderline cases in which the propriety of certain conduct is unsettled or ambiguous in the mind of the actor. The importance of this role should not be underestimated; in a society with the complex interdependencies characteristic of ours, an apparently harmless action can have destructive consequences. When the action is criminalized by the legal system, one wants the citizen to respect the law in such an instance even though he or she does not immediately

poorly coupled to the actual code provisions. See John Darley et al., *Community Standards for Defining Attempt: Inconsistencies with the Model Penal Code*, 39 AM. BEHAV. SCI. 405, 414 (1996).

36. See Robinson & Darley, *supra* note 34.

37. See *id.* at 468-71.

38. See *id.* at 471-74.

intuit why that action is banned. Such deference is facilitated if citizens are disposed to believe that the law is an accurate guide to appropriate prudential and moral behavior.³⁹

The extent of the criminal law's effectiveness in both these respects—in facilitating and communicating societal consensus on what is and is not condemnable, and in gaining compliance in borderline cases through deference to its moral authority—is to a great extent dependent on the degree of moral credibility that the criminal law has achieved in the minds of the citizens it governs. Thus, the criminal law's moral credibility is essential to effective crime control and is enhanced if the distribution of criminal liability is perceived as “doing justice,” that is, if it assigns liability and punishment in ways that the community perceives as consistent with the community's principles of appropriate liability and punishment. Conversely, the system's moral credibility, and therefore its crime-control effectiveness, is undermined by a distribution of liability that deviates from community perceptions of just desert.

The central point is this: The criminal law's power in nurturing and communicating societal norms and its power to have people defer to it in unanalyzed cases are directly proportional to criminal law's moral credibility. If criminalization or conviction (or decriminalization or refusal to convict) is to have an effect in the norm-nurturing process, it will be because the criminal law has a reputation for criminalizing and punishing only that which deserves moral condemnation, and for decriminalizing and not punishing that which does not. If, instead, the criminal law's reputation is one simply of a collection of rules, which do not necessarily reflect the community's perceptions of moral blameworthiness, then there is little reason to expect the criminal law to be relevant to the societal debate over what is and is not condemnable and little reason to defer to it as a moral authority.

We now need to turn to the task of discovering how the community thinks about the cases that discriminate a reasons theory of justification from a deeds theory, whether there are describable principles that match the community's judgments, and whether there is some degree of consensus among the judging community.

39. *See id.* at 474-76.

II. TOOLS FOR TESTING CRIMINAL LAW THEORIES: SOCIAL PSYCHOLOGY SCENARIO RESEARCH

A. *Scenarios and Measures*

The method we chose to probe subjects' moral intuitions in this study was the scenario or vignette method. Subjects are presented with a short description of a person's conduct and are asked whether the actor should receive liability for the conduct and, if so, how much. Subjects next are given another scenario, and assess liability and punishment for that actor, then another scenario, and so on. The scenarios are varied by the researchers in ways suggested by the theories being tested, and the patterning of liabilities assigned each scenario provide differential support for the competing theories.

Rather than having the subjects work their way through what can quite quickly become a large number of differing scenarios, why not just ask the subjects whether they think a reason-centered or a deed-centered theory of justifications is appropriate? Because psychologists have discovered that subjects often do not have mental access to the principles they use to make decisions and thus they cannot accurately articulate those principles.⁴⁰ Instead they are often driven to report principles that seem plausible to them at the time but demonstrably do not match their actual decisions.⁴¹ Therefore researchers carry out what is called a policy-capturing study, in which

40. See Paul Slovic & Sarah Lichtenstein, *Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment*, 6 ORGANIZATIONAL BEHAV. & HUM. PERF. 649 (1971). In this article, the authors describe an early demonstration of this proposition that is relevant to the present study. They reviewed studies in which the subject's task was first to make judgments and second to articulate how they came to the judgments they made. See *id.* at 683-84. An example of such a judgment would be what weights a stockbroker assigned to various items of information in forming his judgment about the desirability of a particular company's stock. In the studies they reviewed "all found serious discrepancies between the subjective and [objective] relative weights." *Id.* at 684. That is, the judgment-makers were inaccurate in reporting the weights they in fact placed on the various dimensions when they made their actual judgments. See *id.*

More general evidence for this proposition is reviewed in Richard E. Nisbett & Timothy Decamp Wilson, *Telling More than We Can Know: Verbal Reports on Mental Processes*, 84 PSYCHOL. REV. 231, 231-59 (1977). The authors' study shows that individuals sometimes fail to acknowledge the influence of aspects of the stimulus that actually make a difference in their judgments, see *id.* at 243-45, and other times report as influential the effect of aspects that in fact do not enter their judgment processes at all, see *id.* at 245-46.

41. See Nisbett & Wilson, *supra* note 40, at 247-59 (suggesting that people report what their own implicit theories of decision-making mark as important, rather than what is actually important to them in practice).

subjects judge actual cases. The researchers then infer the subjects' judging principles from the resulting patterning of responses to different cases.⁴²

This is what we did in the present research. We presented subjects with short scenario descriptions of potentially criminal actions. Because the focus of our research was on contrasting the reasons theory with the deeds theory of justification defenses, we designed the variations in our scenarios to reflect those different theories. Generally, two cases differed in a way that would "make a difference" to, for instance, a person who held a reasons-centered view but not to a person with a deeds-centered view. "Make a difference" here means that the two cases would generate different liability judgments if the subject took one view of the theory of justification, but not if she took the other view.

In other words, we conducted an experiment. Experimentation is an unusual tactic in research concerning legal issues; other empirical techniques such as opinion surveys, or the examination of existing records, or other archival procedures are more common. Part of what we seek to demonstrate to criminal law theorists and code drafters is that experimentation, used to capture individuals' patterns of liability assignment, can provide useful information on their issues of debate.

Subjects first read a paragraph of core information that gave the background to the various scenarios:

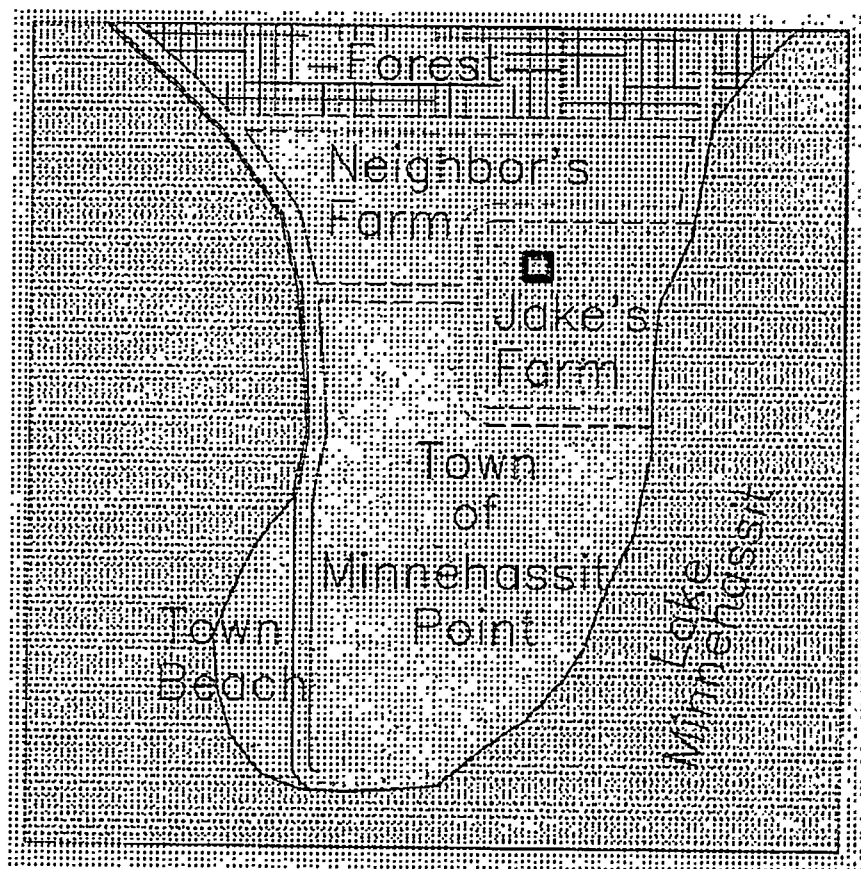
Jake is a farmer who has already harvested his corn crop. His neighbor has not done so, so his three acres of corn are still in the fields. The corn crop makes the difference, for these farmers, between having a profitable season because they have winter feed for their animals, or going into debt.

Running around several sides of Jake and his neighbor's fields are dirt roads. Jake's farm and his neighbor's farm are on a neck of land that stretches out into a lake. Out on the end of the neck of land is the local town. Jake's neighbor's fields cut the town off from the mainland, but Jake's fields do not. The following map shows you this layout.

42. See Robert S. Billings & Stephen A. Marcus, *Measures of Compensatory and Noncompensatory Models of Decision Behavior: Process Tracing Versus Policy Capturing*, 31 ORGANIZATIONAL BEHAV. & HUM. PERF. 331, 331 (1983) (commenting that "one of the oldest and most widely used techniques (for making inferences about the decision process) is that of policy capturing, wherein the model guiding the decision process is inferred from the relationship between the cues provided and the judgment of the subject").

At this point, the subjects were given the map below in Figure 1, which makes clear the essential point: The neighbor's fields, but not Jake's fields, if burned, would create a successful firebreak for the town.

FIGURE 1



Next the subjects read a specific scenario and assigned a liability to the perpetrator described in it. For instance, one offense scenario read as follows:

Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, and there is no wind, it is clear that the burning will create no danger beyond that of destroying the corn. The fire destroys the entire crop.

This scenario is obviously a case of an offense of moderate seriousness, and we use it to establish the sentence that subjects would give to this particular offense. It provides a point of comparison for later cases, a "control" or "contrast" case, to see whether subjects think the subsequent cases, in which justifications for the burning are given, deserve less liability and punishment than the prototype of the unjustified offense.

Some people may assign consistently higher liabilities than others. These differences are not the focus of our concerns. In the experimental design we chose, in which subjects responded to several scenarios, it is not relevant whether each subject was a generally harsh or easy sentencer. Our interest is in the patterning of the difference in liability between specific scenarios, not the absolute amount of liability in any scenario.

As noted above, scenarios differ in ways designed to elicit one pattern of liability assignments if the subject uses a reasons theory of justification, and another if the subject uses a deeds-theory. For instance, another scenario read as follows:

Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, it is clear that the burning will create no danger beyond that of destroying the neighbor's corn. The fire destroys the entire crop. *Unbeknown to Jake*, lightning has started a fire upwind from his and his neighbor's fields and the local town and the fire is burning toward the town and endangering the people who live there. His burning the field creates a firebreak: The town and its inhabitants are saved.

From a reasons perspective, the perpetrator is equally liable in both cases—he intended a harmful act. The fact that the act had a second, helpful, unintended, consequence is irrelevant. But it is not irrelevant to a person holding a deeds-based theory. Therefore an individual assigning a significantly lower liability to this second case is revealing that she holds a deeds-based theory of justification, while an individual who assigns this case the same liability as the previous one is revealing a reason-based theory.

Notice that we have attempted, and we hope succeeded, in making the two scenarios differ in only one way, the way that is relevant to the theoretical comparison in question. The subjects should have perceived the different scenarios as having the same

overall characteristics, so that any differences in liabilities assigned can be attributed to the one characteristic that is varied between the contrasting scenarios.

The task of each subject, then, in response to each scenario was to assign a degree of liability to the protagonist in the scenario—in their view, to assign punishment to a wrongdoer. Subjects did this by marking their judgment on the scale shown below, a scale with which they quickly became familiar:

LIABILITY SCALE

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprisonment	death
	but no punish.											

As can be seen, the scale gave subjects a choice of assigning to the protagonist no criminal liability, liability but no punishment, or eleven levels of punishment, prison sentences ranging from one day in jail to the death penalty. Notice that the difference between two adjacent prison sentences becomes greater as one moves to the right end of the scale. For instance, an assignment of punishment level 2 is an assignment of two weeks in prison, an increase of only thirteen days over punishment level 1. An assignment of punishment level 9 is a fifteen-year increase from the punishment represented by level 8.

We constructed the scale in this way for two reasons. First, and primarily, the differences correspond to the differences in grading categories used in typical American criminal codes.⁴³ This correspondence is quite important because it means that we can translate a difference between two liability units into a difference of

43. See, e.g., FLA. STAT. ANN. § 775.082 (West Supp. 1998) (defining five grades of felonies and two grades of misdemeanors, carrying statutory maximum punishment terms of the death penalty and life imprisonment, 30, 15, 5, and 1 year, and 60 days imprisonment); 730 ILL. COMP. STAT. 5/5-8-1, 5/5-8-3 (West 1993) (defining seven grades of felonies and three grades of misdemeanors, carrying the maximum terms of death penalty and life imprisonment, 30, 20, 15, 7, 5, 3, and 1 year, 6 months, and 30 days imprisonment); N.Y. PENAL LAW §§ 60.06, 70.00, 70.15 (McKinney 1998) (defining five grades of felonies and two grades of misdemeanors, carrying the maximum terms of the death penalty and life imprisonment, 25, 15, 7, 4, and 1 year, and 3 months imprisonment); VA. CODE ANN. §§ 18.2-9, 18.2-10, 18.2-12 (Michie 1996) (defining six grades of felonies and four grades of misdemeanors, carrying the maximum terms of the death penalty and life imprisonment, 20, 10, 5, and 1 year, and 6 months imprisonment).

one offense grade in a criminal code.⁴⁴ Second, this scheme is useful because the differences correspond, roughly at least, to what ordinary people perceive as equal differences between sentences. Thus these sorts of differences are the ones available to code drafters when they decide how to grade an offense, the ones juries and judges must deal with when sentencing a convicted offender, and perhaps the ones that come to the minds of citizens when they read and think about criminal sentences.⁴⁵

In designing the scenarios, our task was to create as many as were needed to provide a reasonably complete test of the implications of the reason- and deed-based theories and their differences. We found that eleven scenarios were needed. The full text of these scenarios is presented in the Appendix to this Article. What is revealed by contrasts between various scenarios we will describe in detail in Part IV.

Pilot testing indicated that the eleven scenarios could be read and evaluated by a subject in approximately half an hour. Further, the subjects were able to maintain concentration; their reports indicated that they found the task quite interesting and were intrigued by thinking about what differences in the cases "made a difference" to them. Each of our subjects responded to all of the cases. In the experimental design literature, a study having these characteristics is referred to as a "within-subjects design." This design focuses the subjects' attention on the differences between the scenarios. The danger is that they think that the existence of a difference implies an instruction from the researcher that the difference should "make a difference," that is, that it should provoke different liability assignments from the subject.⁴⁶ To counter this possibility, we told subjects that we did not expect that different scenarios necessarily should get different liability judgments and that they were to give us their own judgments about what differences mattered. Looking over the individual response protocols from this experiment and other similar ones we have conducted, we note that

44. The utility of this correspondence will become clear in later discussions of the interpretations of the results. *See infra* text accompanying notes 57-70.

45. The scale used here is the same as the one developed and first used in PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* (1995), in which we report the results of a number of studies that map the community's perceptions of the appropriate liabilities to assign in various criminal situations. For a more lengthy discussion of the scale and its properties, see *id.* at 223-25.

46. For a fuller discussion of the strengths and weaknesses of within-subjects designs in this sort of research, see *id.* at 221-22.

subjects did rate some cases alike as to the liabilities they generated.

As is usual in these designs, the order in which the cases were given to the subjects was randomized. To make the subjects' contrasting task simpler for them, cases with one dimension of variation were grouped together. The order of cases within these groups, as well as the order in which the groups were presented, was randomized in order to prevent results from being undetectably dependent upon the order in which the scenarios were judged.

B. The Sample of Subjects

Any research study must select subjects from the population about which the research generalizations are intended to apply. Our concern is with the moral intuitions of the community of citizens governed by the laws in question. Given that this research is about differences in the rationale for criminal sentencing that exist at the national level, eventually one would want to construct a national sample of subjects. For this initial study of the issue, practical considerations limited our selection of one set of subjects to the lists of jury-eligible citizens in a town in New Jersey.⁴⁷ The second set consisted of college students who were readily available for research. It is sometimes suggested that students are atypical, in that their responses would differ from so-called ordinary people. Since we had both students and ordinary people in our research, we were able to test this contention.

We tested twenty-seven students (average age 19.2) and twenty-one jury-eligible community members (average age 50.8); men and women were equally represented in both samples, as were the major religious affiliations of Protestant, Catholic, and Jewish. Two of each sample were African-Americans, for a total of four in a panel of forty-eight. Most subjects in both samples identified themselves as politically moderate, with the students leaning a little more to the liberal side of the continuum. Students filled out the questionnaires in a room on campus; for the community members, questionnaires were mailed out to them, and occasionally after a telephoned reminder, the questionnaires were mailed back to us. As expected, conservatives assigned slightly higher liabilities to the various scenarios we presented. Our jury-eligible community members also assigned slightly higher liabilities, over and above the fact that

47. Of those contacted, 56% agreed to participate. This rather high rate of participation was probably the result of the subjects' agreement with our explanation of the goals of the research.

conservatives did so. What is important, for our purposes, is that the two groups of subjects did not show any significant difference in their *pattern of relative liability* assigned among the scenarios.

III. LIABILITY PREDICTIONS

In justification defenses, we are in the minority of commentators.⁴⁸ We believe the community's views are more accurately reflected in doctrine based upon a deeds theory of justification. It matters to lay persons whether a net societal harm actually occurs or not, we think, just as it matters to them whether a prohibited result, such as a resulting death, occurs or not. In particular, we think the community sees the unknowingly justified actor as deserving the reduced liability of attempt rather than the full liability that would come from denying a justification defense.⁴⁹ As to the reverse case of mistake as to a justification, we think the community views unreasonable mistakes as to a justification as deserving mitigation, in contrast to the majority rule in the United States.⁵⁰ We describe below exactly how these general claims translate into specific predictions with regard to the liability results of the scenarios used in the study.

The first six scenarios are contrast cases, the responses to which established benchmarks for each test subject. These scenarios provide the full range of possible liability, as well as a variety of intermediate points. Not only do they give us results of considerable intrinsic interest in their own right, but more importantly for the present purposes, they allow us to interpret the liability results of the last five scenarios, the test scenarios. For each test scenario, we used as a point of comparison the contrast scenario most relevant with respect to the competing theories, from a case of an intentional unjustified act to a completely justified act, or any one of the many possibilities between those two extremes on the continuum of liability. Taken together, the following six contrast cases represent all of the obvious variations of a non-justification case to which the subjects' responses might be compared.

48. See *supra* notes 2-10 and accompanying text.

49. See *supra* text accompanying note 21 (discussing attempt liability).

50. See *supra* notes 22-28 and accompanying text.

A. *The Contrast Cases*

Scenario 1. Intentional (Unjustified) Burning

Scenario 1 presents the prototype case of a burning that is intentional and for which no claim of justification exists:

Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, and there is no wind, it is clear that the burning will create no danger beyond that of destroying the corn. The fire destroys the entire crop.

We expect the liability here to be somewhere mid-scale because the offense is against only property. Our past work suggests that it is not likely to inspire the heavy penalties at the higher end of the scale.⁵¹ Its purpose is to give us a liability rating against which we can compare the liabilities assigned to other scenarios.

Scenario 2. Attempted (Unjustified) Burning

Scenario 2 is similar to scenario 1 except that the harm intended does not actually come about:

Just as Jake sets the fire, the neighbor unexpectedly returns and puts it out before it does any harm.

A crime like this one typically is treated as an attempt, as compared to that in scenario 1, which is referred to as the substantive offense. From our past work, we expect the liability here to be substantially less than that imposed in scenario 1, even though the actor's conduct and intention are identical in the two cases.⁵² The fact is, the vast majority of lay persons share a strong intuition that whether or not the planned harm does or does not occur makes a difference and that the occurrence of harm increases the punishment deserved. No claim of justification is at issue in scenario 2.

Scenario 3. Created Risk of (Unjustified) Burning, Realized— Reckless Commission

Scenario 3 differs from scenario 2 in that the actor only risks the burning, rather than intending it. But as in scenario 1, here the harm actually comes about:

51. See, e.g., ROBINSON & DARLEY, *supra* note 45, at studies 6, 8, 11, 18.

52. See *id.* at studies 1, 17.

Like all of the local farmers, Jake routinely piles dry cornhusks near where they are cut and eventually burns them. Jake has one such pile near his neighbor's fields. Jake wants to get the pile burned quickly; the previous year he waited and the pile got soaked by rain. He is aware that high winds are forecast for today; winds that create a real risk that his fire will jump the gap between his trash pile and his neighbor's corn fields. Despite this danger, Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. The winds come and the fire jumps to his neighbor's crop. The fire destroys the entire crop.

We expect the liability here, as in scenario 2, to be less than that in scenario 1. Such an offense typically is termed a "crime of recklessness."

Scenario 4. Created Risk of (Unjustified) Burning, Unrealized— Endangerment

Scenario 4 is similar to scenario 3, but here, luckily, the harm does not come about:

Despite [the danger from the high winds], Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. The winds come but, Jake is lucky, the fire does not jump to his neighbor's crop.

We think the liability will be less here than in scenario 3. Liability also will be less than in scenario 2, we think, because here the actor does not intend the harm but only risks it. This is an application of the principle noted in our earlier work that greater punishment is due for greater culpable state of mind.⁵³ Intending to burn is more culpable than intending to create a risk of burning, all other things being equal. Offenses like this commonly are termed "endangerment" offenses. No claim of justification is at issue in the scenario.

These four cases present variations of the culpability and harm variables, as evidenced in Table 1.

53. See *id.* at studies 8, 9, 16.

TABLE 1. INTERRELATION OF CONTRAST CASES 1 THROUGH 4

	Harm	No Harm
Intentional	Scenario 1 Substantive Offense	Scenario 2 Attempt
Reckless	Scenario 3 Reckless Offense	Scenario 4 Endangerment

Scenario 5. Attempted Risk Creation—Attempted Endangerment

In scenario 5 the actor thinks he is creating a criminal risk, but in fact no such risk is created. In other words, it is a case of attempted endangerment rather than the actual endangerment of scenario 4:

Despite [the danger of high winds], Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. It turns out that the weather forecast was in error about the wind. Jake's burning never creates any danger to his neighbor's field.

We think it will have even less liability than scenario 4. In scenario 4, a risk of the harm was in fact created, while here no such risk is created; the actor only mistakenly believes that it is created. Scenario 5 (risk intended but no risk occurs) bears the same relation to scenario 4 (risk intended and risk occurs) that scenario 2 (burning intended but does not occur) bears to scenario 1 (burning intended and occurs). No claim of justification is at issue in the scenario.

Scenario 6. Intentional Justified Burning

In scenario 6, the final contrast case, the burning occurs but is clearly justified, under both a reasons and a deeds theory. Not only do the objective circumstances actually exist that make the burning the right thing to do, but the actor knows of the justifying circumstances and acts because of them:

Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. He can see the

smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire crop. Because of his quick work the town and its inhabitants are saved.

We think Jake will get a complete defense in this scenario.

B. The Test Cases

We now turn to the scenarios that discriminate between the two competing theories. As described above, a subject's response to each of these cases is compared to her response to one or more of the contrast cases relevant from the point of view of the competing theories. It is from this comparison that we infer the subject's views.

Scenario 7. Unknowingly Justified Burning

Scenario 7 is the case of the unknowingly justified actor:

Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, it is clear that the burning will create no danger beyond that of destroying the neighbor's corn. The fire destroys the entire crop. *Unbeknown to Jake*, lightning has started a fire upwind from his and his neighbor's fields and the local town and the fire is burning toward the town and endangering the people who live there. His burning the field creates a firebreak: The town and its inhabitants are saved.

Recall that the deeds theory, which we think better represents community views, predicts this liability to be similar to that of scenario 2, the attempt case. The actor's liability is based entirely upon his intention to burn without justification, the classic rationale for punishing an attempt. The reasons theory predicts that the actor will have no defense and therefore will be liable for the full offense, the same liability as in scenario 1.

Scenario 8. Knowingly Justified Burning but with Bad Motive

Scenario 8 is a case in which the justifying circumstances exist and the actor knows about them, but he acts for a bad motive, rather

than for a justificatory purpose:

Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. He can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake has no interest in saving the town; the townspeople have always been unfriendly to him. Further, Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. He decides to use the fire as an excuse to burn his neighbor's corn crop. Without asking his neighbor for permission, he burns the fields. The fire destroys the entire crop. Because of his quick work the town and its inhabitants are saved.

This scenario is an interesting case because, recall from Part I.B., most jurisdictions implement the reasons theory by defense formulations that require only that the actor "believe" that the justifying circumstances exist.⁵⁴ They do not require that the actor act for the justifying "purpose," even though "purpose" is a standard culpability level commonly required by other criminal law doctrines.

Contrary to the legal rules, which give the same complete defense for both a justificatory purpose and mere knowledge of the justifying circumstances, we think the community will find a difference between the two cases. We predict that the person who acts for the justificatory purpose, as in scenario 6, will receive a complete defense (no liability), whereas the person who acts knowing only of the justifying circumstances but with a purpose other than to avoid the greater harm, as in scenario 8, will have some level of liability imposed.

A strict reasons theory might give no defense here. While the actor knew of the justifying circumstances, they were not his reason for acting. But the reasons theory as adopted in current law treats this actor as fully justified, thus imposing no liability.

The deeds theory would give this actor a significant discount from full liability, at least as great as that given the unknowingly justified actor, because while his motive may be bad, his conduct is objectively justified. On the other hand, the actor in scenario 8 at least does not think that he is causing a net harm, and therefore we

54. See *supra* text accompanying note 22.

think will have less liability than the unknowingly justified actor of scenario 7, who does think so.⁵⁵

These first two test scenarios present variations on the unknowingly justified actor. The next three scenarios consider the reverse case of the actor who mistakenly believes he is justified. The person setting the fire thinks that he has a justification for doing so, but his reasons for thinking this become increasingly poorly grounded.

Scenario 9. Mistake as to Justification, Reasonable

Scenario 9 presents the case of a reasonable mistake as to a justification:

Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. In the past, Citizen's Band radio reports have often been true, but also often false. Jake stops two cars that are racing into town, and both confirm that "there is a big, out of control fire, heading this way." (Any reasonable person would think there was a destructive fire coming.) Jake can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire crop. It turns out that the radio report was in error. The smoke was from a controlled burn being done by a crew of local foresters and presented no danger to the town or any of the surrounding area.

Both deeds and reasons theories would give a complete defense. Only the labeling would be different. The deeds theory would consider the actor excused; the reasons theory would consider the actor justified.⁵⁶ All jurisdictions agree that a full defense is appropriate in this case; no issue of mitigation arises. Both the deeds theory, which we support, and the reasons theory predict no liability.

55. Recall that the deeds theory relies upon attempt liability as the source of liability for the unknowingly justified actor. When an actor does not think that he is acting unjustifiably (he knows of the justifying circumstances), it is unclear that attempt liability is appropriate.

56. See *supra* text accompanying notes 22-23.

Scenario 10. Mistake as to Justification, Negligent

Scenario 10 is a case of a negligent mistake as to a justification, differing from scenario 9 as follows:

In the past, Citizen's Band radio reports have often been true, but also often false. Jake doesn't think of this, and although a reasonable person would do so, Jake doesn't think to check on the truth of the report, but there is no doubt in his mind that it is a dangerous fire. Jake can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire.

Those jurisdictions that require a reasonable mistake for a defense, a majority of reasons theory jurisdictions, would deny any defense here and would impose full liability, as in scenario 1. The jurisdictions that do recognize a mitigation for an unreasonable mistake as to a justification, which includes the few deeds jurisdictions, will impose more liability than the complete defense in scenario 6, but notably less than the full liability of scenario 1 that the majority view predicts.

Scenario 11. Mistake as to Justification, Reckless

Scenario 11 is a case of a reckless mistake as to a justification, when there is greater culpable state of mind than in scenario 10, but the actor still honestly and sincerely believes that he is acting justifiably. It differs from scenarios 9 and 10 as follows:

In the past, Citizen's Band radio reports have often been true, but also often false. Jake remembers this fact, and realizes there might not be a dangerous fire, but doesn't check on the truth of the report. He can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire.

The majority all-or-nothing approach again would give no defense and would impose full liability, as in scenario 1. The mitigation approach again would give a mitigation from full liability, although not as much as the mitigation given in scenario 10. Thus, we predict that the liability here will be greater than that in scenario 10, but markedly less than that in scenario 1.

Our predictions and those consistent with current law are summarized in Table 2. Generally, we think the subjects will agree with the predictions of the deeds theory and with mitigations for unreasonable mistakes as to a justification.

TABLE 2.
SUMMARY OF LIABILITY PREDICTIONS

CONTRAST CASES

Scenario	Expectation
1. Intentional commission of substantive offense	Baseline
2. Attempt	< 1
3. Reckless commission	< 1
4. Endangerment	< 2 and 3
5. Attempted endangerment	< 4
6. Justified commission	No liability

TEST CASES

Scenario	Our Predictions (Deeds Theory)	Predictions Consistent with Current Law (Reasons Theory)
7. Unknowingly justified burning	= 2 (attempted burning)	= 1 (no defense; full liability)
8. Justifying knowledge without justificatory purpose	> 6, but < 7	= 6 (complete justification)
9. Reasonable mistake as to justification (MJ)	= 6 (complete excuse)	= 6 (complete justification)
10. Negligent MJ	> 6, but < 1	= 1 (no defense; full liability)
11. Reckless MJ	> 6 and 10, but < 1	= 1 (no defense; full liability)

IV. LIABILITY RESULTS

The mean liability for each of the scenarios is set out in Table 3.

TABLE 3.
LIABILITY MEANS

Scenario	Liability Mean	Imprisonment Equivalent
Contrast Cases		
1. Intentional (unjustified) burning	4.65	~ 10 months
2. Attempted (unjustified) burning	3.52	~ 4 months
3. Created risk of (unjustified) burning, realized	2.69	~ 6 weeks
4. Created risk of (unjustified) burning, unrealized	0.48	essentially no punishment
5. Attempted risk creation	0.42	essentially no punishment
6. Intentional justified burning	0.57	essentially no punishment
Test Cases		
7. Unknowingly justified burning	3.63	~ 4 months
8. Knowingly justified burning but with bad motive	2.10	~ 2 weeks
9. Mistake as to justification, reasonable	1.10	~ 2 days
10. Mistake as to justification, negligent	2.02	~ 2 weeks
11. Mistake as to justification, reckless	2.33	~ 4 weeks

A. *The Contrast Cases*

We begin with an examination of the results for the cases that were designed to provide comparative information for the test cases. The scenario 1 contrast case of intentional burning with no claim of

justification has a liability mean of 4.65 (equivalent to about ten months imprisonment).⁵⁷ This result is what one might expect given the nature of the offense, a property offense in which no risk to persons is created.

The scenario 2 contrast, the attempt case, has a liability of 3.52 (just over four months). This result is consistent with our expectation of substantially reduced punishment based solely on the fortuitous absence of the intended harm. Indeed, the ratio of penalties between scenarios 1 and 2 is consistent with those jurisdictions that set the grade of an attempt as one grade less than or half the penalty of the substantive offense.⁵⁸ Recall that on our exponential penalty scale, one unit is equivalent to one offense grade in a typical modern American criminal code and that each higher grade typically doubles the penalty of the previous grade.⁵⁹

Scenario 3, in which the actor creates a risk of burning that is realized, has a liability of 2.69 (6.2 weeks). As predicted, it is less than the liability in scenario 1; here the actor does not intend the harm, but only risks it. The importance of this difference in culpability level often is reflected by corresponding differences in penalties. For example, in homicide cases, this same culpability difference results in an intentional killing being punished as murder with long-term or life imprisonment or death, while a reckless killing is punished as manslaughter with a maximum penalty more in the range of ten years.⁶⁰

Scenario 4, a case of risk creation in which the harm risked goes unrealized, receives a liability of 0.48, which is essentially no punishment (a liability mean of 1.0 is equivalent to one day imprisonment). No liability was assigned by 36.5% of subjects. Another 42.3% gave liability but no punishment. The remaining 21.2% gave punishment ranging from one day to six months. The deeds theory predicted low liability, from the concurrence of both the discount for no resulting harm seen in scenario 2, and the discount for lower culpability level seen in scenario 3. We have discussed elsewhere this additive nature of different discounts from the full

57. More precise translation from liability means to imprisonment terms can be obtained by using the table in ROBINSON & DARLEY, *supra* note 45, app. C at 283.

58. See PAUL H. ROBINSON, FUNDAMENTALS OF CRIMINAL LAW 297 (2d ed. 1995) (citing statutes).

59. See *supra* note 43 and accompanying text.

60. See, e.g., MODEL PENAL CODE § 210.3(2) (1985) (making manslaughter a second degree felony); *id.* § 6.06(2) (setting the maximum term of imprisonment for second degree felonies at 10 years).

intentional substantive offense.⁶¹ The liability result here seems a neat accumulation of the 1.1 discount from the full offense seen in scenario 2 and the 2.0 discount seen in scenario 3. Thus, a perfectly additive discount would be 3.1. The liability result here shows a 3.2 discount. It seems unlikely, however, that this additive discounting always will be quite so neat. This substantial reduction is consistent with current law's treatment of such matters. In homicide, for example, creating a risk of death that is unrealized is punished as endangerment, which typically carries a maximum sentence of only one year,⁶² compared to life imprisonment or death for murder.

Scenario 5 presents the case of attempted endangerment. As expected, the liability mean is low, 0.42. The result is only slightly less than the result in scenario 4 (endangerment), and that small difference is not statistically significant. We predicted a difference between the two, with scenario 5 less than scenario 4, to reflect the absence in scenario 5 of the risk that in fact is created in scenario 4. Our assumption is that the difference does not appear because the scenario 4 liability is already so low no further reduction is possible. In scenario 4, 78.8% of the subjects imposed no punishment. That leaves little room to distinguish scenario 5 as a case of even less blameworthiness. (In scenario 5, 86.5% imposed no punishment.) If scenarios with a more serious base offense were used, such as homicide, the distinction we expected here might appear.

Scenario 6, the final contrast case, is an intentional justified burning. As expected, it received essentially no punishment. Its liability mean was 0.57. No liability was assigned by 38.5% of the subjects. Another 40.4% gave liability but no punishment. The remaining 21.2% gave punishment ranging from one day to one year. This baseline is not as low as we might have guessed but still reflects the predicted judgment that the vast majority of subjects see this as a case of little or no blameworthiness, despite the fact that an intentional harm is caused.

To summarize the contrast case results, the results came out as we predicted. Those predictions, it will be remembered, were based on two principles: a liability discount given when the harm risked did not actually occur, and a liability discount given as the harm risked

61. See Paul H. Robinson & John M. Darley, *Objectivist vs. Subjectivist Views of Criminality: A Study in the Role of Social Science in Criminal Law Theory*, 18 OXFORD J. LEGAL STUD. (forthcoming 1998).

62. See, e.g., MODEL PENAL CODE § 211.2 (making reckless endangerment a misdemeanor); *id.* § 6.08 (setting the maximum term of imprisonment for misdemeanors at one year).

was altered from intentional, to reckless, to justified.

B. The Cases for Which the Theories of Justification Have Different Predictions

The first two test cases examine the community's views on cases critical to the deeds-reason debate. In scenario 7, presenting the unknowingly justified actor, the two theories predict starkly different results, and the deeds theory predictions are confirmed. The perpetrator in scenario 7 received a liability mean of 3.63 (just over four months). This result is not statistically different from the attempt contrast case in scenario 2, as the deeds theory predicts. Such liability is dramatically less than the 4.65 liability (about ten months) for the substantive offense that the reasons theory predicts. The deeds theory is clearly more consistent with community views on this matter.

Scenario 8 presents the case of the actor who knows of the justifying circumstances but who acts for other, non-justificatory motives. Recall that current law would give a complete defense in such a case, although logic would seem to suggest that a strict reasons theory would give no defense.⁶³ The liability mean is 2.10 (2.6 weeks), not the complete defense that current law would provide—only 7.8% of our subjects assigned a verdict of no liability—and not the full liability that the reasons theory logically would seem to suggest. It is consistent, however, with the deeds theory prediction of liability being somewhat less than that of the unknowingly justified actor. The actor is entitled to at least the discount given the unknowingly justified actor because his act is objectively justified; a greater harm is in fact avoided. Unlike the unknowingly justified actor, however, this actor's liability for attempt is less clear. His knowledge of the justifying circumstances may suggest to him that his conduct is not in fact criminal, thus he does not have the clear intention to violate the law that the unknowingly justified actor has. He might be viewed less as breaking the law than as taking advantage of it. In any case, the results again are consistent with the deeds view and inconsistent with the reasons view.

Turn next to the three scenarios in which the perpetrator, mistakenly believing that the town was in danger of a fire, set his neighbor's fields on fire to provide a firebreak. The reasonableness of the mistake varied across these three scenarios. Scenario 9 presents the case of a reasonable mistake as to a justification. Both

63. See *supra* notes 54-55 and accompanying text.

reasons and deeds theories would give a complete defense. Our subjects assigned liability of 1.10 (2.3 days), which was higher than we expected. Further, only 17.3% gave the complete defense verdict of "not guilty." On the other hand, 42.3% assigned liability but no punishment. Perhaps these subjects were concerned about the implications of giving a complete defense in a case in which the conduct in fact is not justified in an objective sense. There is reason to think that they should be concerned, as discussed in Section V.C. below, which presents our proposal to revise acquittal verdicts.

Scenarios 10 and 11 are cases of unreasonable mistakes as to a justification. In scenario 10, the actor honestly believes his conduct is justified but is mistaken, and his mistake is negligent rather than reasonable. That is, a reasonable person in the actor's situation would have been aware of a risk that the contemplated conduct was not justified. In scenario 11, the actor similarly honestly believes his conduct is justified and is similarly wrong. But here his mistake is more culpable; he is reckless. That is, he is aware of a risk that his conduct might not be justified, although, on balance, he concludes that it is justified. He disregards the risk (that the conduct might not be justified) and proceeds with the conduct. In other words, he makes a reckless mistake as to a justification.

As expected, the subjects imposed greater liability in these two cases than in the case of the reasonable mistake. Further, liability was greater in the case of greater culpability in making the mistake: 2.02 (two weeks) for the negligent mistake, 2.33 (about four weeks) for the reckless mistake. But this range of liability is considerably less than that imposed by current law's majority rule, which denies any defense or mitigation and imposes full substantive liability. In the context of this burning offense, current law's assignment of no defense would give the perpetrator ten months imprisonment, as imposed in scenario 1, not the two weeks and four weeks that scenarios 10 and 11, respectively, actually received. We conclude that the subjects would very much support recognition of mitigations for unreasonable mistakes as to a justification.⁶⁴ These results suggest

64. Recall the correlation in law between the reasons theory liability for the unknowingly justified actor and the all-or-nothing approach to mistake as to a justification (denying a mitigation for an unreasonable mistake as to a justification). See *supra* text accompanying notes 24-28. We reasoned that nothing in the reasons theory logically requires adherence to the all-or-nothing view. Our study results seem to confirm this speculation. The persons in our sample closest to the reasons theory—those that gave the smallest discounts to the unknowingly justified actor in scenario 7, as against the full liability of scenario 1—were neither significantly higher nor lower in their liability assignments in unreasonable mistake scenarios, 10 and 11, than the other subjects.

that the Model Penal Code's mitigations of this sort should not have been rejected so regularly by state criminal code drafters.⁶⁵

V. IMPLICATIONS FOR CRIMINAL LAW REFORM

With the results current in the reader's mind, we turn immediately to the question of how criminal codes might be modified in light of the community intuitions reported here. The results reported in the previous section confirm that much is right in current criminal law formulation. But the results also frequently challenge criminal code formulations and suggest a variety of criminal law reforms. These alterations would involve a reformulation of offense definitions, altered formulations of what count as defenses, and an altered system of trial verdicts.

A. *The Formulation and Grading of Offenses*

While the study was designed to examine defenses related to

65. For scenarios 9, 10, and 11, we added one more dependent measure that provides some illumination of our respondents' reactions to these scenarios, and perhaps of their reactions to other scenarios as well. In all of the scenarios in which the neighbor's fields actually burned, there is one individual who is obviously the innocent victim of events, and that of course is the neighbor who lost his crop. We asked respondents "what should be done" about the neighbor's loss. A number of respondents wrote that the neighbor should be compensated for his loss. (Recall that the instructions made clear that the loss was a significant one, moving the farm from a profit to a loss.) As the culpability for the loss in these cases altered, so too did the identity of the individuals who owed the neighbor compensation. When the perpetrator made a reasonable mistake, respondents thought that the town should share in the task of providing compensation, although the perpetrator, who had made the mistake, also owed compensation. When the mistake was described as negligent, and then reckless, more of the respondents thought the burden of providing compensation fell solely on the perpetrator, and not on the townspeople.

As we noted before, we had expected some judgments of no liability and instead found judgments of liability of a very minor sort. An impulse toward finding compensation for the victim may explain this difference. Testing this possibility, we added a question about compensation to scenario 6, in which setting fire to the neighbor's fields was completely justified because it prevented the oncoming fire from destroying the town, and we gave this scenario to six new respondents, who first responded to the compensation question and then to the liability question. The responses of these new subjects were quite revealing. As to compensation, all thought compensation was due and that the townsfolk should be the major source of it. Several suggested, as a more than token gesture of community, that the farmer who set the fire should give some of his crop to the neighbor. Whether they felt that this was "owed" or simply a wise and neighborly gesture on the farmer's part was not clear. After dealing with the compensation issue, respondents felt that the question of liability was moot. Pressed to answer, respondents generally decided that "not guilty" was the appropriate verdict. One suggested "no liability." What those verdicts suggest is that, having required the defendant to pay some compensation, the respondents thought that adding criminal liability would generate excessive total punishment.

justification, the results, specifically those in the contrast cases, reveal something about how the law ought to define offenses. In many respects, the results support the general approach of current law.

A comparison of scenarios 1 and 2 illustrates that, although the actor's conduct and culpable state of mind are the same in the two cases, the existence of a resulting harm matters greatly in assessing liability and punishment. This outcome confirms findings in our earlier studies.⁶⁶ It provides grounds to criticize the minority of state criminal codes that follow the Model Penal Code in grading attempts the same as the substantive offense.⁶⁷ The Code would have graded the offenses in scenarios 1 and 2 the same, but our subjects gave the attempt less than half of the punishment of the completed offense.

A comparison of the results in scenarios 1 and 3, as well as in scenarios 2 and 4, confirms current law's view that the actor's culpability level ought to have a large effect on degree of liability. Again, these findings are consistent with the findings in different contexts found in our previous studies.⁶⁸ Intentionally causing or trying to cause a harm is dramatically more blameworthy than being reckless as to causing the same harm. This norm supports current law's grading of offenses according to culpability level, as in homicide—intentional killing (murder) is graded more seriously than reckless killing (manslaughter).

But given the near universality of this rule, it also may be appropriate to criticize current law for limiting the use of culpability levels in grading to a few serious offenses. The results suggest that an actor's culpability level is significant in offenses far less serious than homicide. Even in the pure property offenses tested here, the effect of culpability level was dramatic. The intentional burning received more than seven times the punishment of the reckless burning.⁶⁹ Each point on our liability scale is equivalent to approximately one grade in a modern American criminal code. Thus, if the results here were followed, reckless burning would be graded two grades less than intentional burning. Current law, in contrast, typically grades intentional and reckless (and negligent) burning the same.⁷⁰

66. See ROBINSON & DARLEY, *supra* note 45, at studies 1, 17.

67. See, e.g., MODEL PENAL CODE § 5.05(1); CONN. GEN. STAT. ANN. § 53A-51 (West 1971); N.J. STAT. ANN. § 2C:5-4(a) (West 1978); 18 PA. CONS. STAT. ANN. § 905(a) (West 1972 & Supp. 1997).

68. See ROBINSON & DARLEY, *supra* note 45, at studies 8, 9, 16.

69. Compare scenario 1's ten months to scenario 3's six weeks.

70. See, e.g., MODEL PENAL CODE § 220.3(1)(a). Grading differences are based exclusively on the value of the property damaged. See *id.* § 220.3(2).

B. *The Formulation of Defenses*

The results in the contrast cases also tell us something about defenses. A comparison of the results in scenarios 1 and 6 shows clearly that a "lesser evils defense," as it is called, has strong intuitive support among the subjects. A plain language version of the defense might read like this: "You may act in a way that would otherwise be a crime if your conduct is necessary to avoid a more serious harm or evil than that caused by your conduct."⁷¹ About half of American jurisdictions do not yet recognize such a defense,⁷² and many of those that have recognized the defense in one case or another have declined to codify it, leaving its availability and formulation in doubt.⁷³ The strength of intuitive support for the defense suggests that it ought to be formally recognized through codification everywhere.

The five test cases offer the most important new information with implications for criminal law reform. As noted in the previous section, the results in scenario 7, as compared to scenarios 1 and 2, suggest that the unknowingly justified actor ought to be treated as an attempter, not as a perpetrator of a full substantive offense. He has in fact avoided a greater harm; there is no net harm. An objective formulation of justification defenses, like the one quoted in the paragraph above, would achieve this result, for it would give a justification defense to the unknowingly justified actor, who would then be liable only for attempt under a provision like the Model Penal Code's: "A person is guilty of an attempt to commit a crime if, acting with the kind of culpability otherwise required for commission of the crime, he purposely engages in conduct which would constitute the crime if the attendant circumstances were as he believes them to be"⁷⁴ Under the circumstances as the unknowingly justified actor believes them to be, he is committing the crime; hence, he is liable for an attempt to commit the crime.

In addition to justification defenses objectively defined, the law must provide a defense provision governing mistake as to a justification that would give a defense to the actor who mistakenly believes her conduct is justified. The results of scenario 8, as compared to scenario 6, suggest that such a provision should be

71. Paul H. Robinson et al., *Making Criminal Codes Functional: A Code of Conduct and a Code of Adjudication*, 86 J. CRIM. L. & CRIMINOLOGY 304 app. A, § 67, at 344 (1996).

72. See 2 ROBINSON, *supra* note 3, § 124, at 45 n.1 (citing authorities).

73. See 2 *id.*

74. MODEL PENAL CODE § 5.01(1)(a).

formulated to give a complete defense only if the actor acts for the justificatory purpose; it ought not be enough that she simply knows of the justifying circumstances, if these were not her reason for acting. As noted above, only 7.8% of our subjects gave a defense to an actor who knew of the justifying circumstances but acted out for a different, malevolent motive. This result suggests that a complete defense ought to be given not when an actor "*believes* her conduct is necessary to avoid a threatened greater harm," but only when she engages in the offense conduct "*in order to* avoid a threatened greater harm."

The question remains, however, whether such an increased defense requirement has practical utility. It is not impossible, but certainly difficult, for a court to know whether an actor acts for the proper purpose. An actor who knows of the justifying circumstances generally would have little difficulty persuading a court that those circumstances are the source of her motivation. Further, the case in which the justificatory purpose is not present, as in scenario 8, will be rare. All things considered, it may not be worth the trouble to have the defense formulation distinguish between purpose and simple belief.

The real dispute in formulating the mistake-as-to-a-justification defense is whether to allow a mitigation for an unreasonable mistake. As noted above, the results in scenarios 10 and 11 show that our subjects give a significant mitigation in such cases from the full liability given in scenario 1. The current law's majority rule, then, is badly out of step with our subjects' views. In the case of a reckless mistake, our subjects would give one-tenth the liability given for the full offense.⁷⁵ For a negligent mistake, our subjects would give one-twentieth of that for the full offense.⁷⁶ These are substantial mitigations, in cases in which current law commonly gives none.

How might code drafters incorporate this mitigation approach into defenses for mistake as to a justification? The basic defense might read something like the following: "An actor is excused for her conduct constituting an offense if her conduct would be justified had the attendant circumstances been as she believed them to be." The effect of this provision would be to provide a mistake-as-to-a-justification excuse to a person who honestly believed her conduct to be justified. Another provision would then impose liability upon those actors whose mistakes were culpable, varying the level of

75. Compare results of scenarios 1 and 11.

76. Compare results of scenarios 1 and 10.

liability with the level of culpability of the mistake.⁷⁷ Such a provision might provide: "When an actor is reckless or negligent in assessing the circumstances that justify her conduct, the mistake-as-to-a-justification excuse [quoted above] is not available for an offense for which recklessness or negligence, as the case may be, is sufficient to establish liability."⁷⁸ Thus, a person who intentionally kills another believing that such killing is justified but who is reckless in having such a belief, would be liable only for reckless homicide (manslaughter), not intentional homicide (murder). A person who is negligent in so believing would be liable only for negligent homicide.⁷⁹

C. *Reform of Acquittal Verdict Forms*

Recall the peculiar results in scenario 9, in which the actor makes an entirely reasonable mistake, is blameless, and few subjects imposed any significant punishment, yet only 17.3% gave the actor a defense. If the subjects thought that no punishment is appropriate, why would they impose liability?

One argument made in support of the deeds theory is that it helps make a distinction that is important to effective operation of criminal justice: the distinction between (1) conduct not punished because it is the right thing to do, it avoids a greater harm, and we would want it to be performed under similar circumstances in the future, and (2) conduct not punished because, while it is wrong, it does not avoid a greater harm, and we would not want it performed in similar circumstances in the future, the actor in fact is blameless for performing the wrongful conduct. Recall from the introduction of this Article that this is the classic distinction between a justification and an excuse. The deeds theory allows this distinction to be made manifest by distinguishing cases of mistake as to a

77. This is the structural approach of the Model Penal Code. See MODEL PENAL CODE § 3.09(2).

78. This language is modeled after Model Penal Code § 3.09(2). See *id.* It suffers from a number of technical problems that are beyond the scope of this Article. See ROBINSON, *supra* note 1, at 463-64 (discussing these technical problems).

79. Unfortunately, this approach to drafting a mitigation provision is dependent upon the criminal code having different culpability levels for most offenses. That is a suggestion that we urge above, but it is not true of most modern criminal codes. When no lesser grade for a lower culpability existed, the actor would get a complete defense under this approach, even for a reckless mistake. That would be a very undesirable result. Another approach, not dependent on the proper structuring of offense definitions, would give a set mitigation (for example, one offense grade—for a reckless mistake) and a greater mitigation (two offense grades) for a negligent mistake.

justification from cases of objective justification, labeling the former excuses and only the latter justifications.

Under this approach, an actor acquitted under a justification defense provides an example to others of conduct that they are free to repeat in similar circumstances in the future. An actor excused under a mistake-as-to-a-justification excuse, in contrast, provides an example to others of conduct that they ought not perform in similar circumstances in the future. The actor is being acquitted *despite* her wrongful conduct. The reasons theory, by combining truly justified conduct with mistake as to a justification, terming both "justified," makes it impossible to make this distinction.

This tension between judging the *actor* and judging the *act* may well have influenced our subjects in scenario 9, in which 71.2% gave the actor essentially no punishment—either no liability, liability but no punishment, or one day imprisonment, which might have been seen as a symbolic gesture.⁸⁰ If no punishment is the strong majority view of the group, why did only 17.3% give the complete defense verdict of "not guilty," thus imposing no liability at all? They may well have been concerned about the precedential effect of such outright acquittals, the message that it would send to others. Would it be taken to weaken the prohibition against such burnings generally?

That is certainly a danger in a system like the current one, which does not distinguish between justified conduct, which the law is happy to have repeated by others in similar circumstances, and excused conduct, which the law does not want repeated. Both cases are acquitted under current practice with the same verdict, "not guilty." If the only choice available is "not guilty," with no justification or excuse distinction, jurors are likely to feel uncomfortable acquitting in cases of excuse, for fear of the ease with which the verdict can be misunderstood. On the other hand, they also would feel uncomfortable exposing a person they thought blameless to substantial punishment with a "guilty" verdict.⁸¹

We offered our subjects a way out of this dilemma, and they embraced it. They were offered a choice of "liability but no punishment," which gave them the opportunity to avoid punishing a blameless actor but also to condemn the conduct as something that

80. Of all the subjects, 86.5% fit into these three categories for the contrast case of actual justified burning.

81. See PAUL H. ROBINSON, *STRUCTURE AND FUNCTION IN CRIMINAL LAW* 146-48 (1997).

ought not be repeated.

While this approach is a useful research device and has revealed interesting information, it is not a solution to the practical problem in real life, for juries have little role in the sentencing process. Once they give a verdict of "liability" of any sort, it is for the court to sentence. The offense for which the jury convicts sets a statutory maximum above which the sentencing judge cannot go, but, unless they give a defense, they cannot otherwise assure that the actor will get no punishment or only symbolic punishment, as our study allows its subjects. Further, the "liability but no punishment" option has the disadvantage of imposing liability on a blameless defendant. In the real world, where criminal conviction can bring moral condemnation and stigmatization, as well as other collateral disadvantages in jobs, licensing, and the like, such liability is unfair. Thus, even if jurors had sentencing power, this solution to the problem—liability with no punishment—has the effect of imposing the condemnation and stigma of criminal conviction on a blameless offender who does not deserve it. We force jurors between the two bad choices of doing injustice or undermining the prohibition against such conduct in the future.

The better resolution is to recognize formally distinct acquittal verdicts of "justified" and "excused," in which the former approves of the actor's conduct and the latter disapproves of it. An objectively justified actor receives a verdict of "justified," thereby approving of the conduct, while the actor who mistakenly believes she is justified is "excused," thereby disapproving of the conduct. One of us has elsewhere offered the details for such a verdict system.⁸²

VI. CONCLUSIONS

The results reported here illustrate the potential usefulness of social science research for illuminating issues concerning the formulation of criminal codes. If the code drafters are interested in knowing the moral intuitions of the community that the codes will govern, then the sort of careful, empirical social science study of the sort conducted here is the preferred mechanism for discovering those intuitions. Properly constructed studies can resolve competing claims among criminal law theorists over which theory or rule better accords with people's intuitions of justice. Here we conclude that the deeds theory of justification better accords with community views than does the reasons theory.

82. *See id.* at 204-07.

Further, such studies also generate specific reform proposals that would make the criminal justice system more just in its operation. The results in this study suggest a reformulation of many aspects of offense definitions, reforms to justification defenses and mitigations for mistake as to a justification, and reforms of jury acquittal verdicts. On our examination, and, we hope, on the reader's examination, these suggestions for reform appear coherent and are ones that adequately balance the competing considerations that govern judgments about these difficult cases in which the reason for and the outcome of the perpetrator's acts are in conflict. They are, in other words, reasonable candidates for code adoption. Reforms of this sort, that bring criminal law's principles of justice closer to those of the community, we argue, increase the law's moral credibility, which in turn increases its long-term effectiveness in crime control.

APPENDIX: STIMULUS STORIES

As we all know, in different circumstances, some actions can generate criminal liability while very similar others do not. Below are a number of cases in which a person sets a fire, for a number of different reasons, and under a number of different circumstances. Your task is to judge whether the act, in each of the specific stories that you read, should count as an offense generating criminal liability or not. If you decide that it is a offense that should generate liability, you will then assign it a punishment of whatever magnitude makes sense to you, or you may decide that even though it is a criminal act, you want to assign it no punishment.

Here is how you will register your judgments. You will always make your judgment by responding to the scale that we furnish below each case. Glance at the sample scale just below this paragraph now. After you read a specific scenario, circle "N" if you think the person has committed something that ordinarily would be considered a crime, but he has an acceptable justification for what he did and so should get no criminal liability. Circle "O" if you think the person has done something that generates criminal liability but should not receive any punishment. Otherwise choose a sentence from the other options. Work through the set of cases, giving us your opinions—there are no right answers. Take as much time as you need to go through the set of cases. (The numbers in front of each scenario are random, and simply tell us the source of the scenario. Ignore them.)

Background Information for all of the scenarios.

Jake is a farmer who has already harvested his corn crop. His neighbor has not done so, so his three acres of corn is still in the fields. The corn crop makes the difference, for these farmers, between having a profitable season because they have winter feed for their animals, or going into debt.

Running around several sides of Jake and his neighbor's fields are dirt roads. Jake's farm and his neighbor's farm are on a neck of land that stretches out into a lake. Out on the end of the neck of land is the local town. Jake's neighbor's fields cut the town off from the mainland, but Jake's fields do not. The following map shows you this layout.

[See FIGURE 1.]

Now read the stories. Please circle the rating that corresponds with YOUR OPINION about what the appropriate sentence (if any) should be for Jake in each case. These cases will differ slightly, so it

is important that you read the entire case before making a judgment about sentencing. If you think that the difference between two scenarios is important, you should assign different amounts of punishment to Jake as a result of that difference. But you may find some differences between scenarios to be unimportant, in terms of the amount of punishment, and it is quite all right if you assign them the same amount of punishment.

Some people have trouble thinking about punishments in terms of prison sentences. Our real question to you is what punishment Jake deserves for the act he committed, using the scale as a vehicle to express your beliefs. So you may want to think about the amount of punishment you think the act deserves as equivalent to a prison sentence of a particular length, and then assign that length sentence. For example, you may think a two-week prison sentence is equivalent to a \$10,000 fine, and hence circle a "2" to indicate that relative amount of punishment.

After reading and assigning a sentence to a later case, you might want to change your punishment ratings of one or more previous cases. You are free to do so. Remember, we are interested in knowing the liability and sentence YOU THINK SHOULD BE ASSIGNED in each case: there are no right and wrong answers and your responses will be kept completely confidential.

1. Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, and there is no wind, it is clear that the burning will create no danger beyond that of destroying the corn. The fire destroys the entire crop.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

2. Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, and there is no wind, it is clear that the burning will create no danger beyond that of destroying the corn. Just as Jake sets the fire, the neighbor

unexpectedly returns and puts it out before it does any harm.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

3. Like all of the local farmers, Jake routinely piles dry cornhusks near where they are cut and eventually burns them. Jake has one such pile near his neighbor's fields. Jake wants to get the pile burned quickly; the previous year he waited and the pile got soaked by rain. He is aware that high winds are forecast for today; winds that create a real risk that his fire will jump the gap between his trash pile and his neighbor's corn fields. Despite this danger, Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. The winds come and the fire jumps to his neighbor's crop. The fire destroys the entire crop.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

4. Like all of the local farmers, Jake routinely piles dry cornhusks near where they are cut and eventually burns them. Jake has one such pile near his neighbor's fields. Jake wants to get the pile burned quickly; the previous year he waited and the pile got soaked by rain. He is aware that high winds are forecast for today; winds that create a real risk that his fire will jump the gap between his trash pile and his neighbor's corn fields. Despite this danger, Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. The winds come but, Jake is lucky, the fire does not jump to his neighbor's crop.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

5. Like all of the local farmers, Jake routinely piles dry

cornhusks near where they are cut and eventually burns them. Jake has one such pile near his neighbor's fields. Jake wants to get the pile burned quickly; the previous year he waited and the pile got soaked by rain. He is aware that high winds are forecast for today; winds that create a real risk that his fire will jump the gap between his trash pile and his neighbor's corn fields. Despite this danger, Jake burns his trash pile, hoping the fire will not jump to his neighbor's crop. It turns out that the weather forecast was in error about the wind. Jake's burning never creates any danger to his neighbor's field.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

6. Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. He can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire crop. Because of his quick work the town and its inhabitants are saved.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

7. Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. When he knows that his neighbor is away, he sets fire to the neighbor's three acres of corn. Because the field is bounded on all sides by dirt roads, it is clear that the burning will create no danger beyond that of destroying the neighbor's corn. The fire destroys the entire crop. *Unbeknown to Jake*, lightning has started a fire upwind from his and his neighbor's fields and the local town and the fire is burning toward the town and endangering the people who live there. His burning the field creates

a firebreak: The town and its inhabitants are saved.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprisonment	death
	but no punish.											

8. Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. He can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake has no interest in saving the town; the townspeople have always been unfriendly to him. Further, Jake is angry with his neighbor over a dispute about use of water from a creek that the two share. He decides to use the fire as an excuse to burn his neighbor's corn crop. Without asking his neighbor for permission, he burns the fields. The fire destroys the entire crop. Because of his quick work the town and its inhabitants are saved.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprisonment	death
	but no punish.											

9. Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. In the past, Citizen's Band radio reports have often been true, but also often false. Jake stops two cars that are racing into town, and both confirm that "there is a big, out of control fire, heading this way." (Any reasonable person would think there was a destructive fire coming.) Jake can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire

crop. It turns out that the radio report was in error. The smoke was from a controlled burn being done by a crew of local foresters and presented no danger to the town or any of the surrounding area.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life	death
	but no										imprison-	
	punish.										ment	

10. Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. In the past, Citizen's Band radio reports have often been true, but also often false. Jake doesn't think of this, and although a reasonable person would do so, Jake doesn't think to check on the truth of the report, but there is no doubt in his mind that it is a dangerous fire. Jake can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire crop. It turns out that the radio report was in error. The smoke was from a controlled burn being done by a crew of local foresters and presented no danger to the town or any of the surrounding area.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life	death
	but no										imprison-	
	punish.										ment	

11. Jake hears over his Citizen's Band radio that lightning has started a fire upwind from his and his neighbor's fields and the local town and that the fire is burning toward the town and endangering the people who live there. In the past, Citizen's Band radio reports have often been true, but also often false. Jake remembers this fact, and realizes there might not be a dangerous fire, but doesn't check on the truth of the report. He can see the smoke from the approaching fire and calculates that if he burns his neighbor's corn crop he can create a firebreak that will stop the fire. (Remember that Jake's own

field is not located where it could serve as a firebreak.) Jake knows that his neighbor is not available to ask for permission, and he burns the fields. The fire destroys the entire crop. It turns out that the radio report was in error. The smoke was from a controlled burn being done by a crew of local foresters and presented no danger to the town or any of the surrounding area.

N	0	1	2	3	4	5	6	7	8	9	10	11
no liab.	liab. but no punish.	1 day	2 wks	2 mo	6 mo	1 yr	3 yr	7 yr	15 yr	30 yr	life imprison- ment	death

