

The Insanity Defense and Game Theory: Reflections on *Texas v. Yates*

Emily A. Keram, MD

J Am Acad Psychiatry Law 30:470–3, 2002

On March 14, 2002, the murder trial of Andrea Yates, a woman with a psychotic illness who drowned her five children, ended in a guilty verdict and a life sentence. Ten days later, *A Beautiful Mind*, the movie about John Nash, a man who recovered from a psychotic illness, won the Academy Award for Best Motion Picture. In what might be called “a beautiful irony,” Nash’s Nobel Prize-winning work,¹ which revolutionized the understanding of rational human behavior during conflicts, serves as a useful framework for an exploration of the current state of the insanity defense in the United States. The result suggests that our adversarial system of justice may not provide an effective means of arriving at fair and meaningful insanity determinations. In addition to providing a basis for understanding how the adversarial process may lead to unjust insanity trial outcomes, Nash’s work also encompasses concepts helpful in creating a system that better facilitates thoughtful examination and understanding of the complex interaction of mental illness and behavior in specific criminal cases.

Game Theory

Game theory describes conflicts in which the outcome is dependent on the strategies employed by players who are both rational (not self-defeating) and who reason strategically (taking into account the other player’s likely behavior).² Prior to Nash’s involvement in the field, game theory was understood

only in relation to “noncooperative” (zero-sum or total-conflict) games—that is, conflicts in which every gain made by one player comes at the expense of the other. In a noncooperative game, the outcome for any player depends on both his or her actions and the independent actions of the other player. Noncooperative games, by their definition, end in a win for one side and a loss for the other. Because the goal of the game is monolithic—to win—there is no incentive for the opponents to consider factors other than those that facilitate the achievement of that goal. Each opponent formulates a strategy that he or she believes will most likely ensure success. At trial, the adversarial system of justice may therefore be analogized to a noncooperative game. The prosecution and defense deploy the strategy each believes will result in a jury verdict favorable to its side.

Interpersonal conflicts in the real world are rarely zero-sum games. More often, they resemble “cooperative games”—that is, conflicts that offer the possibility of an outcome involving benefit to both parties. Game theory, as originally conceived, is therefore limited in its applicability. John Nash’s contribution came in the creation of a theory of cooperative games that allows for an analysis of strategy in conflicts in which there is a possibility of mutual gain. Nash had the insight that in this type of conflict “the game would be solved when every player independently chooses his best response to the other player’s best strategies.”³ Nash’s theory, or Equilibrium, was soon applied to a variety of conflicts, from peace talks to the \$7 billion Federal Communications Commission auction of airwaves. Cooperative game theory envisions an outcome more complicated than a simple win or loss—an outcome in which the op-

Dr. Keram is Assistant Clinical Professor, Psychiatry and the Law Program, University of California San Francisco, San Francisco, CA. Address correspondence to: Emily A. Keram, MD, University of California San Francisco, 401 Parnassus Avenue, San Francisco, CA 94143. E-mail: emilyk@lppi.ucsf.edu

ponents work together to fashion a result that is acceptable to both sides.

The Insanity Defense and the Adversarial System: A Noncooperative Game

In the United States, the insanity defense is raised in the context of an adversarial system of justice. An insanity trial is analogous to a noncooperative (zero-sum) game. Every gain made by the prosecution is a loss to the defense and vice versa. The noncooperative (adversarial) nature of insanity defense trials intensifies the importance of determinations of responsibility and appropriate punishment, for these determinations are the stakes of the contest. There is nothing else to gain or lose.

The purpose of an insanity trial is to arrive at an accurate determination about an inherently subjective and ambiguous phenomenon: the influence of mental illness on the defendant's behavior. However, the politics, laws, and very nature of the adversarial system influence the conduct and outcome of the trials in ways that are unrelated to, and in fact may distract from, the purpose of the trial.

Elected Judicial Officials

The adversarial system is about winning and losing. Yet the concepts of winning and losing do not provide a sufficient basis for a thorough, free, and unbiased assessment of a potentially insane defendant. Concerns about victory and defeat extrinsic to the trial itself contaminate the process. For example, jurisdictions in which judges and the district attorney are elected create an inherent conflict between an incumbent's political self-interest, which aligns with public opinion, and the interests of justice, which, in any given case, may have little to do with popular will. A judge or a prosecutor who diverges too far from popular opinion is all too likely to have an opponent raise this tendency toward independent thought and action in the next election cycle.

Jurisdictional Variation of Insanity Standards and Interpretation

The jurisdictional definition of the insanity standard obviously influences trial outcome. Among the 46 states that continue to retain an insanity plea, the trend increasingly has been to narrow the scope of the defense.⁴ The more narrow the definition, the greater the need to educate the jury regarding its interpretation. For example, the Texas insanity stan-

dard, which represents the most restrictive variation of the M'Naghten rule, requires that, "at the time of the conduct charged, the actor, as a result of severe mental disease or defect, did not know that his conduct was wrong."⁵ The Texas Penal Code and case law are largely silent as to the meaning of wrongfulness, leaving to the jury the decision of whether to apply a moral or legal wrongfulness standard. In other words, in jurisdictions such as Texas, wrongfulness is in the eye of the beholder. When the death penalty is sought, the necessity of impaneling a death sentence-qualified jury (one in which the *voir dire* process has ensured that each juror seated can agree to the death sentence) allows the prosecutor more than the usual opportunity to influence the selection of beholders.

A closer examination of the Texas insanity standard raises the possibility that, in the absence of some interpretation beyond the concrete meaning of the words used in the penal code, the standard may be largely meaningless. Who does not know that killing is against the law? Whether a defendant kills a person by accident, in self-defense, with malice aforethought, or in a delusional state, he or she almost certainly knows that the police will take a serious interest in those actions. The Texas standard does not allow for an insanity plea to be based on a defendant's lack of knowledge of the nature and quality of his or her actions.

Jury's Lack of Awareness of Dispositional Options

An additional source of outcome bias arises in jurisdictions that prohibit jurors in insanity trials from learning about the laws governing the disposition of insanity acquittees. Those in favor of keeping this information from jurors argue that in insanity trials, as in all other criminal proceedings, the process of fact-finding should not be influenced by considerations of disposition. This argument is naïve at best. Americans appear to have an unquenchable fascination with crime and criminal trials. Witness the popularity of Court TV, television shows such as "Law and Order," and the media attention generated by high-profile criminal cases. There is a common cultural perception of the nature (that is, incarceration in a prison), if not the precise duration, of punishment for criminal conduct.

In contrast, Americans do not have a firm grasp of the laws regulating the disposition of insanity acquit-

tees and how they are applied in practice. There is a popular misconception that an insanity acquittee will probably be “out on the street” immediately after a verdict of not guilty by reason of insanity. More experienced devotees of real-life crime may have some awareness that treatment is an option for insanity acquittees. They may be unaware, however, that this occurs in a forensic setting, with release conditioned on a lack of dangerousness. Insanity juries tend to have an ill-formed and incorrect sense of the impact of a verdict of not guilty by reason of insanity on the public safety.

If there should be no distinction between juries in insanity trials and those in ordinary criminal trials with respect to the knowledge of possible dispositions, if the two juries are truly to be put on equal footing, then insanity trial jurors should know the likely range of dispositional options for insanity acquittees in all jurisdictions. These juries should know the specifics of how long insanity acquittees are confined or involved in conditional release programs. The danger of prohibiting insanity trial juries from knowing the dispositional options for insanity acquittees is that, in the absence of this knowledge, the jury may do its own assessment of dangerousness and choose incarceration as a means of avoiding what it perceives to be an unacceptable or overly ambiguous disposition.

Prosecution and Defense Strategies

The adversarial system necessarily creates a climate in which any tactic is used that can help to achieve the desired outcome. The adversarial nature of insanity trials therefore shapes the development of strategies by the prosecution and defense in ways that have more to do with ensuring victory than with arriving at an accurate assessment of the defendant’s mental state at the time of the offense. Although these strategies are most frequently expressed in the way in which each side portrays the defendant during the trial, there are other opportunities to introduce bias into the process. For example, both sides may posture for the media in an attempt to manipulate a result by biasing potential jurors. The defense may leak records of a lengthy history of psychiatric treatment, while the prosecution may provide shocking and disturbing details of the crime.

Perhaps the most problematic tactic that the adversarial process promotes with its emphasis on winning arises from the fact that death sentence-quali-

fied juries tend to be more conservative and less likely to deliver an insanity verdict. This increases the probability that the death penalty will be sought by the prosecution, even in cases in which it is not its ultimate objective.

The Insanity Defense and the Inquisitorial System: A Cooperative Game

In other countries, notably Australia, cases in which the insanity defense is raised are transferred from an adversarial to an inquisitorial system of justice, one in which the judge is charged with finding the truth. The judge identifies what evidence is needed to arrive at this determination and (theoretically) has the resources both to acquire this evidence and examine it fairly. In some jurisdictions, judges are assisted in the inquisitorial process by a panel of mental health professionals who share the responsibility of identifying and neutrally evaluating the evidence necessary to make an assessment of legal sanity.

Insanity defense determinations in inquisitorial systems may take on the characteristics of a cooperative game. As noted above, the Nash Equilibrium states that a cooperative game is solved when every player independently chooses the best response to the other player’s best strategies. When the goal of the proceeding is a determination of the truth as opposed to winning, when external political pressures are minimized, when the determination of legal insanity is made by someone who is aware of the dispositional options, the interests of the judge, prosecution, and defense may become less polarized, and more accurately reflect their true duty: the delivery of a just outcome. The “best strategy” of each participant may be surprisingly similar.

In an inquisitorial system, the opportunity exists for the proceeding to focus on understanding the defendant’s mental state at the time of the offense and on making an appropriate disposition. Winning and losing and the prescription of punishment may be less of a focus than they are in the adversarial system.

Future Directions

The verdict of not guilty by reason of insanity in the trial of John Hinckley heralded the reform of insanity defense laws in most states and in the Federal system.⁴ Recently, a constitutional challenge to one

of these resultant reforms has been successful at the state level. A Nevada Supreme Court decision struck down the state's abolition of the insanity defense, recognizing a criminal defendant's constitutional right to present evidence of mental state that negates *mens rea*, and preserving the concept of an exculpatory insanity defense.⁶

Perhaps the guilty verdict and life sentence in the insanity trial of Andrea Yates will provide impetus for further review of the insanity defense. In the current climate, local politics and jurisdictional differences in laws regulating insanity defense trials conspire to influence unduly the conduct and outcome of insanity trials.

Undertaking a new phase of insanity defense reform necessitates a re-examination of the fundamental concepts of criminal responsibility, punishment, and the adversarial system. The barriers to insanity defense reform in the United States are numerous. For example, one of the unspoken obstacles is the national ambivalence regarding whether punishment should be excluded as an objective for disposition of insanity acquittees. Americans are not comfortable with the concept that a determination of legal insanity may carry a presumption that the defendant does not deserve to be punished, regardless of how horrible the crime may have been.

In addition, the American right to a trial by jury may also present a challenge to meaningful reform. We may not be willing to create a system that ex-

cludes jury participation from insanity determinations. However, insanity trials present issues that are unique. The jury system should enhance the fairness of a trial by ensuring that the defendant is judged by a jury of his or her peers. How often does this happen in the case of a truly insane defendant? Jurors may consider themselves experts on human emotion and behavior by virtue of their life experience. The natural tendency to extrapolate from known experience would be hard to overcome in an insanity trial. It is unrealistic to expect that a jury can form a reasonable understanding of the complexities of mental illness and its effect on behavior in the few weeks or months of a trial.

Perhaps the biggest impediment to insanity defense reform is more fundamental. We may not be ready for a system of justice that allows for adherence to the values of compassion and justice and for their consistent application, regardless of the circumstances.

References

1. Nobel Seminar: The Work of John Nash in Game Theory. Les Prix Nobel 1994. Stockholm: Norstedts Tryckeri, 1995
2. Myerson RB: Game Theory: Analysis of Conflict. Cambridge, MA: Harvard University Press, 1991
3. Nasar S: A Beautiful Mind. New York: Simon & Schuster, 2001, pp 14-15
4. Practice Guideline: Forensic Psychiatric Evaluation of Defendants Raising the Insanity Defense. J Am Acad Psychiatry Law 30:(suppl):S6-7, 2002
5. Texas Penal Code § 8.01
6. Finger v. State, 27 P.3d 66 (Nev. Sup. Ct. 2001)