On the Duty to Protect: An Evolutionary Perspective

Thomas J. Rudegeair, MD, PhD; and Paul S. Appelbaum, MD

Psychotherapists' duty to protect potential victims from their patients' violence has evolved in recent years toward a narrower set of obligations. This reformulation of the duty appears to us to be consistent with a sociobiological analysis of the reasonableness of compelled altruism. Altruistic behavior (e.g., rescuing a potential victim) takes place rarely in the animal world, and even among humans usually occurs only in situations in which reciprocity is likely. The *Tarasoff*-like duty to protect violates this sociobiological rule by requiring therapists to place the interests of an unknown victim over a known patient, and even to subordinate their own interests to the victim's. This has never been a socially tenable position. Psychotherapists appear to have escaped from this situation by avoiding potentially dangerous patients. The changes in the duty to protect have mitigated this dilemma, by moving the duty in a direction consistent with the evolutionary theory of altruism.

In 1976, the California Supreme Court, in *Tarasoff v. Regents of the University of California*, ruled that therapists have a duty to protect the potential victims of their patients.[1] Concern for the safety of the public proved more compelling than misgivings about the disruptive effects of such a duty on the therapist-patient relationship. Some version of a therapist duty to protect has since been adopted in the majority of jurisdictions.

The duty to protect, however, has not been a static concept. Courts experimented with broadening the duty defined by the *Tarasoff* court by: removing the requirement that the potential victim be identifiable before the duty was invoked[2]; imposing liability even when the therapist-patient relationship was limited to the emergency setting[3]; extending the duty to require protection of those, other than the victim, who might foreseeably be harmed by the patient's violence[4]; and holding that the risk of property damage, as well as personal injury, may give rise to an obligation to act.[5]

At the same time, other courts offered restricted versions of the *Tarasoff* duty, including: requiring an overt threat before the duty was said to exist[6]; negating the duty when a victim already had reason to know of the possible danger[7]; and limiting the duty to patients over whom the therapist had a right to exert control, that is, who were committable.[8]

These alternative approaches to the duty to protect vied with one another

Dr. Rudegeair is Assistant Professor of Psychiatry, University of Massachusetts Medical School, where Dr. Appelbaum is A. F. Zeleznik Professor of Psychiatry and Director, Law and Psychiatry Program. Dr. Rudegeair is also staff psychiatrist, Worcester State Hospital. Address correspondence to Dr. Rudegeair at Worcester State Hospital, 305 Belmont St., Worcester, MA 01604.

for more than a decade for acceptance in the case law. Ultimately, however, legislatures took the initiative and identified the most reasonable version of the duty. In general, they have narrowed the duty to protect, more clearly defining and limiting when it is evoked (i.e., when a threat has been made to harm an identifiable victim) and what actions are required to discharge the duty (i.e., generally warning the victim and/or police, or acting to hospitalize the patient).[9]

There are a variety of ways to explain these developments. One can view them as a result of therapists' political success (although therapists have been a group notably lacking in political clout in the past), or as part of a general trend to restrict physicians' malpractice liability. Many psychotherapists probably believe that the legislative initiatives simply represent a recognition of the unreasonableness of the more expansive versions of the doctrine. This may, in fact, be the case, although to draw this conclusion is not to say in what way the original doctrine was unreasonable.

We suggest in this article that evolutionary theory, in particular that branch that has been denominated "sociobiology," has something useful to say about the unreasonableness of a broad version of the therapist's duty to protect. Although we cannot demonstrate that legislative actions to narrow the duty were in any conscious way based on a sociobiological analysis, the new statutes are certainly in keeping with the conclusions of this approach. This essay constitutes an initial attempt to employ evolutionary theory to provide insight into the inarticulate (and perhaps even unconscious) impulses that have pushed the duty to protect in the direction of a more focal obligation.

Altruism in Evolutionary Perspective

Few would dispute the assumption that human beings are a product of an evolutionary process, yet attempts to explore the implications of this are frequently met with suspicion.[10] Examining the behavioral propensities of a species, the work of ethologists or behavioral ecologists, is not a controversial pursuit for any species except *Homo sapiens*. The dangers that accompany attempts the explore "human nature," however, are legion. All too often, descriptions of how people "are" merely serve the theorist who is concerned with convincing an audience of how people "ought to" be. These diatribes, usually more political than scientific, provide skeptics with all of the evidence they need to dismiss the evolutionary perspective entirely. Unfortunately, politics often plays as important a role in these premature dismissals as it does in the generation of self-serving theories.

Over the past 20 years, several theorists have articulated a neo-Darwinian view of the evolution of social behavior[11] and others have developed models of human evolutionary history.[12] Our brief summaries of relevant theories that follow are not meant to be a substitute for their careful exploration.

Altruism is rare in the animal kingdom. Selfishness appears to be the preferred mode of social interaction. Behav-

iors that enhance the survival and reproductive success of the actor result in the transmission of that individual's genes into the next generation. Given this axiom of evolutionary theory, it is not surprising that behaviors that enhance the survival of another to the detriment of the actor (a reasonable operating definition of altruism) are rare. Generosity among animals is largely confined to close relatives (i.e., those individuals who share an enhanced proportion of genes in common with the actor). W. D. Hamilton[13] articulated the theory of "kin selection" and explained rather parsimoniously the vast majority of apparent exceptions to the overt selfishness expected from Darwin's formulation of the "survival of the fittest."

There is no doubt that kin-centered generosity occurs in humans; our treatment of our children, for example, is remarkably consistent with Hamilton's model. But is it equally apparent that we have extended our generosity beyond the limits of the nuclear, or even the extended, family. The entity of the "best friend", for example, illustrates our tendency to bond with and invest in members of our social group outside the family. Robert Trivers[14] has presented an elegant and persuasive model for the emergence of "altruism" between nonrelatives. In his model of "reciprocal altruism", he explores the evolutionary wisdom of an individual's selective assistance of group members despite immediate costs, when such acts increase the likelihood that others in the group will behave in kind toward him/her in the future. He concludes that before an individual will realize a selective advantage (i.e., the enhancement of his/her genetic contribution to the next generation) from the tendency to perform "altruistic" acts, s/he must confine such acts to a very select subset of recipients under very strict social conditions. For example, the recipient must be someone who is willing and able to reciprocate, and the act must not exact a cost greater than the likely benefits to be ultimately enjoyed when reciprocity occurs. It follows from this that individuals who are perceived as unreliable or ineffectual, and thus unlikely to reciprocate effectively, will not be popular objects of altruistic behavior, and that group members will work hard to avoid being labeled in this way (even to the point of deception).

This view of generosity as emerging from a subtle calculation of individual self-interests is viewed by many as cynical as well as inaccurate.[10] A common reaction is to say that it simply *doesn't* feel right. Most of us are convinced that we are often generous without the slightest thought of any potential benefits beyond the resultant satisfaction. In response to this criticism, Trivers postulates an elegant theory of "self-deception." Given that the altruist is perpetually vigilant for any indications of insincerity in a potential partner, selection has favored any advances in the efficient communication of sincerity. The signs of anxiety that accompany deliberate deception are a dead giveaway, however, and can only be controlled, Trivers argues, at the source. We look sincere only when we are sincere!

The successful deception of others is accomplished by the deception of the self—we believe in our basic altruistic motives and are appalled by any suggestion of selfishness. The implications of this concept of pervasive self-deception will not be explored further here, except to point out that none of us can be relied upon to accurately report on our own motives (a theorem with considerable support in cognitive psychology[15]).

The *Tarasoff II* Ruling

The *Tarasoff* case presented a disturbing scenario. A mentally ill man had murdered an innocent young woman after confiding to his therapist that he was entertaining this brutal plan. Although the therapist and his supervisors had contacted police in an effort to have the man confined, they failed to follow through on this strategy of protection. The victim's family felt that the therapists had been negligent in not warning them of the man's intentions, for by remaining silent they had deprived the family and the victim of an opportunity to respond to the threat.

The California Supreme Court, considering a challenge to the basis for the plaintiff's suit, was faced with a dilemma. Although therapists are obligated to respect the privacy of potentially dangerous patients, they are in a unique position to protect potential victims, if patients disclose violent intent. There was no shortage of advice to the court during the extended process in which it considered and reconsidered this issue. The American Psychiatric Association, in an amicus brief joined by other mental health groups,[16] argued that: (1) therapists are no better than laypersons at predicting future violence (and hence an expectation by the court that they could make such predictions would be unrealistic); (2) the violation of patient confidence that a duty to warn would mandate would disrupt therapy in a profound way, decreasing disclosure and therapists' capacity to intervene, and thereby increasing the risk of patient violence; and (3) when therapists were faced with the " . . . twin ethical responsibilities of preserving the patient's confidentiality and preventing any harm to his patient, the effect of imposing a duty to warn may well be to deny the class of persons seen as potentially violent all right to psychotherapy" (i.e., it would lead the therapist, when faced with an insoluble ethical conflict, to avoid such patients). Fleming and Maximov,[17] two lawyers, wrote a law review article that seemed to move the court more profoundly than the APA's effort. In it they argued that the impact of disclosure by therapists has an unpredictable effect on therapy, not the inevitably negative one predicted by the *amici*, and that, in the absence of data on the subject, they regarded it as an open question. Although clearly against the imposition of an enhanced duty to commit dangerous patients, they argued for a duty to warn potential victims, while eloquently articulating the double bind thereby constructed. They stated, "In sum, the therapist owes a legal duty not only to his patient, but also to his patient's would-be victim, and is subject in both respects to scrutiny by judge and jury."

Balancing the interests involved, the court concluded that " . . . the protective privilege ends where the public peril begins," and recognized a duty to protect third parties. The court argued that the existence of a "special relationship" between the therapist and the patient placed a burden on the therapist to protect third parties from the patient's violence. It placed the responsibility for the recognition of this implied bright line squarely with the therapist and, in so doing, appeared ready to accept the therapist's double bind.

Tarasoff, Altruism, and Evolution

In our view, the demands placed on therapists by the *Tarasoff* court are a recent manifestation of an ancient and relentless societal effort to control individual behaviors that appear to be detrimental to the well-being of the group. Tension between the interests of the individual and those of the group is an inevitable product of the evolution of social behavior in any species, but has been particularly so throughout human history due to the remarkable degree of interdependence required for survival in a hostile environment. Under ancestral conditions of small group size and stable composition, control of individual behavior was accomplished without reliance on laws or anonymous authority. Study of modern foraging societies (living under conditions most closely approximating the ancestral) reveals an informal control, exerted by implicit expectations and enforced by group sanctions.[18] The ancestral view of social responsibility was clearly hierarchical, with relatives expecting the bulk of an individual's generosity in proportion to their degree of relatedness. Nonrelatives received little, unless they stood in some "special relationship" to an individual. They may have been related by marriage, for example, and gained the benefits of spousal kin-based affiliations or, more relevant to our discussion, they may have had a reciprocally altruistic relationship with a resourced individual and thereby had a claim to some measure of beneficence.

But strangers had no such claim. Ancestrally, this status was confined to nongroup members with remote allegiances and dubious characteristics. Such people were owed no obligations and likely were viewed with considerable suspicion (at those rare times when they were viewed at all). With the emergence of agriculture, population size increased and, with it, anonymity between close neighbors. This compelled the codification of laws and explicit consideration of the rules governing interactions between strangers. The Western common law tradition of punishing malfeasance while tolerating nonfeasance between strangers appears accurately to reflect the nature of the ancestral relationship. Societal interests required a significant infringement on the tendency of individuals to exploit strangers (malfeasance) but stopped short of unrealistically demanding active altruism. Yet the "special relationship" identified by the *Tarasoff* court exists between the therapist and his/her patient, not with the patient's potential victim. The therapist

feels an alliance with the patient, a form of reciprocity, but has no such feeling for the stranger s/he now must protect at any cost.

The *Tarasoff*-like duty, however, violates the maxim that closer relationships have a greater claim on altruism than more distant ones. True, the special relationship identified by the court was deemed to exist between the therapist and his/her patient, with whom some reciprocity might be expected. But the duty was imposed on the therapist to protect the victim, often at the expense of the patient, putting the stranger—from whom reciprocation is unlikely—ahead of someone who ordinarily would stand higher in the altruistic hierarchy. Even more important, the duty to protect places the interests of the victim above those of the therapist, in a situation in which reciprocity is unlikely.

Therapists will not long remain in this sociobiologically untenable position. As the Fleming and Maximov quote so clearly articulated, the therapist of a potentially dangerous patient is subject to loss of a patient, or even threat of harm, if s/he moves to protect a potential victim, and vulnerable to litigation regardless of the course of action selected. A prudent therapist under this no-win set of conditions will refuse to play the game. Anecdotal data suggest that this has, in fact, occurred. We know many therapists who report avoiding "dangerous" patients, in part because of their concern over the potential for liability. Empirical data on point are less clear. Givelber *et al.*,[19] in a questionnaire survey of therapists, asked them if *Tarasoff* had resulted in a reduction in the number of dangerous patients they were seeing. They concluded that predictions of desertion of dangerous patients (e.g., as in the APA *et al.*, amicus brief) were overblown. It does not appear, however, that this conclusion is warranted by their data. Beck's[20,21] more intimate studies of post-*Tarasoff* therapist experience also suggest that the disruptive effects of the duty may be manageable. Unfortunately, each of these studies relies on therapist responses to direct questioning. Even if therapists say that they are not affected by *Tarasoff*, the expectation that individuals necessarily will be in touch with their ambivalence about treating such patients under post-*Tarasoff* conditions (or will be forthcoming about it) is naive, given the human propensity for self-deception (14), as well as social pressures to provide "acceptable" answers.

Given that therapists are likely to be reluctant to describe any history of "abandonment" when asked on a questionnaire, the demonstration of this predicted outcome would require a more oblique research instrument. We suggest embedding questions on the impact of patient dangerousness and the "duty to protect" in a research instrument with an ostensibly different focus. One might, for example design a questionnaire exploring "termination" in therapy, an outcome seen as ubiquitous (inevitable in fact), mutual, and appropriate when properly negotiated. A focus on termination would be less likely to engender the unconscious evasion anticipated from an overt probe of therapist tend-

encies to "abandon" patients. The demonstration of enhanced therapist avoidance of potentially dangerous patients in jurisdictions operating under stringent *Tarasoff* criteria by such an instrument is predicted by the evolutionary model described here, and would constitute strong supporting evidence.

It is appropriate, of course, for the courts to be attempting to mould the behavior of private citizens, and the *Tarasoff* court's concern for the safety of the public in these violent times is understandable. But, as Levmore[22] points out, the proper manipulation by the court of rewards and punishments is critical for success. In this case, Levmore suggests, the courts have strayed from the tendency in the United States to refrain from either punishments (sticks) for failing to act altruistically or rewards (carrots) for doing so by imposing a formidable stick for nonrescue (i.e., failing to protect) while offering no carrots for successful rescue. Levmore concludes, as we have, that such an arrangement will result in the abandonment of "rescue spots" by a potential rescuer in order to avoid potential rescuees. Clearly, Levmore recognizes the same tendency of "human nature" toward the enhancement of self-interest in the face of excessive social demands.

Conclusions

Whether the evolution of the *Tarasoff*-like duty (for it does seem that laws, like living organisms, evolve) was propelled by a recognition of its incompatibility with deep-seated sociobiologic constraints is not ascertainable. The directions in which it has moved, however, are entirely compatible with what might be expected on the basis of the evolutionary theory of altruism described above. Furthermore, this analysis suggests that to the extent that some jurisdictions still cling to a more expansive model of the duty to protect, they are swimming against a powerful current of sociobiologic tendencies and their likelihood of success is slim.

References

1. Tarasoff v. Regents of the University of California, 551 P. 2d 334 (Cal. 1976)
2. Lipari v. Sears Roebuck, 497 F. Supp. 185 (D. Neb. 1980)
3. Jablonski v. United States, 712 F. 2d 391 (9th Cir. 1983)
4. Hedlund v. Superior Court of Orange County, 669 p. 2d 41 (Cal. 1983)
5. Peck v. Counseling Service of Addison County, 499 A. 2d 422 (Vt. 1985)
6. Brady v. Hopper, 570 F. Supp. 1333 (D. Colo. 1983)
7. Votteler v. Heltsley, 327 N.W. 2d 759 (Iowa 1982)
8. Hasenei v. United States, 541 F. Supp. 999 (D. Md. 1982)
9. Appelbaum PS, Zonana H, Bonnie R, Roth L: Statutory approaches to limiting psychiatrists' liability for their patients' violent acts. Am J Psychiatry 146: 821–8, 1989
10. Lewontin R, Rose S, Kamin L: Not In Our Genes. New York, Pantheon, 1984
11. Dawkins R: The Selfish Gene. Oxford, England: Oxford University Press, 1976
12. Leaky R, Lewin R: Origins. New York, Dutton, 1977
13. Hamilton WD: The genetical evolution of social behavior. J Theoretical Biol 7:1–52, 1964
14. Trivers R: The evolution of reciprocal altruism. Q Rev Biol 46:35–57, 1971
15. Nisbett R, Wilson T: Telling more than we know: verbal reports of mental processes. Psychol Rev 84:231–59, 1977
16. Brief in Support of Respondent's Petition for Rehearing *Tarasoff v. Regents of University of California*, 551 P. 2d 334 (Cal. 1976)
17. Fleming J, Maximov B: The patient or his

victim: the therapist's dilemma. Cal L Rev 62:1025–68, 1974
18. Lee R, DeVore I: Man the Hunter. Chicago, Aldine, 1968
19. Givelber D, Bowers W, Blitch C: *Tarasoff*, myth and reality: an empirical study. Wis L Rev 1984:443–97, 1984
20. Beck J: When the patient threatens violence: an empirical study of clinical practice after *Tarasoff*. Bull Am Acad Psychiatry Law 10:189–201, 1982
21. Beck J: Violent patients and the *Tarasoff* duty in private psychiatric practice. J Psychiatry Law 13:361–76, 1985
22. Levmore S: Waiting for rescue: an essay on the evolution and incentive structure of the law of affirmative obligations. Va L Rev 72:879–941, 1986