

Harmful Thoughts

Meir Dan-Cohen*

If there is one article of faith all liberals share, it is quite definitely an aversion to thought control. This aversion is closely linked to another fundamental liberal tenet, the harm principle: Law should concern itself exclusively with preventing harms that people may inflict on each other; people's thoughts ought to remain their own affair. Although the inviolability of thoughts and the harm principle reinforce each other, the connection between them is not as simple as may appear at first sight.

In a well-known article,¹ Professor Herbert Morris probes one dimension of this connection: the interest that criminal law may take in thoughts consistently with the harm principle. From the fact that thoughts are not harmful by themselves, Morris points out, it simply does not follow that the harm principle will not authorize their legal suppression. After all, the law commonly punishes conduct that is not by itself harmful, when the conduct increases the likelihood that harm will occur. One example, given by Morris, is of reckless behavior that may be punished even when it does not result in any harm. Precautionary prohibitions, such as gun control, are another. Finally, and most pertinently, attempts are commonly punishable in the absence of harm. The same logic that licenses legal intervention before the criminal act is fully consummated would also seem to apply to the even earlier stage in the criminal sequence at which the criminal intention is formed. To be sure, intentions present special

* I'd like to thank my friends and colleagues, Paul Hoerber, Sandy Kadish, Chris Kutz, Robert Post, Sam Scheffler, and Jan Vetter, for their helpful comments.

¹ "Punishment for Thoughts," in Herbert Morris, On Guilt and Innocence (Berkeley: University of California Press, 1976), 1-29.

epistemological problems. But here they are shielded from punishment by their inaccessibility rather than by their harmlessness. Conduct is still called for, not by the harm principle, but rather as an evidentiary imperative. The difference is not only of philosophical but also of practical significance: for the latter purpose mere verbal behavior such as a confession, will do. The conclusion that the harm principle might be in principle consistent with the punishment of thoughts is striking, perhaps scandalous, and Professor Morris proceeds to hedge and qualify this conclusion in many subtle and interesting ways.² My aim in this article is not, however, to examine Professor Morris's conclusions, but rather his basic premise: that unless acted upon, thoughts by themselves can have no extra-personal effects, so that if thoughts are of public, specifically legal concern, it is only because of their link to harmful action.³

The same premise underlies other legal areas as well. The thoughts Morris focuses on are for the most part intentions, and the only way they supposedly can affect others is by being carried out. But intentions are not the only kind of potentially harmful thoughts, and executing intentions not the only way by which that potential can be realized. Think of areas such as defamation and hate speech. Here the potentially harmful states of mind are not intentions, but rather beliefs, opinions, and emotions, and the harmful behavior in which they issue is typically speech that expresses those states of mind. Speech, accordingly, becomes in these cases the agent of harm, its regulation licensed by the harm principle. Of course, regulation of speech conflicts with another liberal commitment, to freedom of speech, leading to

² For some further reflections on this theme see Douglas Husak, Philosophy of Criminal Law (Totowa: Rowman & Littlefield, 1987), pp. 99–97, 103–05.

³ For an explicit statement of this assumption, see, e.g., R.A. Duff, Criminal Attempts (Oxford: Clarendon press, 1996), pp. 313–14: “even practical thought [deliberation, intention-formation, choices] does not, by itself, impinge on the world: it must be translated into action; and we can say that it is by action (as distinct from mere thought) that agents impinge on or make a difference to the world.”

a wide ranging debate concerning the permissible constraints on harmful expression. But all parties to the debate seem to take for granted that the potential for harmfulness of the states of mind in question is only realized when those states of mind are publicly expressed.⁴

In this paper I question this assumption. Though the assumption is quite entrenched, part of the very “grammar” of mental states, I argue that it rests on untenable beliefs and is inconsistent with other shared judgments and usages in neighboring linguistic fields. I conclude that thoughts can have extra-personal effects, including negative ones, all by themselves, unaided by action of any kind, executory or expressive. The argument is laid out in sections 1–4. Section 5 examines some of its normative implications, specifically with regard to the controversy over the legal treatment of hate speech. The point I raise has broader philosophical ramifications as well: it bears on the limits of individualism by suggesting what I believe is a novel, I call it ‘active,’ form of externalism about the mind. I deal with these issues in section 6.

Before I proceed, it is worth pointing out that my main claim is not as striking as may appear at first sight. First, we are often intensely and legitimately interested in the states of mind of other people quite apart from their speech and behavior: Does she love me? Do these people think well of me? Does the physician think I am in good health? Indeed, even when such queries are ostensibly answered by the respective people’s reassuring speech and behavior we may not be quite satisfied: we sometimes worry about the sincerity of their avowals, thus confirming that it is the others’ thoughts themselves, rather than their expression, that is of primary significance to us. To be sure, our interest in others’ thoughts does

⁴ See, for example, Joseph Tussman, Government and the Mind (New York: Oxford University Press, 1977), p. 88.

not show that those thoughts by themselves affect us: I may be interested in the preservation of the rain forest or in whether there is life on Mars, and yet logging operations in the one case or protozoan stirrings, or their absence, in the other will not by themselves be thought to affect me one way or the other. Still, the reminder that an interest in others' thoughts is pervasive makes the claim about the extra-personal effects of thoughts seem less far-fetched. Secondly, other writers have maintained that people's thoughts can be evil, thereby making the world a worse place, and that thoughts can wrong others and perhaps even harm them.⁵ As they stand, these judgments appear dark and mysterious, but they express an intuition that even to a secular morality some thoughts can be by themselves of extra-personal interest and concern.

1. THE ARGUMENT FROM KNOWLEDGE

The most likely basis for the assumption that one person's thoughts by themselves cannot affect another is probably the alleged inaccessibility of the other's mind. Unless the other person manifests her thoughts by speech or action one cannot know them, the argument goes, and what one does not know cannot affect one. In syllogistic form:

- (1) What one does not know cannot affect one;
- (2) in the absence of any external manifestation, another's thoughts are not knowable; hence,
- (3) in the absence of any external manifestation, another's thoughts cannot affect one.

⁵ See e.g. Joel Feinberg, Harmless Wrongdoing (New York: Oxford U.P., 1988), pp. 23–24; D. Husak, Philosophy of Criminal Law, p. 105; R.A. Duff, Criminal Attempts, p. 314.

Call this the argument from knowledge. As it stands the argument is obviously fallacious, since the first premise is clearly untenable: Unbeknownst to me, there may be at present a cavity in my tooth, causing decay that will eventually result in acute pain. Surely the cavity is “affecting me” even though I don’t know about it or, indeed, even if I never realize that a cavity will have been the reason for my toothache. The same could be said of thoughts as well. If you punch me in the face, the intention that led you to do so is obviously harmful to me though I never learn what it was. To avoid this all-too-easy refutation, the first premise must be amended by adding the proviso “unless it has consequences that I do know,” and correspondingly the second premise must preclude knowable consequences. But even with these amendments the premises will not withstand scrutiny.

Starting with the second, it is not always true that we cannot know through reliable inference another’s thoughts. If I see you standing in front of an elephant cage with your eyes wide open, in broad daylight, with the right orientation, and in the absence of any unusual circumstances, I will conclude, with practical certainty, that whatever else may be in your mind at that moment, there is an elephant-thought in it. I could also attain such knowledge by myself planting the thought in your mind by telling you, under ordinary, auspicious auditory conditions: “Think of an elephant.” Indeed, far from requiring interpersonal communication in order to be acquired, knowledge of others’ minds in circumstances such as these is a pre-condition of the acquisition of language and of the very possibility of successful communication.⁶ Minimal information about the other person concerning her eyesight in the one case

⁶ Cf. Willard Van Orman Quine, Word and Object (Cambridge, Mass.: M.I.T. Press, 1960), chs. 1–2.

and hearing ability in the other will normally suffice for the kind of rudimentary knowledge of her thoughts here involved. But the better we know someone, the more elaborate the thoughts we can confidently ascribe to her on the basis of the external circumstances we know her to be in. Occasionally we can be even better at divining another's state of mind than he himself is. Psychologists and parents are among those who tend to exaggerate and capitalize on this possibility.

Shaky as the second premise turns out to be, it is the first premise that seems to me the argument's weaker link. This premise conjures up a conception of humans as essentially psychological beings whose welfare and well-being is at bottom entirely a matter of the quality of their experiences. It is for this reason that things which do not eventually induce any experience in me cannot be said to have a bearing on me. Such a conception of the self has of course played a prominent role in moral philosophy as the springboard of utilitarianism, at least in Jeremy Bentham's classical version.⁷ One way, accordingly, in which we can challenge the first premise of the argument from knowledge is by recalling arguments made by utilitarianism's critics to discredit this underlying conception of the self. Let me just mention one well-known argument to that effect: Robert Nozick's thought experiment of the "experience machine."⁸ He imagines a device to which you can hook up via electrodes stuck in your brain that will provide you in the course of a lifetime with a constant stream of pleasurable experiences of whatever kind you choose in advance. Such a device is the fulfillment of the utilitarian's dream, and yet it would seem to strike most people as a nightmare. To those who respond to the prospect of an

⁷ Jeremy Bentham, The Principles of Morals and Legislation (New York: Hafner Press, 1948), p 1.

⁸ Robert Nozick, Anarchy, State, and Utopia (New York: Basic Books, 1974), pp. 42–5.

experience machine with horror rather than delight, the thought experiment demonstrates that there is more to our lives than the experiences they contain. We can formulate this conclusion in a way more congenial to our present purpose. Those aspects of our lives that the experience machine cannot in principle replicate are the ones in which we are vulnerable to being affected by things quite apart from the experiences, if any, those things induce.

The experience machine provides an indirect line of attack on the argument from knowledge by challenging a conception of the self and a related conception of value that seem to underpin its first premise. A more direct line of attack involves looking at some counter-examples: situations in which we strongly incline to say that one is affected, indeed harmed, by an event of which one does not know. Consider two legal cases first. In Breithaupt v. Abram⁹ a blood sample was taken at the police's behest from an unconscious driver suspected of drunk driving. Were his constitutional rights infringed? Though the Supreme Court divided on this issue, no one doubted that the needle prick, though administered to an insensate person and thus having no experiential effects, affected Breithaupt, thus triggering an investigation into the scope of his constitutional rights. The second case, People v. Minkowski,¹⁰ involves rape by deception. On numerous occasions, the defendant, a gynecologist, had sexual intercourse with his patients, unbeknownst to them. Though the physician was eventually found out, all would agree, I suppose, that the patients had been raped even before their suspicions arose, and whether or not they would have ever discovered the truth.

These cases are naturally seen as involving a violation of the respective victims' autonomy: it's

⁹ 352 U.S. 432 (1956)

¹⁰ 204 Cal.App.2nd 832; 23 Cal.Rptr. 92 (1962)

after all non-consensual sex in the one case, and an involuntary drawing of blood in the other. But appeal to the victims' autonomy begs the question here. People do not in general hold a veto power, not even a prima facie one, over actions they would not approve. To engage their autonomy, the action in question must first be said to affect them. But then, if the individuals don't even know about the actions, in what sense can they be said to be affected by them? On what ground is their assent required? A simple answer to these questions is readily available, though. We recognize that the respective victims of the two cases were affected, because their bodies were affected. A bodily intrusion by itself counts as affecting the person though no experience be involved. Despite its obviousness, the significance of this answer should not escape us. It implies a conception of human beings and what might be said to affect them that runs counter to what the argument from knowledge assumes. It does so by identifying an aspect of persons—their bodies—that can be the locus of effects on them without their knowing it and without any subsequent experiences. By generalizing these examples we get an expanded notion of effects and potentially of harm. Just as experiential effects impinge on a person's mind, other effects impinge on other aspects or constituents of the person, such as her body. Moreover, though embodiment is an obvious aspect or property of the self, thus providing a natural arena for non-experiential effects, it need not be the only one. What counts as affecting the person, and as potentially harming her, must ultimately depend on the picture of the self we have and on what other aspects or constituents it might be thought to comprise.

2. THE ARGUMENT FROM CAUSATION

By demonstrating that we can be affected without being aware of it, the examples I've just considered help refute the argument from knowledge. But they also suggest an alternative basis for the belief that thoughts cannot have extra-personal effects. These examples fit into a familiar paradigm, namely a common-sense causal paradigm: X can be said to affect me insofar as there is a physical causal link between X and an aspect or constituent of myself. These examples fit this paradigm particularly well, since the constituent they concern is the body, and X is a physical object or event impacting the body no differently than the proverbial billiard balls hitting against each other. Indeed, once we discover this causal paradigm at work in the cases just considered, instances of experiential effect can be assimilated to it as well: our experiences are simply one kind of result that external factors can cause in us. Our psychological make up is in these cases the aspect of ourselves in which the effect takes place, and communication the typical form that the causal chain assumes. But now the belief that uncommunicated thoughts cannot by themselves affect us will seem to rest on a different and if anything firmer foundation than that suggested by the argument from knowledge. The reason that thoughts cannot affect us, and a fortiori cannot harm us, it will now seem, lies in their lack of direct causal efficacy outside the agent. How can one person's thoughts affect another if they can't causally engage with her?

We reach accordingly the second argument for the belief that a person's thoughts cannot directly affect anyone else, the argument from causation:

- (1) For X to affect A, X must have some causal efficacy with regard to A;

(2) thoughts by themselves are causally inert with regard to other people;

hence,

(3) a person's thoughts by themselves cannot affect other people.

Is this argument sound? The second premise seems to me unassailable. This is so quite trivially, since to speak of a thought “by itself,” as it exists in a person's mind, is to cut it off simply by stipulation from any causal chain that could possibly lead to another person. The first premise may seem to be just as firm. Doesn't “affecting” entail, analytically, causal efficacy on the part of whatever does the affecting? But as some familiar examples demonstrate, the first premise is nonetheless false, in that we do commonly judge a person to be affected by a thing or an event, even though no causal chain, as ordinarily understood, runs from the one to the other. Consider the father whose child is killed, unbeknownst to him, say, in a car accident.¹¹ It would be natural to report in these circumstances that something terrible happened to the father. But why? How can a car accident that resulted in one person's death be said to affect another person, who, let us assume, lives in a faraway land, and doesn't even know about it? Clearly, unlike the cases of unknown direct bodily impact I mentioned earlier, there is no causal relation between the accident and the father in this case. Is this then a case of some mysterious, distal causation?¹² Of course, the father has a deep concern for and an abiding interest in

¹¹ The example is discussed by George Pitcher in “The Misfortunes of the Dead,” *Am. Phil. Q.* and by Joel Feinberg, *Harm to Others* (New York: Oxford University Press, 1984), pp. 90–1

¹² Pitcher, *ibid.*, diagnoses the problem this case poses in similar terms. I propose a different solution from his, though the two, I think, are consistent. Mine, however, leads more directly to the claim I seek to establish concerning the extra-personal effects of thoughts.

his son's welfare, quite independently of what he learns about it.¹³ But as we saw earlier, such concern and interest are insufficient to ground the judgment that the son's death affects the father: recall the earlier point about a corresponding concern for the rain forest one might have. It may be tempting to distinguish the case of the father from that of the rain forest by invoking here the notion of "having an interest": though I may be avidly interested in remote deforestation, I cannot be said to have an interest in it whereas the father does have an interest in his son's longevity.¹⁴ The distinction is sound, but not very illuminating. In the order of explanation, "being affected by" seems to be a more basic and broader notion than that of "having an interest in." To buttress in a non-question-begging way the judgment that something terrible happened to the father when the son was killed, we need some plausible account of the possibility of one thing affecting another without a causal chain that starts from the one and terminates in an alteration in the latter.

3. EFFECTS AND RELATIONAL PROPERTIES

The Eiffel Tower is 984 feet tall and made of steel; it is in Paris and is illuminated by floodlights every night; it is the tallest structure in Paris and, some believe, the ugliest.¹⁵ These salient properties of

¹³ For a discussion of "vicarious harms" along such lines see Feinberg, Harm to Others, at pp. 70–9.

¹⁴ Cf. Feinberg, *ibid.*

¹⁵ Guy de Maupassant famously stated that the tower was his favorite lunch place in Paris, being the one place from which the tower could not be seen; he is also said to have left Paris permanently to avoid looking at the tower.

the Eiffel Tower fall into three categories: the first two properties are intrinsic properties, since describing them does not mention any other objects; whereas the other four are relational, since their descriptions do mention other objects. Among the relational properties, in the last two the relation involves a comparison between the tower and some other objects, so I will call these properties comparative properties.

The properties I have listed vary in significance: the tower's height is probably more important than its material composition (up to a point: it would have been an altogether different kettle of fish, so to speak, if it were made of cheese), and the latter is perhaps more important than its nocturnal illumination but not quite as important as its general location. Such rankings of properties in terms of their relative significance is natural and common, attesting to some shared background understanding, of a roughly pragmatic nature, about the kind of thing the Eiffel Tower is. Though many of the true statements about the tower describe its properties, attention to our ordinary tower-talk reveals that not all do. This is particularly evident with regard to relational facts: that I climbed it in 1972, and its distance from Mars. Why don't these facts ordinarily count as properties? The answer probably lies in the same criterion of pragmatic significance we just mentioned and in light of which the tower's properties can be ranked. Understanding the kind of thing the tower is also implies the kinds of statements it may be ever profitable to make about it. When they fall below a certain threshold of pragmatic importance, true facts about the tower no longer count in everyday discourse as properties at all.¹⁶

¹⁶ So I don't mean to raise the separate question whether there is a useful logical or ontological sense that would admit the "properties" ordinary speech implicitly excludes, or, for that matter, exclude "properties" that the latter takes for granted.

How can the Eiffel Tower be affected? Clearly, changing any of the properties I have listed would count as affecting the tower. These effects could take place with regard to the tower's intrinsic properties—it would be shortened if its legs were sawed off—as well as with regard to the relational properties. Notice however that unlike the intrinsic properties, the relational ones could change in two very different ways. Take the tower's municipal location. That could change if the tower were extracted from Parisian soil and moved to London. But the tower could also lose its Parisian address while standing still, if the rest of Paris were to be moved a hundred miles down the Loire valley. Changing the tower's relational properties without causally affecting the tower itself need not, of course, be as fanciful as that. Consider the other relational property I've mentioned: being illuminated at night. The tower could be affected in this regard if it were moved over to a dark corner of the city, out of reach of the floodlights. But a much less fanciful way in this case of changing the same property would be to simply turn off the lights.¹⁷ The same is obviously true regarding the comparative properties as well: by sawing off its legs the tower would no longer be the tallest structure in Paris. But the same change in this comparative property would be effected if a taller building were erected on the other side of town. The respective effects on the tower of each member in the pairs of procedures we imagined are the same: in the first case it ceases to be in Paris, in the second, it winds up in the dark, and in the third it is no longer the tallest. Yet the causal chains that led from each member of the pair to these respective effects are, as far as the tower is concerned, radically different. The first procedure in each scenario involves a

¹⁷ This example is actually somewhat out of date: the external floodlighting was replaced in 1986 by a system of illumination that is within the tower's structure. The change corroborates my main point: the significant property of the tower in this case is its being illuminated at night; it doesn't seem to matter much, relative to this property, whether the light comes from within, making it an intrinsic property, or from without, thus rendering it relational.

causal chain that terminates in the tower itself, whereas in the second procedure the causal chain evades the tower, engaging instead with the other objects—call them the relational terms—on which the tower's respective relational properties depend.

The case of human beings, to which I now turn, is not essentially different. We too have relational and comparative properties, as well as intrinsic ones. In describing a person it would be natural to list such intrinsic facts as his height or intelligence, as well as relational properties such as marital status or nationality, and comparative ones: being a valedictorian or the best lawyer in town. Consequently, we too can be affected by things that are not causally linked to us but rather to the relational terms. Consider the case of Bob, who is married to Mary. A serious quarrel causes them to separate. A separation could be a protracted and gradual process, but to simplify matters assume that this one was abrupt and final, so we can date it with precision. If the separation took place, say, on Monday at five, then, all would agree that Bob underwent an important change of status at that time: all at once, he went from being married to being separated. But notice that the causal chain leading from the quarrel to the separation can take two different paths. In one scenario, Bob packs his belongings and leaves in anger, determined never to return. But alternatively, Mary may be the one to leave, with Bob staying put, neither initiating nor desiring nor executing the separation. We can even imagine that Bob isn't at home, but at work, when the separation takes place, so that he does not find out about it until later.

A similar account applies to the example of the bereaved father we discussed in the preceding section. The question How can a car accident that takes place in a faraway land be said to affect the father? can be now easily answered in terms of the relational nature of the property of being a father. It

is this property that makes the father vulnerable to effects due to causal chains, in this case of a fatal nature, that terminate not with him but with his distant son. This account also helps distinguish the case of the father from that of an interest in the rain forest with which I've contrasted it before. Remote deforestation will not be said to affect me no matter how interested I am in it, since no corresponding relational property links me to the rain forest. Why not?

I have mentioned earlier the important role supposedly played by a vaguely defined pragmatic background in constraining the kinds of properties ordinary speech acknowledges. In selecting the kinds of properties it is sensible to ascribe to human beings, it is safe to assume, account is taken of people's typical interests. But though the link between interests and properties may be close, it is neither tight nor direct. Our judgment in a particular case that an individual has been affected will be mediated by our general understandings of what counts as a relational property and its relational term. The availability of standard terms and set phrases such as "father," "spouse," or "widow" that embody putative properties seems relevant though not dispositive here. To be sure, being a parent or a spouse is ordinarily bound up with numerous and highly significant causal and experiential effects on the persons concerned. These facts are doubtlessly relevant to these terms' evolution, but they need not be present on each occasion of the terms' employment. It would be altogether natural, for example, to refer to a man's death as an event that befell his wife by reporting that "she was widowed," even if she had never heard about the event, and despite our belief that she wouldn't have cared one bit if she had. Conversely, our ordinary conceptual framework may place some things about which one may care a lot—the rain forest in my example—outside the range of a person's relational terms, so that their fate will not by itself be described in terms that imply an effect on the individual concerned.

4. THE EXTRA-PERSONAL EFFECTS OF THOUGHTS

Recognizing the pervasive role that relational properties play in ordinary speech explains how a person may be affected by a thing that is not causally linked to him when that thing is instead causally linked to a relational term. But this account may seem not to advance my main claim much. Although we may have now somewhat weakened the grip of the argument from causation, we haven't weakened it enough: applied to the case of thoughts, our model would allow that one person's thoughts can affect another without engaging causally with her, as long as those thoughts do engage causally with a relational term that defines one of her relational properties. If thoughts had no extra-personal causal efficacy, however, they wouldn't be able to do that: they would be no more capable of causally engaging with any relational term than they could causally affect directly the other person herself. Having posed the problem in these terms, the solution, or rather solutions, seem rather obvious. First, although a thought cannot causally affect any relational term that is external to the thinker, the thought itself can simply be another's relational term. Secondly, the thinker can be the relational term that defines another's relational property. Since a thought affects the person whose thought it is, then in such a case, by affecting the thinker the thought will have affected the other as well.

I have just given an example of one person serving as the relational term that fixes another's relational property: the son to his father. Hence the judgment that the accident that killed the son affects the father as well. But other effects on the son, short of death, could similarly be said to affect the father. Suppose that the son barely survived the car accident, remaining however in a permanent vegetative

state. That too, I believe, could be credibly described as a disaster that befell the father, even if he never learns about the event. In a similar vein, imagine now a third scenario: the son, whom the father remembers as a sunny, cheerful lad, has sunk, perhaps due to a car accident, into acute depression, and is relentlessly tormented by suicidal thoughts. Clearly there is no fundamental difference as far as effects on the father are concerned between this scenario, which has to do entirely with the son's state of mind, and the preceding, more physical affliction. In this case, the son's state of mind, his thoughts, can by itself affect the father because of the fact that the subject of those thoughts serves as the relational term that constitutes the father as such.¹⁸

Two clarifications are in order. My argument throughout depends on certain shared linguistic intuitions I assume. I cannot convince the reader who does not share them or at least the belief that others do. But there is one cause of doubt about these intuitions to which we ought to be alert. There is some tendency, I believe, to withdraw one's assent to the judgment, say, that a disaster struck the father whose son died, when the explicit question is posed as to whether the father can be said to have been "affected" or "harmed." This tendency, moreover, seems to increase the more attenuated the effects on the son himself are. Such a tendency does not seem to me to detract from the force of my argument. We know from the outset that we are in general in the grips of a causal paradigm: this is after all our putative explanation of the belief that thoughts can have no extra-personal effects. When verbs

¹⁸ Feinberg considers a similar case, and reaches a similar conclusion: the wife's sinking into a depression can all by itself count as a harm to her loving husband. The reason for Feinberg lies simply in the intensity and duration of the husband's love, which give him a vicarious interest in his wife's well-being. But here again the account seems to leave a conceptual gap: no matter how intense and enduring my love or admiration for, say, Queen Elizabeth, her mood swings, of which I know not, will hardly be said to affect me. See Feinberg, Harm to Others, p. 71.

such as “harm” and “affect” that bear their causal connotations on their sleeves are brought to our attention, we resist the implication that they may apply in the absence of a familiar causal pattern. The relevant responses that ought to guide our investigation are accordingly the ones we can imagine ourselves giving to circumlocutions that are more likely to catch us unawares, while on a linguistic cruise, so to speak, and philosophically innocent. Wouldn't it be altogether natural to report about the father whose son had been run over by a car, that “something awful happened to him” and then add “but he never learned about it”? Such a description would seem clearly in place in the event of the child's death, but it would seem also appropriate, perhaps with a bit more straining, in the second scenario of vegetative survival. Acute depression would probably raise even greater doubts.

These doubts lead to my second clarification, which is to recognize that we're faced here with a sliding scale, and to indicate that nothing in my argument depends on the precise cut-off point at which ascribing to the father effects that derive from mishaps that befall the son would seem linguistically natural. My argument only depends on the realization that in our example, as far as effects on the father are concerned, there is in principle no difference between the son dying, or becoming vegetative or depressed. In the absence of knowledge on the father's part, there is no direct causal link between him and any one of the tragic events. The same relational logic that licenses describing the first two events as affecting him seems to apply to the third case in which the son's mental states are involved.

In the examples I have just considered, it is natural to think of the son as the relational term that fixes the pertinent relational property—fatherhood—by virtue of which changes in the son's states of mind can be said to affect the father. The son's state of mind—his depression—does not directly involve the father; it can be said to affect the father only by virtue of a prior and independent connection that

binds the two of them together. But in other cases the bond between the people concerned is at least in part a matter of the contents of their reciprocal states of mind. Then the thoughts themselves can be plausibly seen as the relational term that defines a relational property in another.

Most people would probably agree that being a cuckold is undesirable even if, as is sometimes the case, one never finds out about the spouse's infidelity. This judgment can be analyzed along by now familiar lines: being a spouse is a relational property, one of whose ramifications, defined in terms of the other spouse's behavior, earns the aggrieved party the said sobriquet. We do not have an analogous single English term to describe the condition of the victim of unrequited love. But the situation seems to me relevantly analogous. Consider Joe, who has been married to Jane for many years. Joe is deeply in love with his wife, and has always believed that his feelings were fully reciprocated. But they were not: Jane in fact despises Joe, meticulously maintaining a deceptive facade, perhaps for the children's sake. Now Joe's situation may be plausibly viewed as worse than that of the betrayed husband: it is quite possible, for example, that given the choice, Joe would prefer an occasional fling by an otherwise loving wife to a wife who despises him but remains faithful all the same. Those who know the truth about Jane's feelings will judge correctly that without even being aware of it, Joe is trapped in bad situation. The account we suggested in the case of the cuckold would seem to apply here as well: the relational property of being a spouse makes Joe directly vulnerable to Jane's feelings of which he knows not, just as he is vulnerable to aspects of his wife's behavior of which he remains ignorant. In the marital context, some of each spouse's feelings as well as behavior are themselves relational terms whose content plays a direct role in the other spouse's life. Moreover, although the marital relation makes this example more compelling, it is not strictly necessary. We can imagine other scenarios of unrequited love, in which a

lover's life will be plausibly said to be marred by the loved one's feelings, whether he knows them or not. Such feelings may turn him into a *disappointed lover*, with the italicized expression denoting not the lover's own state of mind, but the actual state of affairs. There is no point in spinning out such additional scenarios. What is important to acknowledge is that the more attenuated the relationship between the parties, the less likely we are to ascribe relational properties to them that derive from that relationship, and consequently our tendency to view their states of mind as affecting each other will be correspondingly attenuated.¹⁹

The romantic domain is of course not the only one in which one's thoughts can serve as the relational term for another's property. Let me consider a different kind of example. P is a young and insecure painter, who desperately aspires to be a "real artist," that is to say someone whose paintings have genuine artistic value. Whether or not P is indeed a "real artist" is doubtlessly an important fact not just to him but about him. But what determines whether P's art has genuine value? This is obviously a difficult question to which many different answers have been given, but one plausible answer points to the experts' opinion. Ordinarily this supposes a consensus or preponderance of views among a group of experts. But suppose that in our case, only one person's, E's, judgment counts: because the area is so esoteric that experts are rare, or because of E's stature as the dominant and non-controverted authority. "To be appreciated by E," on these assumptions, is tantamount to "being a real artist." If being a real painter would be a property of P, then on this account it would be a relational property, constituted by E's opinion. E's thoughts thus play the part of the relational term that defines another

¹⁹ Cases of this kind are also sometimes discussed in connection with the concept of happiness. See, for example, Richard Kraut, "Two Conceptions of Happiness," *Phil. Rev.* 88 (1979): 167–97.

person's relational property. Suppose now that E's initial reaction to P's work is enthusiastic, but that later, after further observation and reflection, her enthusiasm sags and she dismisses P as a failure. On the assumptions concerning artistic merit that we've made, it would be natural for someone familiar with E's change of heart to describe the situation as follows: "It had seemed for a while that P had made it as a painter, but then he turned out to be a flop." Notice that the accuracy of this description does not depend on E's disclosing her opinion to anyone else. Even if P were to spend the rest of his life under the illusion of artistic success, by his own criteria the truth of the matter would be that he was a failure.

As P's story again illustrates, it is possible for thoughts themselves, rather than the person holding them, to be the relational term that fixes a property of another person. But treating E's opinions regarding P's artistic merit as the relational term that constitutes P as an artist is admittedly artificial: Artistic merit is ordinarily determined not by a single authority, but is rather a matter of some collective determination. To take account of this fact, we must distinguish two kinds of relational properties: individual and collective. An individual relational property is one in which the relational term is or makes essential reference to another individual; in a collective relational property, the relational term is or makes essential reference to a collectivity. My extant examples, such as being a parent or a spouse, belong to the former category; ethnic or racial properties—being French or Caucasian—illustrate the latter.²⁰

As the examples of race and ethnicity suggest, some of a person's most important characteristics, those that form what is often loosely referred to as her social identity, are collective

²⁰ These two categories are not exhaustive: physical objects, such as geographic locations, can also play the part of a relational term. They are however irrelevant to our present concerns.

relational properties. Now many of these properties are not exclusively or even primarily a matter of other people's thoughts. But some are: reputation, popularity, fame, prestige, are all important aspects of one's social identity that are for the most part constituted by others' opinions. Some collective relational properties, perhaps all those just listed, only require that the views on which they depend be held distributively: there must be a sufficient convergence of opinion, but the opinions themselves can be held individually, without any reference to the fact that others hold them as well. Though communication obviously facilitates such convergence of views, it is not, strictly speaking, necessary: the convergence can occur spontaneously, in response to some salient characteristics of the person concerned or some events that involve him. It is enough, for example, for a person to acquire a reputation for heroism if everyone in town watches him save a child from a burning building : no public communication is required.

Other collective relational properties require that the underlying views be held jointly, that is with a reciprocal awareness that others hold them as well. Consider authority and leadership. To be an authority in a certain professional area, one must be recognized as such. For the most part, the recognition in question must take the form of convergent beliefs, each of which holds the putative authority in high professional esteem based in part on the belief that others hold similar views. Here communication is more imperative than in the case of the distributively held beliefs, but it is not mandatory either. As we saw earlier, in our discussion of the alleged inaccessibility of thoughts, we can know each other's thoughts with practical certainty in the absence of any communication, by observing each other observe some salient properties or events. Think, for example, of the charismatic leader of an informal group. One, call her L, can attain spontaneously such a position without any communication

among the members of the group, if the following two conditions are satisfied: most members are captivated by L's charisma, and they believe that the charisma is so overpowering that it is practically certain that most others will be similarly impressed.²¹

I have earlier distinguished comparative properties from other relational properties, and will now indicate how comparative properties provide an additional avenue through which people can be affected by others' thoughts. To recognize this possibility we only need remind ourselves that we commonly ascribe to people mental properties, such as being smart or imaginative, and that such properties often give rise to comparative properties: being the smartest kid in class or the most imaginative member of a department. To acquire or lose such a status is often of considerable importance. One way in which such changes can take place is through an improvement or a decline in the quality of one's own mental processes. But the other way in which one will attain or abdicate such a comparative position is by way of mental changes in others. Joe may no longer be the most imaginative member of the department because he grew duller over the years. But he may have also lost the title because Mary has recently flourished in this respect. Obviously, these judgments assume a sharp separation between matters of fact, including inner, subjective fact—e.g. who, Joe or Mary, has in fact a richer imagination—and the epistemological question of how do we tell. The latter, unlike the former,

²¹ Indeed, strictly speaking, the second belief may suffice to establish L as the leader, and it need not even be a true belief. L may possess a salient characteristic, an imposing appearance perhaps, which on close inspection is not charisma, but resembles it sufficiently to lead most members of the group to believe that the others are bound to be misled by that characteristic into believing that L is charismatic, and hence view her as the leader. Notice that for such mental bootstrapping to successfully establish L as the leader requires the absence of any communication among the members. In the terminology I propose, what we witness here is a situation in which the convergent thoughts of a group of people, uncommunicated and unexpressed, serve as a collective relational property of L, by constituting her as the group's leader.

does for the most part depend on expression and communication. But even here the dependency is only partial. For other members of the department to be able to tell who's most imaginative, both Joe and Mary must publicly display their imagination. But for Mary herself to know the truth about Joe's comparative standing in this regard, no manifestation of her imaginativeness is required: since she has first-hand knowledge of her own imaginative life, she can reach a confident conclusion about Joe's standing with only him publicly manifesting his imagination.

Two other points concerning comparative properties. Some readers may have doubts as to whether these are genuine properties of the individuals to whom they're ascribed. That a change in others counts as a change in such a property may increase such doubts. It may perhaps serve as some reassurance of these properties' good standing that many indisputable properties that do not have a comparative form are in fact comparative properties in disguise. Seemingly non-comparative adjectives such as "tall" or "slim" or "smart" do imply a reference group that supplies them with a baseline and a metric: a particular society, or some specific sub-group, or, quite commonly, humanity as a whole.²²

The other point is that although many comparative properties are well-established, the analysis I have here proposed does not equally illuminate all of them. Suppose Johnny had been for quite a while the saddest kid in his class, his sadness a source of consternation to his teachers, parents and friends. Then one day a teacher reports to the parents that Johnny is no longer the saddest kid in the class. This piece of news would ordinarily come to them as a great relief, so the parents would be justly dismayed to learn that all the teacher meant was that another child, even sadder than Johnny, had just joined the

²² Cf. David Hume, "Of the Dignity and Meanness of Human Nature," in Political Essays, Charles Hendel, ed. (New York, 1953).

class. This example ought to alert us once again to the significance of the ever-present pragmatic background against which the judgments ascribing both properties and effects are made. The aspect of that background that is relevant to the present point is the qualified validation pragmatic considerations provide to a competitive dimension in human affairs: the more legitimate, salient and important the competitive aspect is with regard to the assignment of a given comparative property, the less it is likely to matter whether the change in that property is a product of a change in the subject himself or in the others to whom the subject is compared. Sadness is clearly not a characteristic with regard to which competition is encouraged or accepted. Consequently, what really matters about Johnny's sadness are internal changes in him rather than comparative judgements that reflect shifts in other people's moods.²³

5. SOME NORMATIVE IMPLICATIONS

As I mentioned earlier, thoughts can obviously originate harmful causal chains whose middle terms are actions or speech, thus creating a prima facie tension between the harm principle and the liberal commitment to the inviolability of thoughts. But a strategy that mitigates the tension seems available in these cases: the liberal can preserve the inviolability of thoughts without condoning any of their undesirable consequence by focusing all political and legal attention not on the thoughts but on their

²³ I have distinguished relational properties from comparative ones, but some properties are a composite of both. Being a popular member of the class is, as we saw earlier, a relational property that depends on what the other members of the class think of one. But being the most popular member is also comparative. One can accordingly lose this accolade in three different ways: through a change in the person herself, becoming, say, mean and nasty; through a change in the others' attitudes—they simply grew tired of her; or by the fact that a new member, more attractive than the heretofore favorite, joined the group.

overt, public manifestations. If my argument is correct, however, such a strategy is not always availing: we may not be always able to prevent the negative consequences by suppressing public manifestations of thoughts because in some cases there may not be any such manifestations. This leads to the conclusion that if the liberal is to maintain her commitment to the inviolability of thoughts she must do so not because thoughts can have no negative effects on their own but despite the fact that they can. I think, myself, that this challenge should be easy to meet: there are familiar and weighty arguments that I need not rehearse here that point out the inherent dangers of any attempt at thought control. We must, however, recognize that the Inquisitorial specter is a real one, as are many other anti-liberal ghosts that always lurk in the attic. Their presence is after all what gives liberalism its point and its agenda. In the present case, as in the others, one can only hope, if one is a liberal, that these ghosts will be held at bay by the power of argument and by the strength of commitment, rather than by denying the ghosts' existence.

My main aim in this section is not, however, to consider such general matters, but to focus instead on a more specific implication that my argument has on the current debate concerning hate speech: the injurious expression of prejudice and hatred toward the members of vulnerable, usually minority, groups. To what extent do the undeniably harmful effects of this kind of speech justify legal inhibitions in possible derogation of free speech concerns? Somewhat paradoxically, I think, recognizing the possible harmfulness of thoughts suggests an argument in favor of a more permissive approach to speech in this area.

I start with an analogy. Recall the example of the father whose son is killed. Should the father be informed? The news will be to him a source of great, perhaps devastating pain. And yet most would

agree, I believe, that he ought to be told. Even those who disagree with the conclusion will likely perceive that there is at least a genuine dilemma here, both horns of which have to do with concern for the bereaved father. But what precisely is the nature of this concern? One horn, as we've just seen, is straightforward: the father's experiential well-being will obviously be seriously set back by the disclosure. What's on the other side? It is natural to invoke here the father's right to know. But what is the basis for such a putative right? Autonomy seems a likely candidate, and it might seem to counsel disclosure in two different ways. Neither, however, seems to me entirely satisfactory. The first rests on the view that autonomy requires a capacity to make rational, and hence fully informed, practical judgments. Here autonomy is linked to the prospect for some action, but in the situation we envisage none is likely: there is nothing the father can do about the tragic event, and the validity of his claim to be told the truth does not seem to depend on his engaging in some mourning rites or the like: stipulate those away, and the right to know seems to remain intact. Autonomy can enter in another way, though: we assume that given the choice, the father would have wanted to know the truth, so by telling him we are serving his autonomy simply by satisfying his own presumed wishes. Putting aside the question whether and when satisfying someone's wishes is a matter of her autonomy, this consideration in favor of disclosure is unsatisfactory for another reason: it makes perfectly good sense in this case to maintain that the father ought to want to know the truth, even if we believe that he in fact wouldn't. To be sure, our belief that he wouldn't is itself a reason against disclosure, and this perhaps out of respect for his autonomy. But even here the reason need not be dispositive, and at any rate the question remains as to why it might be felt that the father ought to prefer to be told.

Though autonomy does play an important part in these matters, it cannot tell us the entire story.

The locution I've used—knowing the truth—will seem to provide a clue to the missing parts, but unless we're careful the clue may mislead. Aren't we all committed to the truth, and isn't that commitment enough to argue in favor of disclosure? But this of course won't do as a basis for the father's alleged right to know: he has no comparable claim, mercifully, with regard to the myriads of other deaths that occurred around the world that same day, nor for that matter with regard to most other events. It might be suggested that the claim to knowing the truth is limited to events that are of concern to the person, and that unlike his son's death, all these other deaths are of no concern to the father. But as we saw earlier in a somewhat similar context, "being of concern" won't do here either: the father may have an avid concern regarding the future of the rain forest with no resulting claim to be briefed about any untoward logging that goes on. A person does seem to have a right to know the truth, but the scope of truth in question is narrowly circumscribed; it's narrower even than what is to him of interest or concern. It is the truth about the person himself and his own life that seems pertinent here. An apt expression that seems to capture the point is that one ought not to live a lie. Why not? This, I think, is no longer a matter of one's autonomy, but a matter of one's dignity, and hence a matter of respect: willingness to face up to significant truths about oneself, painful or otherwise unpleasant as that may be, bespeaks self-respect, just as respect for another's dignity counsels disclosure of such facts even when distressing.

Though the balance of considerations may vary in the other scenarios I've discussed earlier—such as the betrayed spouse and the failed artist—the basic dilemma they pose is the same: it's the choice between avoiding psychological pain and hurt feelings on the one hand and respecting a person's dignity-based right to know all significant self-regarding truths on the other. A crucial aspect of

this dilemma is of course the scope of self-regarding truths: which facts are self-regarding? The answer, as we saw, depends in part on what we consider our relational properties to be, and hence on what things we take to affect us by affecting our relational terms.

Now obviously not everything other people think about us can be said to affect us in some way; thus by the criteria I propose, we do not have a general prima facie claim to know such thoughts no matter how understandably curious about them we may be.²⁴ But as I've indicated earlier, on some occasions other people's attitudes toward us do serve as the relational terms that fix important features of ourselves. Particularly relevant here are the attitudes held by a community or a society that define aspects of its members' social identity. If, for example, a community despises some of its members and holds them in contempt when acceptance by that larger community constitutes an aspect of their social identity, those members are negatively affected even if the attitudes remain latent and their objects bask in illusory acceptability. They live a lie, and their dignity is trampled. Our attitude toward hate speech in circumstances such as these, I suggest, ought to resemble the dilemma I've just described in the case of the bereaved father and the related scenarios. Though the speech be offensive and painful, it is for the most part just a messenger of bad tidings. Gagging the messenger will not make these tidings go away, nor does it neutralize their poisonous effects.

It is sometimes advisable to state the obvious in these sensitive matters. Nothing I said implies that those who engage in hate speech ought to get any credit for honesty or for respecting the targets' dignity: typically nothing will be farther from their minds; their goal is to add the injury of hurt feelings to

²⁴ Hence, the examples recently adduced by Thomas Nagel to demonstrate the perils of excessive candor in social interactions fall outside the purview of my argument. See his "Concealment and Exposure," 27 *Phil. & Pub. Aff.* 3 (1998), pp. 10–17.

the insult of bigotry and hostility. Nonetheless, public policy toward hate speech ought to recognize that within the nasty social situation in which it is likely to occur it plays a more ambiguous role than we often realize. If we recognize that the attitudes that hate speech expresses are, if pervasive, by themselves harmful to their targets' lives, and if we also believe that this harm is only exacerbated by the indignity of illusion, then despite the obvious hurt, by bringing things into the open, hate speech performs a redeeming function as well.

6. THOUGHTS AND INDIVIDUALISM

The normative implications I have mentioned, though hopefully of some interest, are not, however, the main intended pay-off of the position I've outlined. Only rarely do other-affecting thoughts remain utterly unexpressed, and when they do, one expects and even hopes that other considerations will overshadow and overwhelm the practical significance of my point. My argument, if correct, does however have a broader philosophical significance quite apart from such direct normative and practical ramifications. The broader significance concerns the limits of individualism, and it may be best brought out by relating my argument to another line of thought with which I think it has much in common.

Individualism means different things to different people, but here it stands for a picture of the mind as a self-contained entity that is enclosed within (the brain of) each individual human being. So described the picture is vague, but not so vague as to prevent recognizing some well-known objections. The specific objection with which I'd like to associate my argument is mostly the joint product of two writers, Hilary

Putnam and Tyler Burge,²⁵ who through a number of famous thought experiments have argued for what has come to be known as “meaning externalism”: the view best summarized by the slogan that “meanings are not in the mind.” I cannot here even begin to do justice to the intricacies and complexities of the issues involved. But the main point is rather simple. It depends on the observation that our use of language is commonly based on an incomplete understanding of the terms we employ. Since our use of terms is as a matter of course based on only partial understanding and mastery of them, the only way in which meanings can be attributed to us must incorporate the way these gaps in our own understanding are filled in by factors outside of us. These factors are of two kinds: facts about the environment and facts about usage in the relevant linguistic community. The main emphasis in this line of reasoning, at least as developed by Burge, is that these factors determine the content of the agent’s mind without causally interacting with her: we can imagine two agents whose inner composition and causal histories are identical, and yet the contents of their minds will be said to vary depending on the natural and social environment in which we imagine them to be.

This brief sketch of the meaning-externalist position will suffice to enable me to draw the following three connections to my argument. First, both views point out breaches in the cocoon in which on the individualist picture the mind is enveloped. Through these breaches the mind transacts with the world outside of it without the mediation of any causal chains. The second connection is by way of contrast in the direction of these non-causal transactions in which the mind is involved. Meaning-

²⁵ The main papers are, Hilary Putnam, “The Meaning of ‘Meaning’,” in Philosophical Papers 2 (London: Cambridge University Press, 1975) p. 215; Tyler Burge, “Individualism and the Mental,” Midwest Studies in Philosophy 4 (1979: 73–121), and “Other Bodies,” in Thought and Object, Andrew Woodfield, ed. (New York: Oxford University Press, 1982), p. 97.

externalism highlights the mind's passive susceptibility to external factors: the world, we can say, participates non-causally in constituting or fixing the content of the mind. My argument documents the logical possibility of a more active externalism: the mind can have direct, non-mediated and non-causal effects on the world outside of it.

Thirdly, the connection between the breaches in the individualist picture highlighted by Putnam and Burge on the one hand and the breaches that I have documented on the other is not coincidental. One way to see how the two views relate is to think of meaning-externalism as at least in part an elucidation of the semantics of terms such as "meaning" and "content." In a similar vein, mine may be seen as an argument about the semantics of "effect" and cognate expressions, such as "disaster" or "mishap." What we find in both cases is that many of our ordinary concepts organize reality in "relational" packages, that is, packages that cross-cut with concepts that form individuals, by combining elements, including mental elements, of different individuals. And since the former types of concepts have in general as much (or as little) warrant as the latter, there is no general reason for privileging the ones that track the boundary of a single individual over those that don't. Indeed, given how pervasive and intermixed both types of concepts are, the very idea of a "boundary of an individual" is thrown into question.