Royzman, E., McCauley, C. R., & Rozin, P.

From Plato to Putnam: Four ways to think about hate

(pp. 3-35), 2004

In:
Sternberg, R. J. (ed.).
The psychology of hate.

Washington: American Psychological Association


The
Psychology
of
Hate

1

FROM PLATO TO PUTNAM: FOUR WAYS TO THINK ABOUT HATE

EDWARD B. ROYZMAN, CLARK MCCAULEY, AND PAUL ROZIN

The English word *hate* signifies a domain of dizzying versatility and breadth. Consider: A child hates a bully, a woman hates her estranged husband, a bigot hates the "inferior race," a fighter for human rights hates injustice, an anorexic hates her calves, a widower hates long weekends, some Jews hate Germans, Sallieri hates Mozart, Mozart hates bogus art, millions hate Saddam Hussein, little Jimmy hates his broccoli. People act out of hate, feel overcome by hate ("I really hated her at that moment!"), endure it as a life-long affliction ("Unfortunately, I hate Aunt Margie, always have, always will . . ."), or elevate it into a virtue ("I hate hypocrisy").

In spite of scores of books, topical discussions, stirring editorials, and hate-fighting initiatives, there is no single, commonly accepted definition of *hate*. What is it precisely that people are discussing, fighting against, or trying not to let into their hearts? In this chapter we review major conceptions of hate, both classic and modern. Finding no consensus, we discern in the welter of proposals four ways of asserting an understanding of hate: categorical

We thank James Russell and John Sabini for their comments on drafts of this chapter.

meta-description, (paradigm-based) causal explanation, stipulation, and Platonic insight. Each of these offers a distinctive basis for defining hate and for evaluating that definition. We conclude that a clearer understanding of hate, its origins, and its consequences will be advanced by paying more explicit attention to the different roles that different authors ascribe to their conceptions of hate.

CONCEPTIONS OF HATE: OLD AND NEW

Classic Definitions of Hate

The classic formulations of hate, those by Aristotle, Descartes, Spinoza, Hume, and Darwin, are notable for their contradictions. For Descartes (1694/1989), hate was an awareness of an object as something bad and an urge to withdraw from it. For Spinoza (1677/1985), it was a case of pain (sadness) accompanied by a perception of some external cause. For Aristotle (trans. 1954), the distinguishing phenomenological fact about hate was that it is pain-free (in addition to being incurable by time and striving for the annihilation of its object). Hume (1739–1740/1980) argued that neither love nor hate can be defined at all, because both are irreducible feelings with the introspective immediacy of sensory impressions. Darwin (1872/1998) also saw hate as a special feeling, one that lacks a distinct facial sign and manifests itself as rage.

The contradictions in these definitions have to do with the weighting of feeling and (affect-free) judgment, with some authors emphasizing a negative feeling toward the object of hatred (Spinoza, Hume, and Darwin) and others emphasizing a negative judgment about that object (Descartes and Aristotle). Aristotle's qualification that hatred does not change with time suggests, in particular, a negative judgment about the character or essence of that which is hated, rather than a reaction to a specific negative trait or action. Another source of contradiction involves the behavioral tendency associated with hatred: Descartes suggested withdrawal, whereas Aristotle suggested attack. Darwin's understanding of hatred also suggests attack, insofar as hatred is manifested as rage, but Descartes made hatred more like fear (or disgust) in its link to withdrawal. In these classic treatments of hatred, then, is already apparent the outline of current contentions about the beliefs, feelings, and behaviors associated with hate and about the relation of hate to emotions such as anger and fear.

Emotion or Disposition?

Recent research understands *emotion* as a set of "multicomponent response tendencies that unfold over relatively short time spans" (Fredrickson

& Branigan, 2001, p. 125). That is, emotion is understood as episodic rather than dispositional, and any particular emotion is experienced as a pattern of specific cognitions, subjective experiences, and physiological reactions. Such garden-variety emotions should be distinguished from *episodic dispositions* (Averill, 1991; Ryle, 1949), or relatively long-term tendencies to emote (in the sense of displaying a distinct episodic emotion) about an object, event, or person, or a category of objects, events, or persons.

A very similar distinction between emotional experience and the disposition to experience emotion was offered in Shand's (1920) far earlier and much-neglected treatment of hate. For Shand, hate was the perfect antimony of love: Love involves a positive alignment between the emotions of the lover and the fortunes of the beloved, whereas hate involves a negative alignment:

> The health and prosperity of the loved object are causes of joy: in hatred, they are causes of bitter sorrow. In place of the delight of being again with one we love, is a peculiar mixture of repugnance and anger when we find ourselves again in the presence of one we hate; the one impelling us to avoid the person, the other to attack him . . . The joy of hate is the opposite of the joy of love, being caused by the suffering, loss of power and reputation of the hated person; and the sorrow of hate is opposite to the sorrow of love, and is caused by his power, reputation, and happiness. (p. 59)

Shand (1920) described hate as a *syndrome*, or a bundle of episodic dispositions united by a common emotional object or a common category of such objects. The key feature of such a syndrome is that a person may be legitimately characterized as having it without being imputed any corresponding episodic state.

Patriotism is a case in point. The term has a dual meaning—it refers both to an ideological stance and an affective syndrome that goes with such a stance; people commonly speak of feeling more or less patriotic at different periods of their lives. What are these feelings like? The obvious answer is that no such feelings need to exist, at least not apart from a person's tendency to experience other kinds of feelings (pride, sorrow, shame) in which his or her affective life appears to be aligned with the fortunes of his or her nation (cf. Ryle, 1949; Sabini & Silver, 1998). Thus, a patriotic person would be expected to feel joy and pride when his or her nation is victorious, sorrow and sympathy when it is facing a crisis, anger when it is unjustly slighted, and despair when it suffers a humiliating defeat. Yet carefully as one may inspect a patriotic person's interior life and daily habits, one will never find in it a trace of the special feeling called "patriotism" that exists apart from all of the above. In fact, in our view, the very expectation of such a discovery should automatically attract the suspicion of the so-called category mistake (Ryle, 1949), the mistake of imagining that a given higher

order phenomenon exists as something separate and independent of its basic constituents. (Ryle's own famous example is that of a foreigner who, after being shown the various lecture halls, libraries, lanes, and gardens comprising the Oxford University, asks, "Yes, all of these are fine, indeed, but where is the Oxford itself?").

The syndrome that Shand (1920) identified with "love" has also been described as *positive identification* (McCauley, 2001), or a tendency to track and react congruently to the fortunes of another person or group. Similarly, "hate" would qualify as the perfectly symmetrical syndrome of *negative identification*. In this view, hate is neither a special emotion nor a blend of emotions, but rather a tendency to emote in a number of ways to a number of situations involving the object of hatred.

Modern Conceptions of Hate

In the modern literature on emotions, Ekman (1992) has advanced a view of hate similar to Shand's (1920) idea of hate as syndrome. Ekman argued that hate is not an emotion but an "emotional attitude," the term he reserved for affective phenomena that are "more sustained [than a garden-variety emotion] and typically involve more than one emotion" (p. 194). That is, Ekman saw hatred as more an episodic disposition than a particular emotional experience. In contrast, Elster (1999) argued that hate is an emotion after all, one that is caused by a judgment that another is evil: "Negative emotions triggered by beliefs about another's character. (Contempt is induced by the thought that another is inferior; hatred by the thought that he is evil)" (p. 21).

Although Elster saw hate as separable from contempt, Sternberg (2003) recently proposed that both disgust and contempt are special kinds of hate, "cold hate" and "cool hate," respectively (see also Oatley & Johnson-Laird, 1987, for a claim that hate is a derivative of disgust). Sternberg's proposal is part of a broad theoretical typology based on the principle that, like love, hate can be characterized in terms of three action–feelings components: (a) intimacy (more precisely, the negation thereof), (b) passion, and (c) commitment. The feelings and actions associated with the first (negation of intimacy) component include revulsion–disgust and distancing, respectively. Fight-or-flight is the action pattern, and anger–fear are the feelings attending the passion element. The last (commitment) component involves an attempt to devalue the target of hatred through contempt. On the basis of this triangular structure, Sternberg posited a variety of hates. There is, for example, the already mentioned "cool hate," composed solely of disgust, and "hot hate," composed solely of the anger–fear combination. There are also "cold hate" (devaluation through contempt alone), "boiling hate" (disgust + anger–fear), "simmering hate" (disgust + contempt), "seething hate" (passion + commitment; also called "revilement"),

and, finally, "burning hate," which includes all three action–feelings components.

Although Sternberg linked hate to negation of intimacy and judgments of inferiority (devaluation), Solomon (1977) argued that to associate hate with malice, viciousness, and denigration of one's opponent is to confuse it with another emotion: resentment. He maintained that, unlike other hostile responses, "hatred is an emotion that treats the other on an equal footing, neither degrading him as 'subhuman' (as in contempt) nor treating him with the lack of respect due to a moral inferior (as in indignation) nor humbling oneself before (or away from) him with the self-righteous impotence of resentment" (p. 324). True hate, he argued, is an emotion of intimacy, respect, and strength—"There can be no hatred in weakness" (Solomon, 1977, p. 326); he saw this equality of power as part of hate's special mythology, ensuring that the antagonism involves an element of "mutual respect." Though Solomon referred to hate as an emotion, the general affective construct that appears to fit best his own characterization of hate dynamics is that of a syndrome.

Another recently popular approach has been to view hate as a kind of "personalized," "generalized," or "globalized" anger. For example, Frijda's (1986) analysis of hate distinguishes between emotions that "involve attaching positive or negative valence to a person or object" (p. 212) and those that ascribe valence to an action or event. This distinction led him to describe hate as "emotion that contains the component of object evaluation" (p. 212) or a highly personalized version of anger (see also Kolnai, 1998). Power and Dalgleish (1997) described hate as "generalized anger." It is anger that ceased to be "about one event or one thwarted goal and has broadened to embrace parts of, or indeed all, aspects of the person or object" (p. 334). The generalized evalution involved in hate has also been stressed by Ben-Ze'ev (2000).

One of us previously proposed that hate could be viewed as a compound of anger and fear (e.g., McCauley, 2002); the object of hate not only is blamed for some past maltreatment of oneself or someone one cares about but also is recognized as a source of future threat. For Beck (1999), hate is also defined, in part, by the belief that the source of the threat lies in some stable (although not necessarily global) feature of the hated person (Beck's examples suggest that he thinks of hate as capable of being both an episodic disposition and an episodic emotion): "Assigning responsibility to another for unjustly 'causing' an unpleasant feeling is a prelude to feeling angry. The persistence of a sense of threat and the fixed image of a malicious person leads to at least a temporary feeling of hate" (Beck, 1999, p. 44, see also the example on p. 11).

In a book-length treatise on hate, Dozier (2002) suggested that hate is a sort of adaptively antiquated anger, "anger phobia" as he calls it. In this view, the primary feature that distinguishes hate from anger is not its abidingness or its global focus, but its irrationality. However, in a dictionary

of psychological lexicon, Reber and Reber (2002) defined *hate* as a "deep, enduring, intense emotion expressing animosity, anger and hostility towards a person, group or object" (p. 315). The stress here is on the intensity and depth, rather than irrationality.

Gaylin (2003) proposed that real hate is a mental abnormality (p. 14) that exhibits obsessive–paranoid ideation and whose emotional core is rage (p. 34): "Hatred is a neurotic attachment to a self-created enemy that has been designed to rationalize the anxiety and torment of a demeaning existence" (p. 240). Gaylin also suggested that "the hate-driven people live in a distorted world of their own perceptions" (p. 202). Another psychoanalytic author (Blum, 1995) also has viewed hate as having pathological overtones:

> [Following Freud], hate is an ego attitude with the intent of destructive aggression. Hatred may be mobilized by need, fear, and frustration and by all unpleasant and noxious experiences. Transient or enduring, it tends to be closely linked to disturbance in psychic structure. (p. 20)

Ostensive Definitions of Hate

All the definitions of hate considered thus far, classic or contemporary, may be considered "direct" in that they handle their definiens as well as their definiendum (hate) in rather explicit terms. Such explicitness, however, is not necessarily essential to the practice of defining things. Philosophers also recognize a form of definition in which a definiens is communicated by either literally pointing to or otherwise indexing a case in which the definiendum is thought to be in evidence. Such definitions "by example" are called "ostensive" (Audi, 1995). Thus, one may give an ostensive definition of "pain" by pointing to a person in the throes of a toothache and saying "This is pain" or "This is what pain is like" (the more fine-tuned definition would entail the use of either statement right after stepping on someone's toe or poking him with a sharp stick).

Affirming the legitimacy of ostensive definitions is important because such definitions appear to be the staple for a number of recent political treatises on hate (e.g., Kaufman, 2001; Kressel, 2002). What may be initially confusing about such volumes is that, notwithstanding the word *hate* or *hatred* crisply displayed in their titles, they fail to offer anything like an explicit formulation of the very phenomenon they seem poised to elucidate. A careful reading, however, will indicate that what may at first pass for the absence of a definition is more generously interpreted as an ostensive definition.

Thus, though Kressel's (2002) "mass hate" does not openly define hate, it gives numerous instances of what hate is supposed to look like. The formulation of hate derivable from such instances is that hate is the motivational force that is responsible for acts of ethnic violence from the Holocaust to Rwanda, barring all those instances in which the violence is merely the result of conformity, blind obedience, or pure profit motive. The advantage

of such a formulation is that it pinpoints a set of real problems (ethnic discrimination, ethnic rioting, genocide) that make the study of hate, whatever it may be, genuinely worthwhile (see Sternberg, 2003, for a deft argument regarding the importance of psychological study of ethnic violence). This delineation of hate, however, leaves no room for doubt as to whether hate is, indeed, responsible for ethnic violence. Granted that *being driven to commit acts of ethnic violence* is an integral part of what it means to *hate*, any claim concerning the link between hate and ethnic violence enjoys all the certainty of a tautology. Furthermore, as long as one acquiesces to define hate in terms of whatever is happening in some presumed paradigmatic instance of hate, one should also be ready to accept that whatever is going on in such instances may turn out to be radically different from one's a priori conceptions of hate.

Hate as a Normative Judgment

A special case of ostensive formulation might be found in the concept of the so-called hate crime. Hate crimes are commonly characterized as "criminal actions intended to harm or intimidate people because of their race, ethnicity, sexual orientation, religion, or other minority group status" (Herek, Gillis, & Cogan, 1999, p. 945; Levin & McDevitt, 1993). In this context, hate means roughly that which motivates a deliberate act of physical violence or intimidation against a member of a minority group by virtue of him or her being a member of that group. In this view, classifying a criminal deed as one of "hate" is compatible with a wide range of psychological states, anything from anger to boredom to fear. Why not speak of, say, "anger crimes," then?

One reason, we think, is that describing a hate criminal as acting out of anger carries some unwelcome normative implications (in this context, *normative* should be taken to mean a statement reflective of the generally accepted norms and values, as in "blackmail is morally wrong"). There has been a strong and honorable tradition in both philosophy and psychology of linking anger to a legitimate defense of one's rights (Aristotle, trans. 1954; Averill, 1982; Baumeister, Stillwell, & Wotman, 1990; Hall, 1898; Rozin, Lowery, Imada, & Haidt, 1999; Scherer, 1997). Thus, citing "anger" as a motive of someone's aggressive action may carry a built-in implication of legitimacy-ascription and evaluative approval that is unwelcome in the context of criminal attacks based on race or sexual orientation. Thus, lurking behind the concept of "hate crime," there seems to be yet another cultural meaning of hate as that which motivates acts of senseless (normatively unjustifiable) violence. Of course, what seems senseless to one person may seem like a justifiable act of force to another. In light of Baumeister's (1997) argument that an attack will generally seem more gratuitous to the victim than it does to the perpetrator, insistence on the use of *hate* in a particular situation may

be less a matter of descriptive characterization than a reflection of normative commitment to identify with the plights of the victims while distancing from the viewpoints of the perpetrators.

The recognition of the normative or evaluative aspect of folk concepts is important because it may indicate the possibility that a disagreement over whether a certain folk psychological concept is applicable in a particular case may really be a disagreement about whether the evaluative meaning of that concept should dominate its descriptive (psychological) meaning or vice versa. A quick illustration will suffice. As Rachman (1978) pointed out, the folk concept of courage maps onto at least two psychological phenomena, including a capacity to remain fearless (with or without autonomic signs of fear) while in an objectively dangerous situation and a capacity or willingness to press on (with or without the associated autonomic arousal) in the face of fear. An ability to press on while being both subjectively and autonomically afraid seems to come closest to the ordinary conception of courage.

Yet few would describe a home burglary as an act of courage, even though each burglary may require the perpetrator to confront afresh and carry on in the face of a very real risk of capture, injury, or even immediate death at the hands of a disgruntled homeowner—as such, a burglar appears to perfectly fit the profile of someone capable and willing to uncouple the behavioral and the subjective–autonomic aspects of fear. The reason that home burglary is not thought to be an act of courage is that the ordinary conception of courage is not exhausted by a description of a certain psychological state, but it also entails a normative message to the effect that the action prompted by that state was a worthy one and that all good people should do likewise (Walton, 1986).

This normative aspect of courage was brought home spectacularly by the public outcry over comedian Bill Mahr's assertions that the 9/11 terrorists were not cowards and, by implication, were men of courage. If people's everyday notion of a courageous act were purely psychological, describing, as Rachman (1978) had it, a willingness or capacity to uncouple the behavioral and the mental elements of fear in the name of what the actor considers to be a worthy cause, then terrorists' suicidal assaults would be considered a paragon of courage.

The depiction of the 9/11 hijackers in terms of hate rather than anger or despair could also be interpreted as not so much a descriptive characterization of their psychological states as a normative avowal that the attack was an act of senseless aggression (essentially a "hate crime," albeit one of gigantic proportions) and that the terrorists' actions were deeply, monstrously immoral. We are confident that had some public figure stated that the terrorists "acted out of anger," the ardor of the ensuing controversy would have approximated (or even eclipsed) that occasioned by Mahr's incendiary comment.

To sum up, despite much recent attention to hate as a topic of discussion and intervention, there currently exists no generally accepted definition of hate. More grievously, there is nothing approaching a consensus on how to delimit the domain within which such a definition would fall. Meanings of hate differ both across and within contexts. Thus, it remains unclear if different authors are indeed discussing or intervening against the same thing.

The situation raises a number of questions: Why this cornucopia of meaning? How are psychologists to characterize the underlying disagreements? How are they to decide which disagreements are substantive and which are purely semantic? How are people to decide who is right and who is wrong? What would it mean to be right or wrong in this context? These are trying questions, to which we turn in the next section.

MAKING SENSE OF DIFFERENT CONCEPTIONS OF HATE

What is hate? In the preceding pages, we reviewed attempts to provide a single, nonarbitrary answer. In this section, we dissect the question itself. This question, we argue, lends itself to four distinct interpretations, each equipped with its own logic and its own evidentiary standard, corresponding to four different types of claims concerning the meaning of hate.

(Categorical) Meta-Description

Imagine opening a book and reading, "Hatred is a self-destructive impulse turned outwards." What would you make of this? One natural interpretation is that the author's aim is to offer a window on his or her own psyche. In this view, the author's purpose is to inform readers about a knowledge structure that the word *hate* prompts in his or her own mind ("Just letting you know what comes to my mind when you say . . ."), irrespective of whether he or she thinks that this opinion is in tune with the consensus of the field or if an analogous knowledge structure would be prompted in the mind of a representative layperson. Though self-descriptions may be aplenty in everyday conversation and memoirs, they may not be a very generous way to read a scholarly text.

A more charitable interpretation is that the author is offering a (categorical) meta-description, a claim that is part of what Russell and colleagues (e.g., Fehr & Russell, 1991; Russell, 1991) called "descriptive analysis" (see Wittgenstein, 1953). To interpret the statement "Hate is a self-destructive impulse turned outwards" in this manner is to assume that the author is not merely describing the pertinent knowledge structure in his or her own mind; rather, the author is offering a meta-description, a tidied-up summary description of what *hate* means to a particular culture, community, or group. For example, statements such as "In my opinion, most ordinary English-speaking

people would agree: Hate is a self-destructive impulse turned outwards" or "Based on my discussions with many persons, there is a consensus emerging within our field: Hate is a self-destructive impulse turned outwards" could be rightfully understood as summative or hypothetical claims regarding the lay and academic meanings of hate, respectively. The validity of the former claim could be established by investigating the meaning of hate within the designated segment of the lay community; the validity of the latter claim could be established via a comprehensive literature review. (One point of contrast for a categorical meta-description is a self-description, another is a meta-description that targets particular individuals; a historian may show an avid interest in what hate meant to Aristotle qua Aristotle, not qua an average Athenian or Macedonian).

Insofar as meta-descriptive claims concern categories ("English-speaking lay people"; "most emotion theorists") rather than individuals, they demand the same background assumptions as self-descriptive claims, plus the assumption that members of the designated category can show some level of consensus on what does or does not constitute hate. The question of whether such a consensus exists is orthogonal to the question of how the concept of hate is organized within the minds of the members of the target category. For example, most of the existing expert conceptualizations of hate seem to be marked by a clear set of defining features. However, given the lack of consensus among various emotion theorists as to what these defining features are, any single meta-descriptive claim intended as a representation of the prevailing psychological understanding of hate is bound to be inaccurate.

The basis for classifying a claim as self-descriptive or (categorically) meta-descriptive is its intent, not its source. For all we know, either claim may begin with a person asking him- or herself, "What sort of image (idea, set of defining or prototypic features) does 'hate' bring up in my mind?" The answer may then be served either as a self-description ("My personal intuition is . . .") or, assuming the person sees him- or herself as a good proxy for some target category, as a meta-description intent on capturing the mentality of that category ("All clear-thinking people will agree . . ."). Of course, meta-descriptions may have an altogether different source: results of direct surveys of those whose mentality one wishes to illuminate.

Meta-descriptive claims are clearly present in at least some analyses of love and hate. For example, Shand (1920) prefaced many of his statements regarding love, hate, and anger with the pronoun *we*. Considering Shand's skepticism about the emotion theories of his day, the most plausible interpretation of this practice is that *we* stands for the collective of theoretically neutral laypersons and that his analysis is to be taken as disclosing the everyday meaning of hate (love, anger), irrespective of the prevailing expert opinion. Likewise, Beck (1999), Ben-Ze'ev (2000), and Power and Dalgleish (1997) appeared to write as if intent on capturing the meaning of hate as everyone understands and uses it. As indicated earlier, the standard for judging the

truth or falsity of meta-descriptive claims resides in how well they pan out when assayed by empirical means vis-à-vis the target category specified by the claim. What follows is an outline of one meta-descriptive conception of hate that emerges from a number of survey- and interview-based studies tapping the perspective of a run-of-the mill, linguistically competent layperson. Spanning 5 decades and three continents, these studies show a remarkable degree of convergence regarding the much-discussed link between "hate" and "anger" and the evaluative feature that sets these two apart.

The traditional view holds that hate entails an intense desire for the annihilation of its object (e.g., see Aristotle, trans. 1954; Kolnai, 1998; Ben Ze'ev, 2000), but hate seems also consistent with a wish that the hated person experience sufferings whose nature and magnitude are roughly proportionate to one's own. Of course, the two wishes may be confounded, as when someone expresses a desire that his or her enemies "rot in hell." But, in principle, the desire for the hated one's distress may also appear in isolation from and even outweigh the impetus toward his or her destruction.

This idea is bolstered by what appears to be the earliest empirical investigation of hate. In 1950, McKellar carried out a series of semistructured 1-hour interviews whose stated purpose was to examine the nature of "hostile attitudes" (dislike, contempt, hate). Each interviewee was asked to think of an incident involving him- or herself and the hated or otherwise disliked person. The participant was then asked to imagine the "ideal resolution" of the incident and describe both his or her current feelings toward the "object of hostility" and how he or she would expect to feel and act toward this person upon encountering him or her in the vicinity of the interview room.

In tallying his results, McKellar (1950) described a participant who admitted to having "suffered prolonged physical pain as a result of the actions of the person whom he now hated . . . [and who was now filled with] malevolence towards the object of hate, [wanting] the other person [to experience] exactly the same amount of pain, no more and no less, that he had himself experienced" (p. 110). McKellar cited other cases that testify to what, in his opinion, is the main wish associated with hate, that is, giving the hated person "a dose of his own medicine." As one female participant put it, "I'd like to hear her cry the way she's heard me. I would still carry on . . ." Yet another participant said, "[I] would like to torment her; have her crawling to me for mercy. If she died it would be a pity . . ." (McKellar, 1950, p. 110).

To complicate the story somewhat, hate seems also consistent with an urge to altogether avoid the hated person. Of course, avoidance is consistent with the desire to punish another if one believes that the other will find one's absence distressing, especially if the reason one "hates" or "is angry" with another is that one has been previously the object of another's neglect. For example, in Fitness and Fletcher's (1993, Study 1) examination of real-life incidents of hate in the context of marital relationships, the participants described *leaving the situation* and *acting coldly toward the partner* as the urges and

behaviors most typical of hate. One possible interpretation for the discrepancy between these results and those reported in McKellar (1950) is that the range of hostile reactions is inherently greater when the object of hatred is a relative stranger rather than an intimate. It could also be that reports of hate in close relationships are more sensitive to social desirability concerns. Indeed, Fitness and Fletcher (1993, Study 2; see also Davitz, 1969) discovered that references to verbal and physical attack increased markedly when the participants were asked to think of an exemplary hypothetical instance of hate rather than to recall a real-life incident. In other respects, the simulated and the remembered incident were strikingly alike. On the basis of some further work, Fitness and Fletcher concluded that the results of the initial recall study could not have been due entirely to social desirability concerns and that avoidance reflected a real hate-associated action tendency, one that separated hate from anger. This link between hate and avoidance was further corroborated by a recent examination (Fitness, 2000) of hate and anger in the workplace; this study suggests that revenge and avoidance may both be part of an all-inclusive hate script.

Also, it is important to recall that self-reports offered by McKellar's participants represent what these individuals felt like doing under idealized conditions, not what they in fact did. Only 2% of Fitness and Fletcher's (1993) participants recalling episodes of hate did anything to hurt their spouses, a rate not much higher than the 0% for participants recalling episodes of anger. However, the anger-driven participants were significantly more likely to engage in what could be described as acts of verbal abuse and instrumental intimidation. It may be that vengeful fantasies are insensitive (perhaps precisely what makes them so "sweet") to the social constraints that limit people's urges as well as their actions in a real hostile episode. It could also be that the prototypic behavioral urges associated with hate are poised to be less extreme when a close relationship is at stake.

These findings intimate that, in the mind of many a layperson, hate is associated with a set of action tendencies that may vary with the context and are largely contradictory to the urge for destruction. Destroying a person stands in the way of making him or her suffer; conversely, making a person suffer depends on keeping him or her alive ("If she died it would be a pity" [McKellar, 1950, p. 110]). And getting away from a person seems to be something else altogether, despite occurring in many recollections of hate episodes.

Another early investigation into laypeople's phenomenological accounts of hate is that of Davitz (1969), who sought to produce "a dictionary of emotional meaning" whose authority would derive from the validated consensus of contemporary laypersons of various ages and backgrounds. Toward this goal, Davitz created a 556-item checklist describing the presumed experiential properties of 50 emotional experiences, including those of anger and hate. The checklist was submitted to a sample of

50 people, "25 men and 25 women, all volunteers, ranging in age from 20 to 50, and including both White and Negro subjects" (Davitz, 1969, p. 8). These 50 individuals were instructed to think of a specific instance when each of them experienced a given emotional state. The participants were then asked to give a brief description of the situation that triggered that state and mark any statement in the checklist that described his or her experiences while in that state. The results of the survey were subsequently analyzed for response frequency.

Davitz's (1969) decision, a decision that he acknowledged being somewhat arbitrary but reasonable, was "to include on the definition of a term every statement that was checked by over one-third of the subjects in their descriptions of the emotional experiences labeled by that term" (pp. 12–13). As a further check on the robustness of his findings, Davitz asked a number of independent judges to rate the adequacy of each newly created definition (with 1 and 4 denoting most and least adequacy, respectively) based on their own emotional experiences. Thus, a 50-item dictionary of emotional meanings was born.

For our purposes, the two most noteworthy items are anger and hate. Considering the profiles of hate and anger side by side indicates that the participants understood them as largely overlapping, if not analogous, phenomena. Both anger and hate were characterized by a pattern of muscle tension, gastrointestinal discomfort, quickening pulse and high blood pressure, feelings of being "overcharged," an impulse "to strike out . . . kick, or bite" (pp. 35, 65), a sense of being overwhelmed or gripped by the situation, the experience of having one's attention fixed on one thing, and thoughts of revenge (see also McKellar, 1950). On the whole, hate looked remarkably similar to anger. This fits well with Russell and Fehr's (1994) more recent report that people view hate as one of the more prototypic subcategories of anger.

Davitz's pattern of findings is also largely consistent with Fitness and Fletcher's (1993) more recent research on emotions in close relationships. In Fitness and Fletcher's Study 1, a group of 160 married participants were asked to recall the various details of a recent episode of hate, love, anger, or jealousy toward their partner (20 men and 20 women were randomly assigned to each emotion); the participants also answered a series of follow-up questions that probed the physiological symptoms that went with these experiences and the remembered evaluations of the triggering events on a variety of dimensions (e.g., pleasantness, predictability, number of perceived obstacles associated with each event, perceived control). The participants' responses for hate and anger revealed considerable overlap, but there were consistent differences as well. For example, hate and anger were both characterized by the similar incidence of "tight muscles" and identical incidence of "tight stomach," "agitation," and "heart palpations" (in opposition to love or jealousy, for which no palpations were reported).

Similar results obtained when Fitness and Fletcher (1993, Study 2) asked the participants to base their descriptions not on actual incidents from their marital lives but on "hypothetical accounts of the most typical love, hate, anger, or jealousy incidents that they could imagine occurring in a marital relationship" (pp. 943–944). Indeed, hate and anger were sufficiently close to be repeatedly confused on the follow-up Study 3. This time Fitness and Fletcher asked the participants to read a series of descriptions of marital interactions and impute one of eight possible emotions (hate, anger, jealousy, worry, love, happiness, relief, and pride) to the story protagonists. The participants were randomly assigned to four "information" conditions. In Condition 1, they received only a very basic description of the emotion-triggering event; Conditions 2 and 3 supplemented this basic description with the information about the characters' evaluations (the appraisal condition) or emotion-prototypical symptoms, urges, and behaviors (the prototype condition), respectively (the information was derived from the participants' accounts in Study 1); Condition 4 brought all of that information together.

For our purposes, the most interesting finding of Study 3 concerns the asymmetry in the participants' tendency to confuse hate and anger when one or the other was the target emotion. When hate was the target emotion, the participants appeared to be as likely to see it as a case of anger as one of hate in both the prototype and the appraisal conditions. Even when, as in the all-information condition, the participants evinced a statistically significant tendency (at the 54% rate) to identify hate accurately, fully 32% thought that anger was the most fitting choice. But the participants hardly ever made the reverse error of characterizing anger as hate in either of the above-mentioned conditions when anger was the target emotion. And anger was never misidentified as hate in the all-information condition. As Fitness and Fletcher (1993) pointed out, this is just the pattern of findings one would expect if the lay concept of hate represented a variation of the more generic anger script.

Contrary to accounts of hatred that emphasize intensity (e.g., Reber & Reber, 2002), there was no difference between anger and hate with respect to intensity (see also Fitness, 2000). And though significantly more hate than anger episodes were reported in the longer-duration categories of days and weeks, 37 % of the participants reported that their hate episodes lasted for only a few seconds or minutes, suggesting that what people ordinarily describe as hate may be experienced both as an episodic emotion and as an episodic disposition. In sum, neither duration nor intensity appears to be criterial to the lay meaning of hate. From the standpoint of lay psychology, hate stands for something other than a stronger, longer-lasting anger. What does it stand for, then? The answer seems fairly consistent across the few available studies.

Going back to Davitz's (1969) study, notwithstanding the many similarities between hate and anger, hate was ascribed some unique features of

its own. Most notably, hate involved "a sense of being trapped, closed up, boxed, fenced in, tied down, inhibited" (also characteristic of depression and frustration), a feeling that "it all seems bottled up inside of me" (also characteristic of depression), and a perception of "the world . . .[as] no good, hostile, unfair" (Davitz, 1969, p. 35). Also, unlike anger, hate was described as "more an 'inner' than an 'outer' feeling." Similarly, in Fitness and Fletcher's (1993) report, hate was more likely to be characterized by "a sense" of weakness, inefficacy, and insurmountable obstacles (but it was *not* accompanied by attributions any more global or personalized than those associated with anger). A similar pattern emerges from Fitness's (2000) later study that examined anger in the workplace. In light of her work on marital emotions, Fitness hypothesized that (a) hate involves anger and that (b) insofar as hate involves a self-perception of powerlessness, hate is more likely when the transgressor is a superior than when he or she is a subordinate or a coworker.

Both hypotheses were corroborated by the participants' reports. As expected, there was a significant gap between the proportions of subordinates (45%) and superiors (71%) who confronted the objects of their anger. The majority of those angered by their superiors reported that they failed to defy their apparently insulting treatment "because they feared the consequences of expressing their feelings to a more powerful offender" (Fitness, 2000, p. 155). There were also negative correlations between hate intensity and perceived self-power. That is, perceiving oneself as having little relative power in an interpersonal conflict was associated with the self-attribution of hate. The results of this investigation led Fitness to construct the following hate-dominated "anger script" which is typical of a lower-power worker: The lower-power workers "are likely to become angry over unjust treatment by higher power workers . . . They experience moderate to high levels of hate for offenders, especially if the offenses involved humiliation, and their immediate reactions involve withdrawal" (p. 159). Like Fitness (2000), McKellar (1950) noted that power asymmetry and, consequently, the "unexpressed hostile emotion" (p. 109) felt by the lower-power person toward the higher-power abuser are central features of a prototypic "hate episode":

> In eleven of the cases studied, the other individual had higher status than the subject. Apart from the possibilities of envy and jealousy it was evident that these relations were such as to render any really satisfying expression of experienced anger inexpedient, and to diminish the possibilities of successful defense of oneself, one's status and values. (p. 109)

McKellar (1950) observed the same apparent lack of "successful defense" in many of the remaining cases. In these cases, however, anger remained unexpressed not by virtue of the opponent's superior status but rather because of his or her "dominant personality" (p. 109). He interprets this finding as follows:

> Apart from the effect of such a relation on the self-esteem and security of the subject, it is reasonable to suppose that the experience of being powerless to make effective retaliation, to express anger and defend oneself against a more powerful individual, favours the development of lasting hostility rather than mere momentary anger. (p. 109)

An alternative interpretation, one that seems more in line with the work of Davitz (1969), Fitness and Fletcher (1993), and Russell and Fehr (1994), is, of course, that "the experience of being powerless to make effective retaliation . . . against a more powerful individual" represents both an important part of the lay meaning of hate as well as one aspect of its eliciting conditions.

One could say, then, that when people are asked to report hate, they are essentially reporting anger, albeit a particularly helpless, ineffectual, inhibited, "too risky to stick my neck out" kind of anger. This supposition is consistent with the fact that humiliation (or something like it) appears to be the most commonly acknowledged antecedent within the hate script.

For example, the psychoanalytically oriented clinician Fred Pine (1995), whose primary interest lies in a type of jealous revenge-oriented hate that he sees as common among women, traces the origins of this response to a child's experience of being an impotent spectator or an object of her mother's wrath. According to Pine, the defining feature of such an experience is feeling "reduced and objectified and the helplessness to affect it" (p. 106). One interesting aspect of Pine's analysis is its intertwining of clinical case studies with examples derived from some distinguished literary works, such as Euripides's "Medea" or Balzac's "Cousin Bette." He argues that being diminished and remaining helpless to defy one's all-too-powerful opponent is the pattern that lies at the root of both literary and clinically based instances of revenge-driven hate:

> it is of interest that the features of being treated like an object and reduced in one's personhood, as well as a perceived or real helplessness to affect this characterize the position of these women [the hate-prone literary heroines] in male-dominated society, just as those features characterize the position of the female patients as children. (Pine, 1995, p. 106)

The theme of hate as the result of being helplessly "reduced in one's personhood" is also prominent in the work of McKellar (1950). McKellar identified two factors that appeared to be "favourable to the development of hostile attitudes" (p. 109), including hate. The first factor concerned the association between hate and "unexpressed hostility." The second factor concerned the nature of the negative experiences that trigger such hostile responses. Personal humiliation was the single largest category of such experiences. "Physical pain" and "threat to values" were tied for the second place, and "physical pain to another person" (p. 109) came in last. Humiliation

and physical pain were the two largest categories for the female and male subgroups, respectively. Moreover, if we consider that the physical pain was usually experienced in connection with being dominated and bullied by a stronger or more aggressive individual, it is likely that such a pain occurred in the context of a humiliating interaction; thus, McKellar's method of classification may have actually underestimated the prevalence of humiliation as the trigger of hate.

The studies of Fitness and Fletcher (1993) give credence to this summary. Their participants' reports revealed that, in contrast to anger, hate-eliciting events were evaluated as more unpredictable, effortful, and less amenable to personal control (hate experiences were associated with greater helplessness and a more negative self-view). Anger incidents were generally elicited by being treated "unfairly," whereas hate incidents were said to follow from being "most often elicited by the perception that the subject had been badly treated, unsupported, or humiliated by the partner" (Fitness & Fletcher, 1993, p. 945). Fitness's (2000) study of anger in the workplace also hypothesized that "humiliating anger-eliciting events will elicit more intense hate than non-humiliating events" (p. 150). The hypothesis has been largely confirmed.

Before drawing out the full implications of these studies, we put them in the context of three conceptual viewpoints, neither of which appears to be specifically about hate.

Essay 1 of Nietzsche's (1969) *On the Genealogy of Morals* represents an account of the "slave revolt" and the ensuing transfiguration of values that Nietzsche saw as the birth of modern morality. The two main protagonists in Nietzsche's dynamic of ressentiment are "the priests" and "the nobles." The priests are portrayed as feeling diminished by the nobles' superiority in commanding the respect and admiration of the people. For our purposes, the most interesting aspect of Nietzsche's account is that the construct of ressentiment (usually interpreted as an impotent grudge-laden resentment; see, e.g., Hampton, 1988) appears to capture some of the lay meaning of hate, and it intimates that this type of reaction may be appropriately ascribed to both individuals and groups.

The second pertinent viewpoint belongs to Sabini and Silver (1998). As part of their more general inquiry into the intersection of the affective and the moral, these authors presented an argument that there may not be a need for a unique psychology of emotion, at least for a subset of emotions they call "the passions" (e.g., anger, envy, fear). Whereas judgments, sensations, and desires are certainly among constituents of the mind, there is not necessarily a fourth entity called "emotions" that is different from the first three. Rather, they claim, "emotion" is a sort of fictional posit that we invoke in two types of situations: (a) the situations when we "do things that we know we shouldn't because we are overwhelmed by desire" (in respect to anger, the *indulged defiance* condition) and (b) the situations in which

"we . . . find ourselves devoting cognitive and other resources to preparations to act even though we are quite certain we will not act and we would prefer not to act" (p. 136; in respect to anger, the *inhibited defiance* condition).

For our purposes, the most interesting aspect of Sabini and Silver's (1998) proposal lies in the second condition, which they illustrate with a case of "anger" directed at one's superior. Sabini and Silver asked readers to imagine an untenured professor who is nettled by the self-righteous demeanor of a senior colleague and is contemplating a cutting remark meant to put "the old fool" in his place. It is fortunate that the untenured professor has the presence of mind to realize that directing public insults at his academic betters will have a crippling effect on his budding career; consequently, he decides against lashing out and resigns himself to impotent silence. Suppose further, Sabini and Silver said, that the junior professor is so good at managing his interior life that the mere realization of the imprudence of the "You fatuous old fool"–style diatribe causes him to call off the corresponding desire. However, Sabini and Silver maintained that few people could ever aspire to regulating their desires with such virtuosity. And if the beleaguered assistant professor is like most people, the chances are that there will be a period in his interior life when he will find himself persisting in his desire (and his bodily preparations) to attack the senior colleague, while maintaining the firm awareness that because of his inferior position, no such attack will and should ever take place. It is in this tale of inhibited defiance that Sabini and Silver glimpsed one of the two sets of conditions in which people are likely to diagnose themselves as "feeling angry."

Remarkably, it is this very situation—an inferiority-grounded inhibition of one's desire to attack—that Roseman (1984) viewed as particularly interesting in its failure to accommodate one cardinal feature of a prototypic anger scheme—the perception of one's strength vis-à-vis the perceived transgressor:

> With regard to the emotions directed at another person, it is not difficult to see that one might get angry at someone who caused negative events if one were in a position of strength vis-à-vis that person, but not if one were in a position of weakness. The hypothesized behavioral components of anger (attack) and dislike (distancing) make sense in light of these perceptions. Dislike may be understood as a negative emotional response that, due to weakness, must not be anger. (According to Izard, 1977, p. 331, anger is accompanied by a feeling of power) . . . As with frustration, people who feel anger when seeming to be weak may be powerless to control the negative event but not weak enough to be endangered by feeling or expressing anger. The common observation that people "displace" anger onto targets weaker than themselves is consistent with this view. (pp. 27–28)

Roseman (1984) saw the issue of power as separate from that of legitimacy, which he viewed as another important constituent of a full-fledged

anger response. For Roseman, what makes "dislike" so similar and yet so unlike "anger" is that the person who dislikes finds him- or herself (a) legitimated in the urge to retaliate and yet (b) self-inhibited in his or her desire for retaliation because of the asymmetry in power.

It appears, then, that Nietzsche (1969), Roseman (1984), and Sabini and Silver (1998) all, in their own ways, remarked the conceptual significance of this inhibited defiance phenomenon. And as the data suggest, it is precisely this apparent departure from the everyday anger script that forms the tidied-up summary formulation for the lay meaning of *hate*.

Apparently, in agreement with Roseman (1984), the lay conception of anger has embedded in it the notion of a certain power to express, implement, or indulge one's legitimate defiance of the transgressor, with those instances of legitimate defiance that lack such power (but fit the anger model in other respects) being relegated to a conceptual niche of their own.

In this sense, the findings that tie humiliation (physical pain or threat to values) with unexpressed hostility may be explicated as a two-part dynamic: First, humiliation, physical abuse, and the like are the sort of things that people of superior status (strength, dominance, rank, wealth) are far more likely to do to those below them than vice versa. Second, precisely on account of their inferior status or power, the recipients of this abuse (or perceived abuse) are likely to inhibit what they think to be a legitimate urge to defy the higher-status person, while experiencing the physical and mental preparations that otherwise fit the lay script of anger. Thus, the less powerful person is both more likely to be abused and less likely to find it expedient to defy the abuser, making such a person more prone to encounter the maltreatment scenarios that match the lay prototype of hate. In this view, hate is primarily a bottom-up phenomenon, a poor man's anger, and as such it is likely to be shrouded in secrecy. This formulation of the lay meaning of hate prompts at least two reflections.

First, this analysis suggests that the academic community and the lay public may differ in the way they parse the affective realm or name the resulting affective categories (see also Nabi, 2002). Thus, the type of knowledge structure that would be prompted in the mind of an ordinary study participant by the word *hate* seems to be the very same one that a researcher steeped in the conceptual analysis of Sabini and Silver (1998) and Roseman (1984) would code as prototypic "anger" or "dislike," respectively. Conversely, if our introductory overview is any guide, a lay reference to hate would be likely to prompt in the mind of an emotion theorist a very different idea than the layperson had sought to communicate.

Second, it remains an open question to what extent the very existence of the lay concept of hate is a culturally specific phenomenon and to what extent the application of that concept (with its emphasis on avoidance rather than attack) may affect individuals' actual emotional experiences and behaviors. As far as the issue of cultural specificity is concerned (see

Wierzbicka, 1986, 1992, 1999), it is an intriguing possibility that hate, as understood here, is an exclusive property of only those cultures that link the prototypic impulse toward defiance with an ability and, indeed, a mandate to stand up for oneself, irrespective of the transgressor's social rank. A cultural context that lacks this association and places little emphasis on the weakness–power dimension might not have a precise rendition for the anger–hate distinction at all. However, members of such cultures may be able to "re-experience" their past emotional episodes in terms of our concepts anger and hate, once properly acculturated to our point of view (Russell, 2003).

Stipulation

As the name implies, a *stipulative claim* stipulates or constructs a content, rather than aiming to capture the content already associated with the expression—for example, "for the purpose of our discussion 'existent' means 'perceivable'" (Audi, 1995, p. 186). As Audi pointed out, any explicit delineation of a new technical term (e.g., "bloompy" for "a retired Swiss banker") is necessarily stipulative, but his own example shows that everyday terms may be co-opted for stipulative use as well (see also Russell, 1991, on the role of prescriptive analysis). Unlike self-descriptions or (categorical) meta-descriptions, pure stipulations are self-validating. No such statement can ever be shown to be wrong in the sense of having external evidence available against it; this is the sense of "wrong" that should be reserved for meta-descriptive or self-descriptive claims. The worst thing that can be said about such a statement is that it is bizarre (moreover, it is not very useful). But barring some obviously bizarre uses (e.g., "for the purpose of this discussion, *hate* means a small, furry animal that lives in the attic"), a number of perfectly sensible stipulations can coexist side by side in specifying different sets of events falling within the liberally drawn boundaries of notions such as hate, love, or happiness.

In this view, the claim that hate is "generalized anger" is neither more right or wrong nor directly contradictory to the claim that hate is "a blend of anger and fear" or "a disgust for another person" or "a form of inverse attachment" or "an agitation-free, abiding wish for the destruction of a person or a collective of persons." However, when encountered in the context of a scientific inquiry, the stipulative claim concerning hate does incur the additional obligation of constructing a theoretically interesting category that may guide further research into some (but not necessarily all or only) phenomena that are usually understood with reference to hate. Stipulative claims seem to be a part of what Russell and colleagues called the *prescriptive analysis* (e.g., Fehr & Russell, 1991; Russell, 1991), the analysis aimed at delimiting and investigating a class of events in reference to some of which a particular emotion term may be used.

In our view, the evidentiary standard relevant to the research-guiding component of a stipulative claim is twofold: useful demarcated domain and useful relations to other conceptions (Cook & Campbell, 1979). Convergent evidence would indicate that the stipulated conception identifies a domain of behavior or experience that is sufficiently homogenous to support theoretical advances in understanding the causes and consequences of the stipulated conception (this harkens back to the notion of exposing nature at its joints). Discriminant evidence indicates that the stipulated conception can be distinguished theoretically and empirically from related conceptions. For example, constructing hate as a strong aversion to anything or anyone would be poor on the grounds of convergent validity: The way one is averse to drinking fish oil is not the same way one is averse to visiting a dentist or sitting through a documentary on the future of agrarian science. The phenomena in question are just too nonhomogenous to be subsumed under a single construct. However, constructing hate as fear seems to slight discriminant validity: If hate is fear, how is it different from fear, and why study it under the name *hate*? (This is, of course, not to say that fear cannot be a good explanation of certain paradigmatic cases of hate behavior, from genocide to hate crimes).

It seems that with some modest refinements, the Shand-inspired conception of hate as a syndrome of inverse caring could do especially well on the grounds of construct validity by positing a form of antagonistic response that is fairly homogenous and clearly distinct from what is typically meant by anger, fear, or resentment. In a nutshell, it would stipulate that hate and love are not single emotions or blends of emotions but dispositions to experience many different emotions, depending on the fortunes of those loved and hated. Both caring–attachment ("love") and inverse caring–attachment ("hate") represent syndromes of episodic dispositions that go with motivational orientations tracking and reacting to the fortunes of significant others (individuals or groups who occupy a special place in our lives, positively or negatively). In the absence of tracking (which may range from mild to obsessive), inverse caring would probably correspond closer to what we call *dislike* or *negative identification* (McCauley, 2001) and involve emoting inversely to the news of another's fortunes or misfortunes as such news comes to one's attention without seeking it out.

Though puzzling in their own right, one reason that cases of inverse caring should be of great interest to psychologists is that they may help us to shed light on caring proper or vice versa. Like Sternberg's (2003) triadic model of hate, explicitly motivated by his triadic analysis of love, our stipulation encourages the possibility that there is a general capacity that allows for the tracking of and emotional engagement with the fortunes of others, the capacity whose impairment could undercut love and hate alike.

In our view, love and hate, as stipulated above, are likely to be characterized by the following key features:

1. Both hate and love are likely to be associated with a perception or attribution of a negative or positive essence. Briefly, the idea of essence is the hidden something that makes a living thing what it is (e.g., Atran, 1990; Gil-White, 2001; Keil, 1989). Essence is a more primitive idea than genetics and is better represented as *nature* or *spirit* than as a biological concept.
2. The negative or positive evaluation of the target of hate or love is likely to be linked with a moral judgment. The evaluation may be the product of such a judgment or, as in the case of hate, rooted in envious admiration; the moral judgment may emerge as a rationalization for the pre-existing pattern of (direct or inverse) caring.
3. Loves and hates may vary in the extent to which they exclude any possibility of compassion or ill wishing, respectively. The implication is that patriotism, nationalism, and ethnic group identification are particularly extreme expressions of group love, as genocide may be the ultimate expression of group hate (Chirot & McCauley, in press).
4. People may be expected to differ in how vehemently they endorse or identify (Frankfurt, 1971) with their hates. At one extreme, there is a *faint-hearted hater*, someone who is shocked and ashamed to discover that he or she hates whomever he or she hates and who wishes that things were otherwise. At another pole, there is the *wholehearted hater*, someone who endorses his or her hate completely, appropriates it with pride and even fondness, and makes every effort to nurture it to its fullest capacity. In the latter case, hate becomes "sanctified" (Blum, 1995) or fully integrated into one's value system. The same principle applies to love.

An empirical inquiry into the construct of hatred as inverse caring might proceed by asking people to identify relationships in their lives that include a tendency to track and react inversely to another's fortunes. One could then form and test a set of hypotheses regarding possible correlates, consequences, and causal antecedents that characterize such relationships. Do people essentialize the people or groups they are inversely attached to or care about? Do they see the objects of such caring as morally bad? Do they feel good or bad about having such tendencies? Would they ever (and under what circumstances) relent?

An empirical analysis may also reveal that some or all of the above-mentioned elements of hate do not reliably correlate with one another. For example, it is possible to be saddened by another's good outcomes without taking pleasure in his or her bad outcomes; it is possible to do either one without attributing essence or passing a moral judgment.

Explanation of Ostensively Defined Cases

The fourth possible interpretation of a statement such as "Hate is a self-destructive impulse turned outwards" is that it represents a causal explanation—that is, an attempt to specify a mental mechanism most directly responsible for certain paradigmatic instances of "hate-related" behavior. The general idea is that insofar as such paradigmatic instances indicate the immediate effects of hate, working backward to their (relatively proximate) causes should give us a glimpse of hate itself. As noted earlier, the approach to defining hate through paradigmatic instances is common in political science. One context within which hate is commonly discussed and thought to be especially problematic is that of ethnic conflict. In this context, "Hate is a self-destructive impulse turned outwards" would represent an explanatory claim concerning the psychological underpinnings of ethnic conflict; as such, it could be tested empirically by asking if those engaged in ethnopolitical violence are, indeed, driven by a (sublimated) self-destructive impulse.

As we hinted earlier, one potential problem with the ostensive approach to defining hate is that the psychological underpinnings of the relevant paradigmatic cases may be shown to be very different from one's initial intuitions about what hate is or is not. For example, many so-called hate crimes seem to be committed out of some combination of boredom and a desire to show off before one's group (Baumeister, 1997). Indeed, in analyzing Boston police hate crime files, McDevitt, Levin, and Bennett (2002) concluded that the majority (66%) of the perpetrators were motivated by a desire to escape boredom and get some quick thrills and bragging rights, with targets being selected because they were perceived as "somehow different" (p. 307). Of the remaining 34%, 25% seemed motivated by an anxiety-laden desire to protect their neighborhood and families from what the perpetrators perceived as the onslaught of dangerous outsiders, with the criminal behavior being seen as a form of self-defense instrumental to "convincing" the victims to relocate elsewhere as well as forestalling future "intrusions." The Boston police files sample did, indeed, contain a form of motivation that fits well with the classic Aristotelian (1954) notion of hate (see also Elster, 1999; Beck, 1999), namely acting out of a deeply ingrained belief that the "others" are inherently evil or inferior and ought to be eliminated as such. However, this intensely other-focused motivation or affective orientation accounted for less than 1% of the entire sample; it was held by a single person.

Conceding the validity of these observations makes for some tough choices. It seems that either society must allow that most "hate crimes" (assisted as the selection of the victim may be by categorical negative judgments) are neither directly motivated by nor involve hate as their dominant affect, or its current understanding of hate must be expanded.

To give another example, Gaylin (2003, p. 14) proposed that "true hate" is a form of mental disorder ("Hatred is a severe psychological

disorder"), and he cited terrorist violence as one of the paradigmatic cases in which such true hate may be found in abundance. In fact, Gaylin appeared to believe that anything short of portraying the terrorists as psychologically disturbed is morally irresponsible—"When we assume that at times we feel like a terrorist, we grant the terrorists a normalcy that trivializes a condition that threatens the civilized world" (p. 22). We fully agree that terrorism is evil and presents a tremendous threat. However, some of the most systematic research into the psychology of terrorism, including the detailed German studies of the Baader-Meinhof Gang, have found psychiatric disorder no more common for terrorists than for the general population that the terrorists emerge from (Konrad, 1998). Again, it seems that something has to give. Gaylin and those who favor his psychoanalytic approach should either surrender the notion that terrorism is a paradigm case of hate or be prepared to revise the concept of hate itself.

The very idea that the conceptual boundaries of hate or any other item of folk psychology can be redrawn on the basis of results of an empirical inquiry may seem foreign to some. However, this type of situation is fairly common in advanced natural science (Griffiths, 1997; Putnam, 1975). For example, some of the earliest definitions of *gold* were, no doubt, ostensive in quality. These definitions could then be unpacked into a prototypical characterization of gold as "that hard, malleable, yellow, shiny stuff we call gold." From the standpoint of an early chemist, the lay meaning of gold merely delimited the domain of inquiry; the scientific meaning of gold was to be sought by investigating the chemical structure manifest in the paradigmatic instances picked out by the commonsense concept gold. The discovery that, chemically speaking, gold is an element with atomic number 79 inaugurated a criterial shift such that it was no longer the lay but the scientific meaning that determined the extension of the concept gold, with genuine gold being clearly distinguishable from "fool's gold" and other substances. Thus, the lay meaning was bent to the results of the scientific discovery.

Consider also the concept of *hysteria*. The phenomenon itself seems real enough and was described by Hippocrates around 460 B.C., accompanied by a theory that the hysterical symptoms are due to wanderings of the uterus and, thus, are restricted to women (Hothersall, 1990, p. 13). Indeed, the term hysteria comes from *hysteron*, the Greek word for uterus. So deeply embedded the uterine theory had become within the concept of hysteria that Freud's 1886 paper "On Male Hysteria" was summarily dismissed by some observers as not being about hysteria at all (Hothersall, 1990, p. 240). Yet neither Freud nor future students of psychoanalysis found it problematic that one may dislodge a certain historically influential account of why paradigm cases of hysteria arise while retaining much of the concept's phenomenological content, thus opening themselves to the possibility of revising the meaning of the concept in the direction of whatever psychic mechanisms could best explain these cases. Similarly, our present folk-theoretical com-

mitment to viewing courage as the opposite of cowardice may simply have to give should we discover that fear, in fact, is the motivation behind many garden-variety courageous acts (think of a female patient who drags herself to a feared and painful medical procedure for the fear that otherwise her life or health may be seriously at risk). There are reasons to believe that the same revisionist logic may be usefully applied to (re)construct the meanings of such folk notions as self-deception (Mele, 1997) and humility (Royzman, Cassidy, & Baron, 2003).

With this in mind, consider Gil-White's (2001) recent argument for a biphasic process whereby cultural differences are first moralized (if you don't act or think like me, you are not just "different," you are bad) and then essentialized (the badness is in your blood and in the blood of all those in your ethnic group) to yield ethnic unrest (see also Hirschfeld, 1995, 1996; see Gelman, 2003, for a broad overview of psychological essentialism). With this in mind, one could argue that, should "hate" be defined as "whatever is going on" in, say, the paradigmatic cases of ethnic conflict, and should "moralization–essentialization" prove to be a good filling for the "whatever" part, the conceptual boundaries of hate may need to be revised accordingly.

Note, however, that what makes this type of criterial shift possible in the case of concepts such as gold or water is not only the overwhelming consensus on what the paradigm instances of these categories are like but also the fact that the domains of inquiry delimited by such instances just happened to match the underlying natural kinds—H_2O in the case of water, a chemical element with atomic number 79 in the case of gold. It is not clear that nominating, say, *ethnic conflict* as a paradigm instance of hate meets either condition. The term ethnic conflict, as currently used, may not pick out a homogenous enough category of natural phenomenon, but there is also no guarantee that it represents the one and only incontestable paradigm case of hate the way that potable, wet, transparent stuff called "water" represents the one and only paradigm case of water. The meaning of hate unearthed in this way may fail to fit other equally compelling paradigm cases identified via other ostensive definitions (e.g., hate as seen in the case of terrorist violence vs. the "admiring hate" of the Sallieri–Mozart type of relationship). Of course, some of these cases, such as the Nazis' genocidal campaign against the Jews, will remain especially central to a cultural consciousness of what hate looks like. It seems that, whenever possible, such core cases should be given preference over their more peripheral counterparts. Still, one may never reach a point when one can say that hate is "really" moralization–essentialization (or something else) with the same tone of confidence with which one can state that gold is "really" an element with atomic number 79 or water is "really" H_2O. Nevertheless, psychologists can recruit all of the above paradigm cases to explore, under the name of hate, the causes of a variety of problems nominally linked to hate from ethnic cleansing to international terrorism, though there is no guarantee that the

uncovered antecedents of all or any of these problems will match either the dominant lay meaning of hate (see the previous section) or any of the expert definitions offered in the opening pages of this chapter.

Platonic Insight

It would seem that the four types of claims presented thus far should be able to fit comfortably any conceivable conceptualization of hate, whether offered in a verbal dispute or as a part of a theoretical review. However, it became increasingly clear to us that our taxonomy contained a serious oversight. In arguing that people who commit acts of mass violence cannot be understood in terms of everyday psychology, Gaylin (2003) made a distinction between true hate and the more pedestrian phenomenon that goes by that name in everyday use: "We are capable of transient extremes of rage that we call hatred, but the true haters live daily with their hatred . . . When we confront the true hater, he frightens us" (pp. 4–5). More specifically, Gaylin goes on to state that true hate is a form of mental illness, characterized by displacement and paranoid ideation. What type of claim is this?

First, Gaylin (2003) does not appear to be venturing anything like a (falsifiable) causal hypothesis concerning the motivational underpinnings of a certain type of nominally hate-associated behavior. Rather, he begins by stating what hate is and then goes on to aver a link between that and mass violence. Second, his use of the qualifier "true" in the early pages of the book signals that he does not intend his conception of hate to be taken as a mere stipulative construction (as in "for the purpose of this discussion, I posit *hate* to be . . ."), but as a statement of some deeper truth. The most obvious interpretation would be that Gaylin's intent is to offer a categorical meta-description—that is, to capture the conventional (English-language) meaning of hate. After all, when one says to a foreigner that the "true" or "right" meaning of a word is slightly different from the meaning she appears to be imputing to it, one should be understood as doing nothing more than instructing her about the (historically derived) linguistic conventions governing the use of that word at a given time and place. But this is clearly not what Gaylin was after; he acknowledged that the conventional lay meaning of hate is extreme anger or rage, but he went on to affirm that true hating is different from what we conventionally understand it to be.

In the annals of Western history, there is one major model of word meaning within which such an argument would make perfect sense. We are speaking, of course, of Plato's theory of the Forms. The Platonic Forms or Ideas are eternal, metaphysical essences that permeate and condition the world of sensory experience. Though we may have no direct access to them at the moment, our souls ostensibly knew them as part of their precarnate existence, and it is this implicit memory that is guiding our eventual

extension of some common name (*love, piety, justice, good*) to a set of seemingly diverse phenomena. In this view, all things called by the same name must share an essence that is an expression of the corresponding Platonic Idea; thus, all the variegated instances of love, be it the love of God or the love of one's neighbor, the love of a mother for her child or the love of a boy for a girl, the love of a philatelist for his stamps or the love of a patriot for her country, must share a core defining feature, which is but a spatio-temporal expression of the Platonic Idea of Love, the capturing of which will give us one and only one true conception of love.

Within the Platonic system, it makes sense that many people think love to be "X" while declaring that "real love" is something else entirely, with the external fact guaranteeing the truth of one's pronouncement being the objectively extent *Idea* of Love. In his discussion of love, Shand (1920) seemed to be following this very path when he averred that it was poets (primarily the classic British poets of the past 4 centuries), not philosophers, who had a true "insight" into the "nature of love" (p. 54). That is, Shand apparently believed that, irrespective of its conventional meaning, love carries some objective essence, true for all times and places, that some persons (those blessed with the powers of Platonic insight) may fathom better than others. Considering that Shand considered love and hate as structurally equivalent opposites, he must have held a similar view of hate. We call this type of apparently essentialist claims "Platonic" not because we believe that either Gaylin or Shand knowingly subscribed to Plato's theory of the Forms, but because this theory represents the only major intellectual system within which such claims can be rendered fully intelligible. (In one of the long editorial notes to his father's magnum opus [J. Mill, 1878], J. S. Mill complained that "as half the conceptual world are Platonists without knowing it, hence it also is that in the writings of so many psychologists we read of the conception or the concept of so and so; as if there was a concept of a thing . . . other than the ideas in individual minds" [p. 237]. Mill apparently believed that the less conscious we are of our [bad] metaphysics, the more likely we are to be trapped by it.) Granted that neither Gaylin nor Shand were conscious Platonists, how are we to explain their apparent Platonic leanings?

One answer may be that the situation is due to a sort of cultural inertia, that is, a tendency toward the continued acceptance of some cultural idea or practice while no longer giving any credence to the background belief system (i.e., Plato's metaphysics) that originally gave it meaning (see MacIntyre, 1981). An alternative account comes from a recent work of Gelman, Hollander, Star, and Heyman (2000; see also Gelman & Heyman, 1999), which suggests that people are poised to essentialize category names that are lexicalized as common nouns. One could speculate, then, that insofar as the name of the superordinate category such as *emotion* is itself a common noun, all instances of emotion are taken to share a single, objectively

discoverable essence, with the result that its subordinate types (love, hate) are understood to be essence holding as well. Assuming that this tendency is linguistically pervasive, it could produce a pattern of thinking about emotion and other folk concepts that fits well with the Platonic worldview.

Whichever explanation proves to be correct, we believe that Platonic musings about one "true," "real," or "essential" nature of hate are scientifically unhelpful and ought to be set aside to make room for the more promising meta-descriptive, explanatory, and stipulative approaches. As recovering Platonists ourselves, we are conscious of the difficulty, but we submit that the resulting gains in clarity and mutual understanding will be well worth the effort.

CONCLUSION

In sum, a statement such as "Hate is a self-destructive impulse turned outwards" lends itself to at least four distinct interpretations. It could be (a) a statement of a categorical meta-opinion, one specifying the predominant meaning of hate within a given linguistic community (e.g., the sense of hate as "personalized anger" among some emotion theorists or the sense of hate as "inhibited defiance" among English-speaking laypersons); (b) a stipulation of a putatively useful theoretical construct (e.g., hate as a syndrome of inverse caring); (c) an explanation of the cause or causes of some nominally hate-related behaviors (e.g., the moralization–essentialization hypothesis); or (d) a statement laying an implicit claim to some form of Platonic insight intent on capturing the objective essence of hate.

There is no reason to expect that any single formulation of hate should be able to satisfy the demands posed by all of these interpretations. Hate as understood by the majority of ordinary English speakers need not be the same phenomenon that motivates "hate crimes" or stirs the fires of ethnic conflict, which, in turn, need not be analogous to the hate that may be posited as a theoretically interesting construct by those with an interest in investigating certain forms of antagonistic motivation.

Thus, the question "What is hate?" must remain largely unintelligible unless the questioner is willing to specify which of the above interpretations he or she has in mind. In the absence of such a specification, it is very hard to interpret either the question or the putative answers. It is this lack of specificity that makes it so difficult to evaluate and compare the various formulations of hate cited in the opening pages of this chapter. We can see now that the disparities among these formulations can be characterized in at least two ways. First, there are substantive differences—is hate a syndrome or an emotional blend? What kind of a syndrome or a blend? Second, there are differences in the nature of the claim being made—is the formulation being offered a stipulation, a meta-description, a causal

hypothesis, a Platonic insight, or some combination of these? It will be helpful if future students of hate can be more direct about which type of claim they are making.

Sternberg and Grigorenko (2001) noted recently that many a dispute within psychology may revolve around "false oppositions" that arise when investigators study what is essentially the same construct from divergent methodological and theoretical viewpoints, each believing that his or her approach is the uniquely correct one. Our analysis suggests that false oppositions may also arise when investigators study what are effectively different constructs—constructs specified in accordance with different types of claims—that, nevertheless, operate under the same name.

Although we have distinguished four different ways of thinking about hate, there should be no implication that these are mutually exclusive. Our point is simply that they need not converge in all (perhaps, many) possible cases. A scholar can concede that the way that most people give meaning to, say, *humility* is also the best bet for a stipulated definition that will support evidence of convergent and discriminant validity, which, in turn, offers the best explanation for the relevant behavioral data to be explained with reference to humility. Indeed, our own experience with hate has been a kind of trajectory of interpretations, in which we began with a personal understanding of what hate means, projected this to a hypothesis about what most people mean by hate, hoped that the meta-opinion of hate would provide an explanation of genocide and other forms of intergroup violence, stipulated hate as inverse caring, and recognized belatedly in ourselves a tendency to see meta-opinion, explanation, and stipulation as shadows of a Platonic notion of what hate "really" is.

Our analysis leaves us uncertain about the much-cited link between hate and intergroup violence such as genocide, ethnic riots, or hate crimes. If hate is defined ostensively through paradigm cases of armed conflict and killing, then the notion that hate is responsible for mass violence is a tautology. Conversely, if hate is to be spelled out in terms of its lay meaning, as a form of inhibited defiance, or in terms of a stipulated meaning, for example, as a syndrome of inverse caring, then the empirical evidence for the link between hate and intergroup violence remains to be seen. That is, the very status of hate as a progenitor of evil rests on a prior conceptual decision about which phenomenon one is willing to probe under the heading of hate and which one will opt to see as being "not about hate at all."

Most generally, the four ways of thinking about hate may be helpful in thinking about many other psychological notions that are rooted in lay experience and everyday language. Disgust, shame, humility, and pride; frustration and aggression; value and virtue; the Big Five personality traits—all grow out of lay meanings. We expect that many such items of folk psychology, and the empirical literatures surrounding them, can benefit from the analysis undertaken in this chapter in the service of explicating hate.

REFERENCES

Aristotle. (1954). *The rhetoric and the poetics of Aristotle* (W. R. Roberts, Trans.). New York: Modern Library. (Original work written ca. 340 B.C.)

Atran, S. (1990). *Cognitive foundations of natural history*. Cambridge, England: Cambridge University Press.

Audi, R. (1995). *The Cambridge dictionary of philosophy*. Cambridge, England: Cambridge University Press.

Averill, J. (1982). *Anger and aggression: An essay on emotion*. New York: Springer Verlag.

Averill, J. (1991). Emotions as episodic dispositions, cognitive schemas and transitory social roles: Steps towards an integrated theory of emotion. In R. Hogan (Ed.), *Perspectives in personality: Vol. 3A. Self and emotion* (pp. 139–167). London: Jessica Kingsley.

Baumeister, R. F. (1997). *Evil: Inside human violence and cruelty*. New York: Freeman.

Baumeister, R. F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology*, 59, 994–1005.

Beck, A. T. (1999). *Prisoners of hate: The cognitive basis of anger, hostility, and violence*. New York: HarperCollins.

Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge, MA: MIT Press.

Blum, H. P. (1995). Sanctified aggression, hate, and the alteration of standards of values. In S. Akhtar, S. Kramer, & H. Parens (Eds.), *The birth of hatred. Developmental, clinical and technical aspects of intense aggression* (pp. 17–37). Northvale, NJ: Jason Aronson.

Chirot, D., & McCauley, C. (in press). *Genocide*. Princeton, NJ: Princeton University Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. New York: Rand McNally.

Darwin, C. (1998). *The expression of the emotions in man and animals*. New York: Oxford University Press. (Original work published 1872)

Davitz, J. (1969). *The language of emotion*. San Diego, CA: Academic Press.

Descartes, R. (1989). *On the passions of the soul* (S. Voss., Trans.). Indianapolis, IN: Hackett. (Original work published 1694)

Dozier, R. W., Jr. (2002). *Why we hate: Understanding, curbing, and eliminating hate from ourselves and our world*. Chicago: Contemporary Books.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169–200.

Elster, J. (1999). *Strong feelings: Emotion, addiction, and human behavior*. Cambridge, MA: MIT Press.

Fehr, B., & Russell, J. A. (1991). The concept of love viewed from a prototype perspective. *Journal of Personality and Social Psychology*, 60, 425–438.

Fitness, J. (2000). Anger in the workplace: An emotion script approach to anger episodes between workers and their superiors, co-workers and subordinates. *Journal of Organizational Behavior, 21*, 147–162.

Fitness, J., & Fletcher, G. J. O. (1993). Love, hate, anger, and jealousy in close relationships: A prototype and cognitive appraisal analysis. *Journal of Personality and Social Psychology, 65*, 942–958.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*, 5–20.

Fredrickson, B. L., & Branigan, C. (2001). Positive emotions. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: Current issues and directions* (pp. 123–151). New York: Guilford Press.

Frijda, N. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Gaylin, W. (2003). *Hatred: The psychological descent into violence*. New York: Public Affairs.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. London: Oxford University Press.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*, 489–493.

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 39, pp. 201–263). San Diego, CA: Academic Press.

Gil-White, F. (2001). Are ethnic groups biological "species" to the human brain? *Current Anthropology, 42*, 515–554.

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.

Hall, G. S. (1898). A study of anger. *The American Journal of Psychology, 10*, 516–591.

Hampton, J. (1988). Forgiveness, resentment and hate. In J. G. Murphy & J. Hampton (Eds.), *Forgiveness and mercy* (pp. 35–87). New York: Cambridge University Press.

Herek, G., Gillis, J. R., & Cogan, J. C. (1999). Psychological sequelae of hate-crime victimization among lesbian, gay, and bisexual adults. *Journal of Consulting and Clinical Psychology, 67*, 945–951.

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition, 54*, 209–252.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Hothersall, D. (1990). *History of psychology* (2nd ed.). New York: McGraw-Hill.

Hume, D. (1980). *A treatise of human nature*. Oxford, England: Oxford University Press. (Original work published 1739–1740)

Kaufman, S. J. (2001). *Modern hatreds: The symbolic politics of ethnic war*. Ithaca, NY: Cornell University Press.

Keil, F. C. (1989). *Concepts, kinds, and human development*. Cambridge, MA: MIT Press.

Kolnai, A. (1998). The standard modes of aversion: Fear, disgust and hatred. *Mind, 107*, 581–595.

Konrad, K. (1998). Ideology and rebellion: Terrorism in West Germany. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, states of mind* (pp. 43–58). Washington, DC: Woodrow Wilson Center.

Kressel, N. J. (2002). *Mass hate: The global rise of genocide and terror*. Cambridge, MA: Westview Press.

Levin, J., & McDevitt, J. (1993). *Hate crimes: The rising tide of bigotry and bloodshed*. New York: Plenum Press.

MacIntyre, A. (1981). *After virtue*. Notre Dame, IN: University of Notre Dame Press.

McCauley, C. (2001). The psychology of group identification and the power of ethnic nationalism. In D. Chirot & M. Seligman (Eds.), *Ethnopolitical warfare: Causes, consequences, and possible solutions* (pp. 343–362). Washington, DC: American Psychological Association.

McCauley, C. (2002). Psychological issues in understanding terrorism and the response to terrorism. In C. Stout (Ed.), *The psychology of terrorism: Volume 3. Theoretical understandings and perspectives* (pp. 3–30). Westport, CT: Praeger Publishers.

McDevitt, J., Levin, J., & Bennett, S. (2002). Hate crime offenders: An extended typology. *Journal of Social Issues, 58*, 303–317.

McKellar, P. (1950). Provocation to anger and development of attitudes of hostility. *British Journal of Psychology, 40*, 104–114.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences, 20*, 91–136.

Mill, J. (1878). *Analysis of the phenomena of the human mind* (Vol. 1, 2nd ed.). London: Longmans, Green, Reader, and Dyer.

Nabi, R. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition & Emotion, 16*, 695–703.

Nietzsche, F. (1969). *On the genealogy of morals*. New York: Vintage Books.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition & Emotion, 1*, 29–50.

Pine, F. (1995). On the origin and evolution of a species of hate: A clinical–literary excursion. In S. Akhtar, S. Kramer, & H. Parens (Eds.), *The birth of hatred: Developmental, clinical and technical aspects of intense aggression* (pp. 105–132). Northvale, NJ: Jason Aronson.

Power, M. J., & Dalgleish, T. (1997). *Cognition and emotion: From order to disorder*. Hove, East Sussex, England: Psychology Press.

Putnam, H. (1975). The meaning of "meaning." In H. Putnam, *Philosophical papers: Vol. 2. Mind, language, and reality* (pp. 215–271). Cambridge, England: Cambridge University Press.

Rachman, S. J. (1978). *Fear and courage*. San Francisco: W. H. Freeman.

Reber, A. S., & Reber, E. (2002). *The Penguin dictionary of psychology* (3rd ed.). New York: Penguin Books.

Roseman, I. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Emotions, relationships and health* (Vol. 5, pp. 11–36). Beverly Hills, CA: Sage.

Royzman, E., Cassidy, K. W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology, 7*, 38–65.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574–586.

Russell, J. A. (1991). Natural language concepts of emotion. In R. Hogan (Ed.), *Perspectives in personality: Vol. 3A. Self and emotion* (pp. 119–137). London: Jessica Kingsley.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology, 67*, 186–205.

Ryle, G. (1949). *The concept of mind*. London: Harper & Row.

Sabini, J., & Silver, M. (1998). *Emotion, character, and responsibility*. New York: Oxford University Press.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology, 73*, 902–922.

Shand, A. F. (1920). *The foundations of character* (2nd ed.). London: Macmillan.

Solomon, R. (1977). *The passions*. New York: Anchor Books.

Spinoza, B. (1985). Ethics. In E. Curley (Ed.), *The collected works of Spinoza* (Vol. 1, pp. 408–617). Princeton, NJ: Princeton University Press. (Original work published 1677)

Sternberg, R. (2003). A duplex theory of hate and its development and its application to terrorism, massacres, and genocide. *Review of General Psychology, 7*, 299–328.

Sternberg, R. J., & Grigorenko, E. L. (2001). Unified psychology. *American Psychologist, 56*, 1069–1079.

Walton, D. N. (1986). *Courage: A philosophical investigation*. Berkeley: University of California Press.

Wierzbicka, A. (1986). Emotions: Universal or culture-specific? *American Anthropologist, 88*, 584–594.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science, 16*, 539–581.

Wierzbicka, A. (1999). *Emotions across languages and cultures*. New York: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.