12 SUMMARY

In contrast to the vague notions of coherence used by many ethical theorists, the theory of coherence as constraint satisfaction can provide a detailed and computable model of how different kinds of coherence can contribute to ethical judgments. Deliberative coherence involves choosing actions and goals on the basis of their coherence and incoherence with other actions and goals. Deliberative coherence is essential to ethical judgments, but deductive, explanatory, and analogical coherence can also contribute. This theory of ethical coherence is intended to be both descriptive of how people make ethical judgments and prescriptive of how they should. Political judgments involving the justification of the state and the choice of a kind of state are based on ethical coherence, particularly on deliberative coherence with respect to the goals of freedom, flourishing, and fairness.

6 *Emotion*

Like most philosophical and psychological writings about inference, my discussion of coherence has so far ignored the important role of emotion in human cognition. This chapter presents a theory of emotional coherence and describes its implementation in a computational model that has been applied to interpersonal trust and other important psychological phenomena that involve both inference and emotion, including empathy and nationalism. The theory and model are then extended to encompass "metacoherence" and the emotional impact of overall assessments of coherence relevant to understanding beauty, humor, and cognitive therapy.

1 THE IMPORTANCE OF TRUST

When Jimmy Carter ran for President in 1976 in the wake of Watergate, he told the voters, "You can trust me." After Tony Blair was elected Prime Minister of England in 1997, he responded by telling the voters, "You have put your trust in me. I will not let you down." In elections, politicians often try to convince the voters that they are more trustworthy than their opponents, and incumbents work to maintain the trust of their constituents (Bianco 1994, Fenno 1978). The political importance of trust is

also seen in international relations, where opportunities for agreement and cooperation can be missed because of distrust between nations (Larson 1997). When U.S. Secretary of State Albright visited Israel in Se7ptember 1997, she told the Israelis and the Palestinians that in order to overcome their conflicts, they need to establish a climate of trust.

Fukuyama (1995) has recently emphasized the economic importance of trust, particularly in the organization of the workplace. Manufacturers such as Toyota have been successful in increasing quality and productivity in part because they have established factories in which there is trust between workers and managers. Business partnerships and deals are much easier to arrange when there is trust rather than distrust among the participants. The sociological importance of trust has been noticed by such writers as Gambetta (1988), Kramer and Tyler (1996), Lewis and Weigert (1985), and Misztal (1996).

Everyday life would be impossible without trust. Dealing with spouses, partners, friends, and myriad other people with whom we interact is immensely facilitated when we can trust them; suspicion and distrust make interactions unpleasant and often unsuccessful. Social psychologists such as Deutsch (1973) and Holmes (1991) have described the central role played by trust in interactions and relationships. Many mundane decisions, such as hiring a baby-sitter to look after one's children, are largely decisions whether the person hired can be trusted.

The concept of trust is also philosophically important. In political philosophy, a focus on trust as both a passion and a policy provides an alternative to the contractarian emphasis on rational egoism (Dunn 1993). In real life prisoner's dilemma situations, for example, the decision to cooperate or defect is typically based not on the abstract logical considerations of game theory, but on informed and

emotional decisions about whom to trust (Deutsch 1973). Trust is also important for epistemology, because knowledge is not just a matter of an individual working out everything alone. Rather, especially in modern science, the development of knowledge depends crucially on collaboration and communication, both of which require epistemic trust (Hardwig 1991, Thagard 1999).

Trust is a matter of both inference and emotion. The inference of whether to trust people depends on combining many kinds of information about them into a coherent system that generates a positive or negative emotional reaction to them. Leaving emotion aside for the moment, I will review the coherence theory of inference developed in chapters 2 and 3.

2 COHERENCE-BASED INFERENCE

The conception of inference familiar since Aristotle and the Stoics is based on formal logic, according to which we infer a conclusion from a set of premises in accord with rules of inference. Probably the most frequently applied rule is modus ponens: if *p* then *q*; *p*; therefore *q*. But this view of inference is problematic, because, as Harman (1986) and Minsky (1997) pointed out, we should not always infer *q* from *If p then q* and *p*, since sometimes it is more appropriate to abandon *p* or *If p then q*. To take a trust-related example based on the 1997 influx of Czech Gypsies to Canada and England, suppose you have the common stereotype that Gypsies are dishonest. You might be prone to make the following inference:

If Karl is a Gypsy, then Karl is dishonest.

Karl is a Gypsy.

Therefore, Karl is dishonest.

But what if Karl has just returned intact the wallet that you dropped on the street? Then you have reason to believe that Karl is not dishonest, so you might consider revising one or both of the premises that led to the conclusion that Karl is dishonest. Alternatively, you might hypothesize that Karl really is dishonest and that he returned your wallet only to ingratiate himself with you. The inference you make about Karl's dishonesty will depend on how the conclusion and the premises generating it fit with everything you know. Inference is a matter of coherence.

Explanatory coherence is highly relevant to trust, because it is the mechanism by which we infer the motives and plans of another. In the Gypsy example, the evidence that Karl returned your wallet led you to consider different hypotheses that would explain why he returned it. In general, you will tend to trust people when you can infer from what you know about them that they have motives and plans that contribute to your own goals. Hypotheses about motives and plans need to be evaluated with respect to how well they explain the evidence about someone in comparison with hypotheses about other motives and goals.

A more automatic kind of inference is performed as the result of conceptual coherence, in which the elements are concepts representing, in the interpersonal case, attributes of people such as stereotypes, traits, and behaviors. The positive constraints arise from observations and positive associations, for example, the prejudiced association that Gypsies are dishonest. Negative constraints arise from negative associations, for example, between returning money and being dishonest. Kunda and Thagard (1996) showed that many psychological phenomena involving impression formation and the application of social stereotypes can be understood in terms of conceptual coherence. Conceptual coherence is relevant to trust when it produces

inferences about stereotypes, traits, and behaviors based on positive and negative associations. We tend to trust people who have characteristics, such as honesty, that are associated with trustworthiness, while we distrust people who have contrary traits, such as mendacity.

Analogical coherence differs from the explanatory and conceptual kinds in that it is primarily based not on general hypotheses or concepts, but on particular cases. In analogical inference, we infer something about a person or situation on the basis of its similarity with other persons or situations. The relevance to trust is that we tend to come to trust people who are similar to other people that we trust, while distrusting people who remind us of people whom we have learned to distrust.

Deliberative coherence involves deciding what to do on the basis of interrelations of competing actions and goals (chapter 5). The actions and goals are elements that are positively constrained by facilitation relations (e.g., an action facilitates a goal) and are negatively constrained by incompatibility relations (e.g., when two actions cannot both be performed). The decision to trust someone involves considering the implications of all that you know about the person that is relevant to the accomplishment of your goals. Deductive and perceptual coherence seem only tangentially related to trust judgments.

A major problem for the theory of coherence-based inference concerns how the different kinds of coherence can be integrated with each other (Thagard and Kunda 1998). How do explanatory, conceptual, and analogical coherence interrelate? How do we integrate possibly incompatible conclusions based on different kinds of coherence? From a coherentist perspective, there is only one rule of inference: accept a representation if and only if it coheres maximally with the rest of your representations. A partial answer to the question of integration, as

or her. This is obviously not intended to be a general theory of emotions, which involve much more than positive and negative valence.

The basic theory of emotional coherence can be summarized in three principles analogous to the qualitative principles of coherence stated in chapter 2:

- Elements have positive or negative valences.
- Elements can have positive or negative emotional connections to other elements.
- The valence of an element is determined by the valences and acceptability of all the elements to which it is connected.

To make emotional coherence more clearly applicable to psychological phenomena such as trust, I will now describe a computational model that specifies a mechanism for combining epistemic- and emotional-coherence calculations.

4 EMOTIONAL COHERENCE: MODEL

As chapter 2 showed, coherence can be computed by a variety of algorithms, but the most psychologically appealing model, and the model that first inspired the theory of coherence as constraint satisfaction, employs artificial neural networks. In this connectionist model, elements are represented by units, which are roughly analogous to neurons or neuronal groups. Positive constraints between elements are represented by symmetric excitatory links between units, and negative constraints between elements are represented by symmetric inhibitory links between units. The degree of acceptability of an element is represented by the activation of a unit, which is determined by the activation of the all the units linked to it, which takes

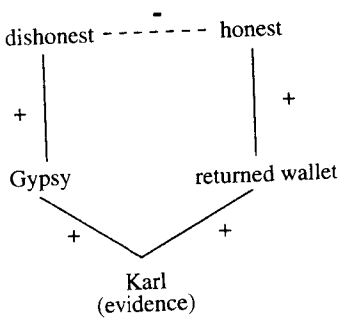

Figure 6.1
Simple connectionist network with excitatory (+) and inhibitory (–) links. All links are symmetric, with activation originating at the evidence node and flowing upward.

into account the strength of the various excitatory and inhibitory links.

For example, in figure 6.1 there are five units representing the Gypsy inference already described. The Karl unit is activated, and then activation spreads to what is known about Karl, i.e., that he is a Gypsy and returned the wallet. Activation then spreads to the units for dishonest and honest, which inhibit each other. Depending on the strengths of the links to these two concepts, one of them may become more active and suppress the other. Activation spreads around the system until all units reach stable activation levels, which typically takes 50–100 cycles. Activations can range between 1 (fully accepted) and –1 (fully rejected), and elements whose units have final activations above 0 are deemed accepted.

It is straightforward to expand this kind of model into one that incorporates emotional coherence. In the expanded model, called "HOTCO" for "hot coherence," units have valences as well as activations, and units can have input valences to represent their intrinsic valences.

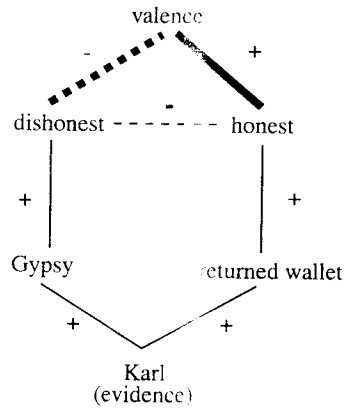

Figure 6.2
The network in figure 6.1 supplemented with valence inputs (thick lines).

Moreover, valences can spread through the system in a way very similar to the spread of activation, except that valence spread depends in part on activation spread. Figure 6.2 shows the network from figure 6.1 expanded to include a valence input to the concepts of *honest* and *dishonest*, the former positive and the latter negative. Just as activation spreads up the network from *Karl* to *honest* or *dishonest*, so valence spreads down the network to *Karl*. If *honest* becomes activated, then its positive valence will spread to *returned wallet* and then to *Karl*. The network of someone who is prejudiced against Gypsies would have a negative valence link directly to the *Gypsy* node.

The valence of a unit u_j is the sum of the results of multiplying, for all units u_i to which it is linked, the activation of u_i times the valence of u_i times the weight of the link between u_i and u_j. The actual equation used in HOTCO to update the valence v_j of unit j is the following:

$$v_j(t+1) = v_j(t)(1-d) + \text{net}_j(\max - v_j(t)) \text{ if net}_j > 0$$
$$= \text{net}_j(v_j(t) - \min) \text{ otherwise}$$

Here d is a decay parameter (say 0.05) that decrements each unit at every cycle, min is a minimum valence (−1), max is maximum valence (1). From the weights w_{ij} between each unit i and j, we can calculate net_j, the net valence input to a unit, thus:

$$\text{net}_j = \Sigma_i w_{ij} v_i(t) a_i(t)$$

Updating valences is just like updating activations (see McClelland and Rumelhart 1989) plus the inclusion of a multiplicative factor for valences. The equation for net valence input, combining both activation (like probability) and valence (like utility), is similar to ones proposed by Anderson (1974), Deutsch (1973), Fishbein and Ajzen (1975), and Lodge and Stroh (1993). The difference in the parallel model HOTCO is that the valence calculation is done locally and interactively, with an overall judgment emerging from the simultaneous application of the valence equation to numerous interconnected units.

To see how this works, we can step through how HOTCO processes the simple example in figure 6.2. Initially, the *Karl* evidence node has activation 1 and the valence input node has valence 1; the other four nodes have low default activation and valence values: 0.01. At the first round of network updating, activation flows from *Karl* to the *Gypsy* and *returned wallet* nodes, and valence flows from the valence node to *dishonest* and *honest* nodes, decreasing the valence of the former and increasing the valence of the latter. At the second round of updating, activation flows from *Gypsy* to *dishonest*, and from *returned wallet* to *honest*, but valence does not flow in the reverse direction because on the previous step *dishonest* and *honest* had not yet been activated. But at the third round

of updating these two units do have both valences and activations, so they can now spread positive valence to *Gypsy* and negative valence to *returned wallet*. Moreover, because of the inhibitory link between *dishonest* and *honest*, they tend to suppress each other's activations and valences. By the fourth round of updating, valence has begun to spread to the *Karl* node, representing an overall emotional attitude toward Karl. What this attitude will be in the end will depend on the strengths of the various links between the nodes in the network. For example, if there is a strong activation link between *returned wallet* and *honest* and there is a strong valence link between *valence* and *honest*, then *Karl* will end up with a positive valence. Typically it takes around 50 to 60 cycles of updating before the network has achieved stable activations and valences.

The term "valence" is borrowed from Gordon Bower's (1981, 1991) model of cognition and affect, but my model differs from his in that the kinds of inference that HOTCO employs are more complex than the simple associationistic ones that Bower discusses and that figures 6.1 and 6.2 display. A full account of emotional coherence has to include not only the sort of conceptual coherence involved in the Gypsy example, but also the contributions of explanatory, analogical, and deliberative coherence. The model is shown more fully in figure 6.3, which indicates more generally how evidence input can meld with emotion input to yield an emotional appraisal of the observed person or situation. HOTCO incorporates the previous computational models ECHO (explanatory coherence), DECO (deliberative coherence), IMP (conceptual coherence), and ACME (analogical mapping). All of these can contribute to the coherence inferences in the middle of figure 6.3 that determine how activation spreads from the evidence input nodes to various nodes representing hypotheses and other elements. Simultaneous with the spreading of

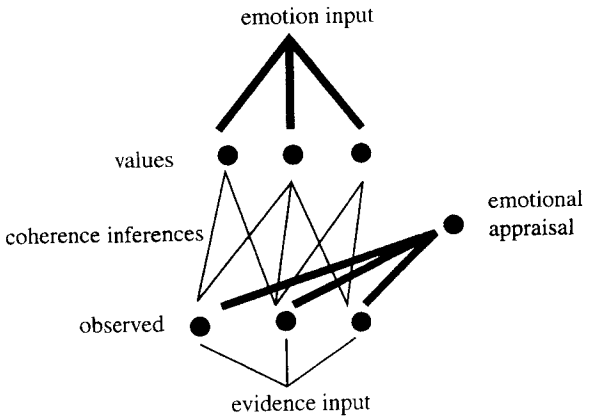

Figure 6.3
A general model of emotional coherence, not showing the many interconnected units that may be involved in coherence inferences. Thick lines are valence links. All links are symmetric, but activation flows up from the evidence input, and valences flow down from the emotion input.

activation determined by links established by explanatory, deliberative, conceptual, and analogical coherence, there is spreading of valences from the emotion input at the top of the diagram. As intermediate nodes acquire both activations and valences, valences spread down to the *observed* nodes that describe a person or situation, and then converge to produce an overall emotional appraisal of that person or situation. The next section shows how this works using a detailed example involving trust.

5 EMOTIONAL COHERENCE AND TRUST

One possible application of emotional coherence to trust would be that the extent to which people trust each other is determined directly by emotional coherence. That is, you

trust a person P to the extent that P has a positive valence. This conjecture is too simple, however, because trust is not always a universal attribute of a person, for it may be relative to a particular goal or situation: I may trust someone to wash my car but not to mind my children or invest my money. In these cases, the elements of emotional appraisal are very specific, not just *Karl* but *Karl as car washer*. Moreover, although positive valence may give a good overall indication of whether you *like* someone, likability and trustworthiness can be independent of each other. You can like affable and charming people without trusting them (if they are unreliable), and trust gruff or awkward people without liking them (if they are reliable with respect to the task to which trust is relevant). Hence there is more to trust than simply attaching a positive valence to a person, although such a valence may well be a large part of what produces the positive valence for the more specific node representing *trusting a person P to do X*.

A more concrete example will help to make these distinctions clear. In 1997 my wife and I needed to find someone to drive our six-year-old son, Adam, from morning kindergarten to afternoon day care. One solution recommended to us was to send him by taxi every day, but our mental associations for taxi drivers, largely shaped by some bizarre experiences in New York City, put a very negative emotional appraisal on this option. We did not feel that we could trust an unknown taxi driver, even though I have several times trusted perfectly nice Waterloo taxi drivers to drive *me* around town.

So I asked around my department to see if there were any graduate students who might be interested in a part-time job. The department secretary suggested a student, Christine, who was looking for work, and I arranged an interview with her. Very quickly, I felt that Christine was someone whom I could trust with Adam. She was intelligent, enthusiastic, interested in children, and motivated to be reliable, and she reminded me of a good baby-sitter, Jennifer, who had worked for us some years before. My wife also met her and had a similar reaction. Explanatory, conceptual, and analogical coherence all supported a positive emotional appraisal, as shown in figure 6.4.

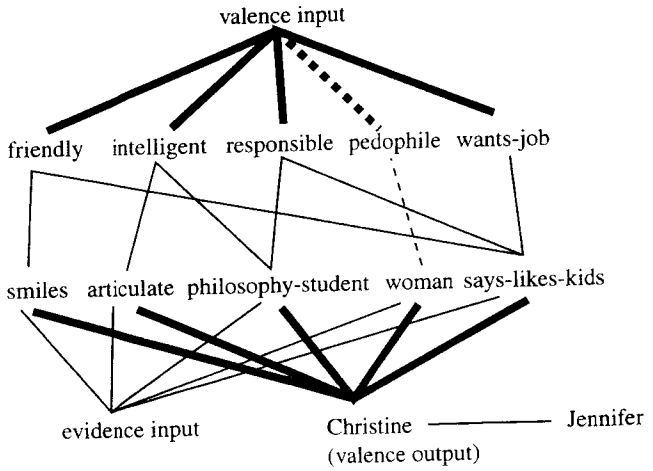

Figure 6.4
Emotional appraisal of a potential baby-sitter. Thin lines indicate activation links, while thick lines indicate valence links. All links are positive except for the two dashed lines, which are negative. Activation spreads only along activation links, but valences spread along both valence links and activation links.

Conceptual coherence encouraged such inferences as from *smiles* to *friendly*, from *articulate* to *intelligent*, and from *philosophy graduate student* to *responsible*. Explanatory coherence evaluated competing explanations of why she says she likes children, comparing the hypothesis that she is a friendly person who really does like kids with the hypothesis that she has sinister motives for wanting the job. Finally, analogical coherence enters

the picture because of her similarity with our former baby-sitter Jennifer with respect to enthusiasm and similar dimensions. A fuller version of figure 6.4 would show the features of Jennifer that were transferred analogically to Christine, along with the positive valence associated with Jennifer.

In the HOTCO simulation of this case, evidence input spreads activation to the units representing what is known about Christine (smiles, etc.), at the same time as valence input spreads valences to the units that have intrinsic valences (friendly, etc.). Then, as these units become active as the result of the coherence-based inferences, they spread valences down to the units that activated them. For example, just as *smiles* spreads activation to *friendly*, *friendly* spreads positive valence to *smiles*, which then spreads positive valence to *Christine*, whose valence is also being affected by other units, including ones representing Christine's analog, Jennifer. The result is an overall positive emotional appraisal of Christine as someone to be trusted. Thus coherence-based inferences, combined with emotional inputs, can yield a kind of emotional Gestalt impression of someone to be trusted. In figure 6.4, the valence output node for *Christine* represents not simply the fact that I like Christine, but also a positive emotional attitude attached to the decision to trust her with Adam. The emotional valence attached to Christine serves to integrate all the information relevant to assessing her as a baby-sitter, blending many coherence and valence considerations into a single emotional reaction accessible to consciousness. A judgment to trust people is more than just a judgment about the probability that they will do what is expected. For most people, trust and distrust are associated with positive and negative emotions, respectively.

Fenno (1978) describes how members of the U.S. House of Representatives present themselves to their constituents in order to gain their trust, attempting especially to convey three impressions: qualification, identification, and empathy. Representatives want their constituents to infer that they are qualified for the job, and accordingly provide brochures listing their background, experience, and accomplishments. With this information representatives provide evidence of competence, and they also try to convey to voters that they are sufficiently honest to qualify for the job. These inferences involve both conceptual and explanatory coherence. The voters can infer that the candidate is competent because that competence is associated with previous accomplishments, and they can infer that the candidate is honest because this attribute is associated with previous behaviors and provides the best explanation of some of those behaviors.

The second trust-generating impression, according to Fenno (1978, 59), is identification, where the message that voters get from the candidate is, "You can trust me because we are like one another." This is a kind of analogical inference, in which voters decide that candidates who are like them on various cultural dimensions are also like them in being trustworthy.

Third, every House member conveys a sense of empathy with his constituents, giving the impression of understanding and caring about their situations. This is a matter of explanatory coherence: the constituents infer that the best explanation of member's expressions of care and understanding is that he or she really does care or understand. Empathy can sometimes fail, as when Canadian Prime Minister Kim Campbell gave a speech at a shelter in Vancouver's Skid Row during the 1992

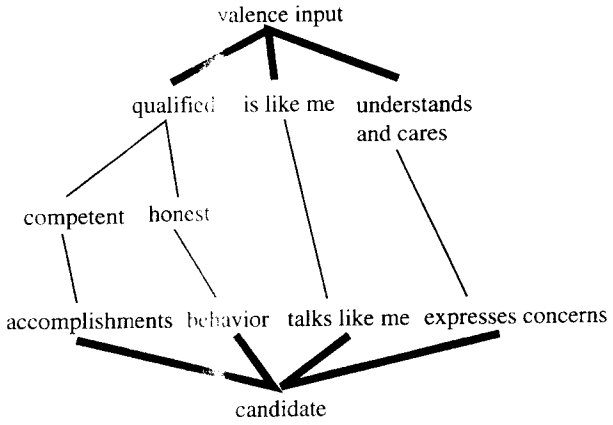

Figure 6.5
Emotional appraisal of a political candidate. Thin lines indicate activation links, while thick lines indicate valence links. Alternative interpretations are not shown in the figure, which does not distinguish the different activation links that derive from conceptual, explanatory, and analogical constraints.

Canadian election. She told the residents of the shelter that she too had known loss and disappointment, for she had once wanted to be a concert cellist. The weakness of the analogy between her history and the condition of the Vancouver derelicts undermined her attempt to convince them of her empathy.

Figure 6.5 illustrates how emotional coherence can generate an inference and feeling that a candidate is to be trusted. It does not show the details of the conceptual, explanatory, and analogical connections that can generate an emotional Gestalt towards a candidate, but serves to display how information converges to generate a positive or negative impression. Empathy is a particularly interesting kind of emotional inference, and I will now show how it can be understood in terms of emotional coherence.

6 EMPATHY

According to Barnes and Thagard (1997), empathy is a kind of emotional analogy, in which person *A* constructs an emotional image of person *B* by mapping *B*'s situation onto a similar situation previously experienced emotionally by *A*. In contrast to the current cognitive models of analogical mapping (e.g., Holyoak and Thagard 1989; Falkenhainer, Forbus, and Gentner 1989), what is mapped is not propositional information, but a feeling or an image of a feeling (Barnes 1998 discusses emotional images). When I have empathy for you, I do not just recognize abstractly that you are similar to me: I actually feel something like what you feel. This account of empathic analogy can be enhanced by viewing it in the context of the theory of emotional coherence.

Here is another example of empathy. When new graduate students arrive from overseas, they are often overwhelmed by arriving in a very different country and university and by having to work in English, if that is not their native language. My best shot at understanding their mental state comes from remembering my own disorientation when I went to study in Cambridge, England, in 1971. Everything seemed different and odd: the colleges, the town, the people, the food, the money, etc. Despite having only minor language difficulties, it was months before I felt I knew what I was doing. Because foreign students' situations are relevantly similar to mine many years ago, I can project my remembered emotional state of bewilderment and anxiety onto them, using my imagination to amplify its intensity because of the greater cultural and linguistic differences that they may face.

From the perspective of emotional coherence theory, empathy is more than just retrieving an emotion-laden

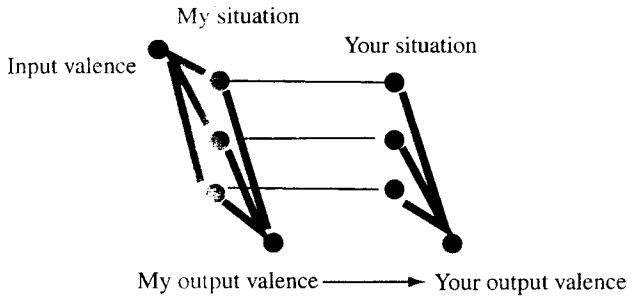

Figure 6.6
Empathy and emotional coherence. My situation serves as a source analog to generate an emotional valence that I transfer to you in your situation. Thick lines indicate valence links, thin lines indicate analogical links, and the arrowhead indicates transfer of valence.

source to map onto a given target. Empathy-producing source analogs can be generated, not just remembered, as when I generate an England-only-worse analog for my foreign student or when I generate a loss-of-English-Canadian-language-and-culture analog to help understand Quebec nationalism (see next section). In both those constructed analogs, I am generating a situation that produces an emotional image based on the emotional coherence of the situation, which has different aspects whose valences contribute to its overall valence. Once I establish a correspondence between my imagined situation and yours, I can ascribe to you an emotional valence for your situation that is similar to the emotional valence that I ascribe to my situation.

Figure 6.6 portrays schematically how I do an analogical mapping that enables me to transfer the valence of my situation to your situation. The elements in my situation have input valences that produce an overall output valence for the whole situation. Once the elements in my

situation have been mapped to the elements in your situation, then the valences of these elements can spread over to the elements in your situation and then produce an output valence for your situation similar to the output valence for my situation. In this way I can feel some approximation to how you feel.

Figure 6.6 is oversimplified in that it portrays empathy as merely a matter of analogical mapping. In fact, a full range of coherence inferences may be involved in (1) understanding your situation, e.g., making inferences about your beliefs and goals, (2) figuring out what elements to add into my constructed analog of your situation, and (3) computing the valence of my constructed situation that serves as a source analog situation. Moreover, empathy is not just a matter of positive and negative valence, but also requires transfer of the full range of emotional responses. Depending on his or her situation, I need to imagine someone's being angry, fearful, disdainful, ecstatic, enraptured, and so on. As currently implemented, HOTCO transfers only positive or negative valences associated with a proposition or object, but it can easily be expanded so that transfer involves an *emotional vector* that represents a pattern of activation of numerous units, each of whose activation represents different components of emotion. This expanded representation would also make possible the transfer of "mixed" emotions.

Empathy is relevant to trust in different ways. The last section described how politicians use empathy to generate trust: people are inclined to trust people who have empathy for them. U.S. President Bill Clinton is noted for his empathic ability, or at least for his ability to appear empathic. If you have empathy for me, leading you to understand me positively by transferring your emotional states, then I will be more likely to trust you. Empathy is often cited as a crucial ingredient in psychotherapy, and

one of its contributions is to enhance trust as well as understanding. Of course, empathy does not always lead to positive valuations: mapping what I know about you onto what I know about me in a similar situation may lead me to project a negative valence onto you if I realize that I would likely act badly in the situation in which you are in.

Trusting people often involves inferring their motives and intentions, but it can also involve inferring their emotional states. I am unlikely to trust someone who I have reason to believe is seething with concealed anger against me. On the other hand, I am likely to trust and like someone who trusts and likes me. Thus if empathy (analogical mapping of the emotion that I would experience if I were in a similar situation) suggests that if someone attaches a positive valence to me, then I can attach a positive valence to him or her. Alternatively, putting yourself in someone else's shoes may strongly suggest that the person's valences are negative, and thus reduce your inclination to like and trust him or her.

To sum up, emotional coherence suggests the following recipe for how to achieve empathic understanding of a person *P*:

1. Take what you know of *P*'s personality and situation, and use explanatory and other kinds of coherence to make inferences about it that supplement the given information.

2. Use analogical coherence to retrieve a similar situation from your own experience. (See Thagard, Holyoak, Nelson, and Gochfeld 1990 for a model of analog retrieval.)

3. Use imagination to enhance the retrieved situation so as to bring it closer to *P*'s.

4. Use coherence-based inferences and emotional coherence to generate a valence for your constructed situation.

5. Project this valence onto *P* as representing *P*'s likely emotional state in that situation.

I now turn to another political application of empathy, arguing that Canada's dealings with Quebec nationalists requires empathic understanding of their goals and of the emotional coherence of their belief system.

7 NATIONALISM

In 1996 the province of Quebec voted in a referendum concerning whether Quebec should separate from Canada and become a sovereign nation. The referendum was defeated, but only by less than 1 percent of the vote, and a substantial majority of those whose first language is French voted in favor of separation. To much of the world, which views Canada as one of the world's best countries to live in, Quebec separatism seems very puzzling; indeed, it seems bizarre to most Canadians outside Quebec. A typical reaction is, "What do these people want? Leaving Canada doesn't make any sense. They're just being emotional."

Nationalism is clearly an emotional issue: many people feel very strongly about the nations and ethnic groups to which they belong, and they often have strong negative emotions towards other nations and ethnic groups (Caputi 1996, Group for the Advancement of Psychiatry 1987, Ignatieff 1991, Kecmanovic 1996, Stern 1995). According to the theory of emotional coherence, however, emotions are not inherently irrational, since they may be tied to coherence judgments that are rooted in evidence, for example via explanatory coherence (see the section on normative issues at the end of this chapter). Without trying to assess whether nationalism is rational or not, I want to try to understand it as a phenomenon involving emotional

coherence. In particular, we can get a better understanding of Quebec separatism by constructing a profile that integrates considerations of emotions and coherence.

A good place to start is the writings of René Lévesque, the first leader of the Parti Québécois, which was formed in 1967 with a platform of achieving Quebec independence. The reasons for establishing the new party were eloquently stated in the book *Option Québec*:

> We are Québécois.
>
> What that means first and foremost—and if need be, all that it means—is that we are attached to this one corner of the earth where we can be completely ourselves: this Quebec, the only place where we have the unmistakable feeling that "here we can be really at home."
>
> Being ourselves is essentially a matter of keeping and developing a personality that has survived for three and a half centuries.
>
> At the core of this personality is the fact that we speak French. Everything else depends on this one essential element and follows from it or leads us infallibly back to it. [Historical background given.]
>
> All these things lie at the core of this personality of ours. Anyone who does not feel it, at least occasionally, is not—is no longer—one of us.
>
> But *we* know and feel that these are the things that make us what we are. They enable us to recognize each other wherever we may be. . . .
>
> This is how we differ from other men and especially from other North Americans. (Lévesque 1968, 1–15)

Lévesque describes how French in Quebec is threatened by the dramatically dropping birth rate among francophones and the strong preference shown by immigrants to the province to learn and work in English rather than French.

The appeal of Quebec separatism can be understood in terms of strong emotional inputs and outputs that are part of deliberative coherence. As the above quotation suggests, Québécois have intense desires to feel at home in their own province, to speak French, and to avoid assimilation into the dominant English-speaking environment of the rest of Canada and North America. Lévesque and his colleagues who started the Parti Québécois strongly believed that sovereignty was the only means to avoid assimilation. Figure 6.7 provides a rough sketch of this attitude. A computational model of this view provides a strong positive valence to feeling at home and speaking French, and a strong negative valence to assimilation. These valences then spread to the options of separation versus staying in Canada, with the former receiving a strong positive valence and the latter reaching an emotional state akin to repugnance.

Of course, the issue is a lot more complicated than figure 6.7 indicates. The valence links presume a number of empirical projections that depend on empirical evidence: critics of Quebec sovereignty deny that staying in Canada will lead to the elimination of French, pointing to such phe-

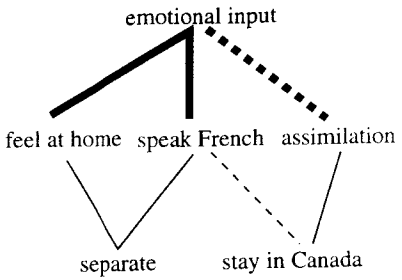

Figure 6.7
A sketch of the emotional coherence of Quebec separatism. Thick lines indicate valence links. Thin lines indicate facilitation relations that are part of deliberative coherence. Dashed lines indicate negative links.

nomena as the bilingualism of the federal government and the growth of French immersion programs in English Canada. In response, sovereigntists point out that the trend outside Quebec among native French speakers is, in fact, strongly towards assimilation. Probably the most effective argument used by antiseparatist forces has been economic: Quebec can prosper within Canada, but risks economic disaster by going on its own. Lévesque and other sovereigntists pointed to a number of models for an economically viable Quebec, particularly the European Economic Community, in which numerous countries are economically drawing closer and closer together for mutual benefit. Analogously, Quebec could be part of a North American common market with the United States and what's left of Canada.

My goal in this section is not to assess the costs and benefits of Quebec separatism, but rather to understand its emotional appeal. For sovereigntists, the economic problem of separation is only a short-term one that can be taken care of by negotiating a new economic arrangement. Hence separatism has only a small negative impact on the goal of economic well-being. In the battle of emotional analogies, separatists look to the European Community, the Scandinavian union (following the peaceful separation of Norway from Sweden in 1905), and the recent peaceful separation of Slovakia from the Czech Republic. They resist analogies with much uglier situations, such as the American Civil war, Northern Ireland, and Bosnia. Blanchette and Dunbar (1997) collected 234 analogies used in the 1996 Quebec referendum. Separatists have a strong emotional attachment to feeling at home and preserving their national personality, both of which are tied in with preserving French as the dominant language of Quebec. They also have a belief that negotiations within Canadian federalism have failed, as in such incidents as the

forced repatriation of the constitution from Great Britain to Canada in 1982 and the failure of the Meech Lake accord in 1990. The result is that separatists arrive at an intense emotional conviction in favor of forming their own country.

Daniel Latouche (1990, 89) wrote, "When will you English Canadians get it through your thick collective skull that we want to live in a French society, inside and outside, at work and at play, in church and in school. Is this so difficult to understand?" English Canadians generally fail to understand why this goal is so strong for Québécois, and why many francophones believe that the goal cannot be achieved within Canada. English Canadians do not feel the anger that arises from the perception of a long history of humiliation, stretching from the conquest of Quebec by England's troops in 1760 through the more recent constitutional wrangles and referendum defeats. Because English Canadians do not have the kind of passionate nationalism found in Quebec or even in the United States, it is difficult for them to have an empathic understanding, based on mapping their own emotions, of why Québécois feel so strongly about their cultural identity. Moreover, although some Canadian cultural institutions are undoubtedly threatened by the American entertainment juggernaut, at least there is no fear that English will be wiped out. In contrast, the Québécois can infer from past behavior and current utterances of many English Canadians that the English do not care about preserving Quebec culture. This inference is based on the high explanatory coherence of the hypotheses that the English have little comprehension and appreciation of the French demands.

Nationalism has often been an evil force in human history, as witnessed in atrocities of the Nazis and in recent Balkan conflicts. But it can have a positive side when it is

directed, as a kind of self-preservation, toward maintaining cultural practices that are important to the people who perceive themselves as a nation. Personally, I have virtually no ethnic identification and, like most anglophone Canadians, only a weak emotional attachment to my native country. Understanding political movements like Quebec separatism requires me to imagine how badly I would feel if I had the prospect, for myself or my children, of being unable live and work in my native language. This empathic understanding involves a kind of analogical coherence in which I transfer my emotional attitude to another in a similar situation. It is probably easier for an Israeli, a German, an Italian, or even an American to understand Quebec nationalism than for an English Canadian, but empathy can be generated if one works at it. It is an interesting question why some French Canadian leaders, such as Pierre Trudeau and the current Prime Minister Jean Chrétien, have little appreciation of the separatist position. Perhaps they do have an empathic understanding that is overruled by other considerations, such as valuing universal rights and freedoms more highly than nationalist aspirations.

Let me emphasize that I am not trying to give a kind of romantic glorification of the sort of aggressive nationalism that urges a people to see themselves as superior to all foreigners and that can be used to justify conquest. The fact is, however, that nationalism is clearly in part a matter of emotion, and it can also be a matter of deliberative, explanatory, and analogical coherence. Defensive nationalism, based on the goal of preserving a culture, is not obviously either irrational or immoral. Convincing Quebec to stay *happily* in Canada will require much more than dire threats about the negative economic and political consequences of separation. Such threats leave untouched the strong feelings about home, personality, and language that

drive separatism. Rather, Canadian unity will require convincing francophones in Quebec that their language and culture are safe within Canada. If this task is accomplished, emotional coherence may point toward accommodating Quebec nationalism without destroying Canada.

8 METACOHERENCE

The applications of the HOTCO model so far described have attached value to particular objects or situations. But emotions also involve more general kinds of evaluations. When a situation "makes sense" to us, we feel a general well-being, whereas a situation that we are unable to comprehend can cause anxiety. The usually pleasant feeling that something makes sense involves an overall assessment of coherence, in contrast to the confusion and anxiety that often accompany incoherence. I call these *metacoherence* emotions, because they require an overall assessment of how much coherence is being achieved.

On the theory of coherence sketched earlier, the coherence of a partition of elements into accepted and rejected is determined by the extent to which positive and negative constraints are satisfied. If the elements are related by highly incompatible constraints, it is possible that the best partition will not be very good, so that the overall coherence of the system is low even though the partition maximized it. Scientists faced with highly conflicting evidence supporting different theories may choose the theory that is best, given the overall evidence, but remain uncomfortable with their conclusion because of low overall coherence. For example, Newtonian mechanics dominated physics throughout the nineteenth century, but some scientists found it to be imperfectly coherent because it gave incorrect predictions about the orbit of Mercury. Similarly, in

everyday life we sometimes make optimal decisions that we are not generally happy with, as when we are forced to make the best of a bad situation. A student, for example, may decide to go to a community college rather than a university because of financial constraints, but be unhappy about not having the chance to pursue more advanced studies. I interpret this as a case where the valence attached to an action is positive, but the emotional reaction to the overall judgment is negative because the best action leaves important goals unsatisfied.

Another metacoherence emotion is surprise, which reflects a judgment that a situation has occurred differently from what was expected. Such failed expectations are noticed when the most coherent interpretation of a situation is replaced by another coherent interpretation that differs from it substantially. For example, if I am watching a hockey game in which one team is leading 5 to 0 at the end of the first period, I will be surprised to find that the game turned out to be a victory for the team that was behind. Surprise is a function of the extent to which elements switch status from accepted to rejected or vice versa, with the greatest surprise contributed by elements that go from being strongly accepted or strongly rejected to the opposite.

A theory of emotional coherence should therefore incorporate overall judgments of coherence and incoherence, happiness and sadness, surprise, and other general emotions. It is easy to expand the HOTCO program by writing functions that calculate the overall coherence and valence satisfaction of the system (chap. 2, section 4), but such global calculations are at odds with the model's connectionist assumptions. Rather, judgments of coherence, happiness, and surprise should emerge from local assessments made by particular units. Figure 6.8 provides a rough picture of how this should work. The various

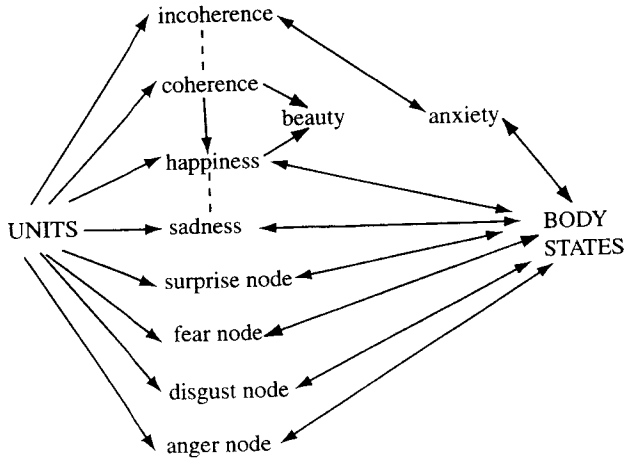

Figure 6.8
Metacoherence nodes (lowercase) in relation to cognitive units and body states. Dashed lines are negative constraints.

cognitive units that represent elements involved in explanatory, conceptual, and other kinds of coherence collectively activate nodes representing coherence, incoherence, happiness, and so on, which also have connections with other emotion nodes and bodily states. It is natural to think of the cognitive units, emotion nodes, and body states as together constituting a dynamic system with a very large state space representing all the different combinations of activations, valences, and values of other variables. Particular emotions, of which there are hundreds if one can judge from the number of emotion words in English and other languages, correspond to regions in this state space.

In the computational model HOTCO, the coherence and incoherence nodes receive activation from each of the cognitive units according to the local coherence of each active unit. An individual unit can assess its own coherence status by determining the extent to which its own

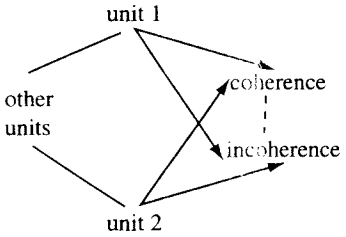

Figure 6.9
Two units affecting the coherence and incoherence nodes, which inhibit each other. The dashed line is a negative constraint.

constraints are satisfied, taking account of its positive and negative links to other units. If a unit is active and it has a positive link to another unit, then the constraint is satisfied only if the other unit is also active. Alternatively, if an active unit has a negative link to another active unit, then that unit must not be active if the constraint is to be satisfied. In HOTCO, each unit has a unidirectional link to the coherence node, to which it passes on its degree of constraint satisfaction, and a unidirectional link to the incoherence node, to which it passes on its degree of constraint *non*satisfaction. Hence the activation of the coherence and incoherence nodes depends on the coherence of the individual cognitive units. Figure 6.9 gives a more detailed picture of the linkages. The coherence and incoherence nodes mutually inhibit each other, so that one or the other will tend to become active, representing an overall judgment of how much the whole situation makes sense. As figure 6.8 suggested, general coherence influences emotions such as happiness, while incoherence influences emotions such as anxiety. There may be individual differences in the strength of the links between the nodes: in people with a high tolerance for incoherence, the link between the incoherence and anxiety nodes will be particularly weak, and in people with a great appreciation for coherence, the link between the coherence and happiness nodes will be particularly strong.

In social psychology, the cognitive dissonance theory of Festinger (1957) has been used to account for a wide variety of phenomena. Shultz and Lepper (1996) presented a computational model of cognitive dissonance in terms of parallel constraint satisfaction, using ideas very similar to those found in coherence models such as ECHO. However, dissonance is not simply cold incoherence, but also has an affective dimension involving negative emotions such as anxiety and discomfort (Cooper and Fazio 1984). Such emotional reactions are more fully modeled using HOTCO's metacoherence nodes than by traditional connectionist systems based solely on parallel constraint satisfaction.

HOTCO also uses a local mechanism to activate the general happiness and sadness nodes, which are affected by both the activation and the valence of each node. If an active unit has positive valence, it affects the activation of the happiness node to an extent that is a function of the unit's activation as well as its valence. On the other hand, if an active unit has negative valence, it affects the activation of the sadness node to an extent that is a function of the unit's activation and magnitude of negative valence. The happiness and sadness nodes inhibit each other, so the system will tend to settle into a state in which happiness is dominant, sadness is dominant, or both are neutral. Figure 6.10 shows the structure, similar to that in figure 6.9. But whereas in figure 6.9 the activation of the coherence nodes depends on the units' calculation of their degree of constraint satisfaction, in figure 6.10 the activation of the happiness and sadness nodes depends on the units' activations and valences.

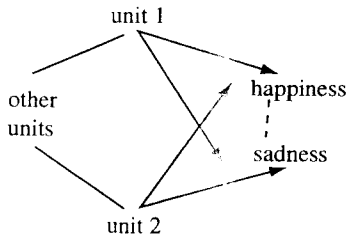

Figure 6.10
Two units affecting the happiness and sadness nodes.

Finally, let us consider how HOTCO produces judgments of surprise. After the network of cognitive units settles, each unit records its final activation. When new information is added to the network and it settles again, each unit compares its new activation with its previous activation, and the difference represents the extent to which the new information has produced a surprising result for that unit (I owe this way of implementing surprise to Cameron Shelley). Each unit conveys to the surprise node the extent to which it is locally surprised, so that many nodes affect the general surprise node shown in figure 6.8, which interacts with body states such as accelerated heart rate. There are both pleasant and unpleasant surprises, so the overall emotional state of the system depends on the activation of other nodes, such as the ones for happiness and sadness.

These extensions to HOTCO show how emotion nodes that represent metacoherence judgments can be implemented in ways that allow local calculations at the level of individual cognitive units to produce general emotional reactions. These metacoherence-based reactions make possible an understanding of the very complex emotional states involved in beauty and humor.

9 BEAUTY AND SYMMETRY

From symphonies to sunsets, beautiful objects produce pleasure and happiness, so beauty obviously has a large emotional component. But it also has a large coherence component, as many philosophers of art have noticed. R. G. Collingwood confidently asserted, "Beauty is the unity or coherence of the imaginary object; ugliness its lack of unity, its incoherence. This is no new doctrine; it is generally recognized that beauty is harmony, unity in diversity, symmetry, congruity, or the like" (1925, 21). The doctrine that beauty is unity in diversity originated with the eighteenth-century thinker Frances Hutcheson, who said, "The figures which excite in us the ideas of beauty seem to be those in where there is *uniformity amidst variety*. . . . The variety increases the beauty in equal uniformity. . . . The greater uniformity increases the beauty amidst equal variety" (Hutcheson 1973, 40–41). The eminent mathematician G. H. Hardy also saw beauty as connected with coherence: "A mathematician, like a painter or poet, is a maker of patterns. . . . The mathematician's patterns, like the painter's or the poet's, must be *beautiful*; the ideas, like the colours of the words, must fit together in a harmonious way" (1967, 84–85). In a beautiful object, diverse elements come together coherently to produce positive emotions, whereas in an ugly object the elements do not fit together and tend to produce negative emotions.

The metacoherence architecture depicted in figure 6.8 above provides a model of how the human mind might generate beautiful experiences. The cognitive units represent different aspects of an object, for example, the features of a human face. Particular features may have input valences attached to them, for example, eyes that are large and colorful, but the beauty of a face depends not just on

the individual features but on how well these fit with each other. Faces have numerous built-in constraints, for example, that the eyes should be the same size and above the nose. If the constraints are well satisfied, then the face generates a high degree of perceptual coherence, which in turn generates positive emotions. A misshapen face, on the other hand, violates conventional constraints on facial structure, producing perceptual incoherence and a negative emotional reaction.

Beauty and ugliness can be intellectual as well as perceptual, as in Hardy's remark about mathematics and in T. H. Huxley's famous complaint about a beautiful theory being killed by an ugly fact. In the poem "Ode on a Grecian Urn," John Keats even goes so far as to identify beauty and truth:

> When old age shall this generation waste,
> Thou shalt remain, in midst of other woe
> Than ours, a friend to man, to whom thou say'st,
> "Beauty is truth, truth beauty,"—that is all
> Ye know on earth, and all ye need to know.

Without going that far, we can still recognize that scientists, like mathematicians, often use beauty as a guide to truth. According to Zemach, "An account that is rich, powerful, dramatic, elegant, coherent, and simple—that is, beautiful (*unity in variety* is the oldest definition of Beauty)—is probably true" (1997, 64).

McAllister (1996, 40) identifies four classes of aesthetic properties of scientific theories: form of symmetry, invocation of a model, visualizability/abstractness, and metaphysical allegiance. The last three of these are easily interpreted as matters of coherence. On McAllister's account, invocation of a model is a matter of analogy between a source domain, such as the solar system, which provides a model of a target domain, such as the atom. Such modeling is a matter of analogical coherence, and an

apt analogy that satisfies the constraints of similarity, structure, and purpose as discussed by Holyoak and Thagard (1995) often inspires positive emotions. Visualizability requires construction of a mental image that guides our understanding of a phenomenon, and thus would seem to combine perceptual and explanatory coherence, as when classical electromagnetic theory portrays the interaction of two electrons as the gradual intensifying of a repulsive electrostatic force. By the metaphysical allegiance of a theory McAllister means its fit with claims about the ultimate constituents of the world and with norms of reasoning about them. For example, a modern physicist might react with anger, disgust, or laughter to anyone who proposed a theory that the planets are carried around the sun by demons.

The remaining aesthetic property of scientific theories is symmetry, which is also related to emotional coherence. Rosen writes, "What makes a theory beautiful? This is, of course, a subjective matter, and in science too, beauty is in the eye of the beholder. But an opinion poll would reveal that simplicity and symmetry play decisive roles in determining whether a theory appears beautiful or not to most scientists" (1975, 121). Simplicity is part of explanatory coherence, which favors hypotheses that accomplish their explanations using fewer auxiliary hypotheses (Thagard 1992b), so its contribution to beauty can be handled in terms of coherence. But what can we make of symmetry?

Some kinds of symmetry can be understood in terms of analogical and perceptual coherence. For example, the bilateral symmetry of human faces consists of an isomorphic mapping between the two sides: the left side of the face is usually analogous to the right side. Symmetry as a kind of analogy is also apparent in McAllister's description of Einstein: "The symmetry

that Einstein valued, and which he judged classical physical theory to possess to an insufficient degree, is one in virtue of which a theory offers explanations of the same form for events deemed physically equivalent" (1996, 43). The principle underlying this value is something like the idea that analogous phenomena should have similar explanations. Such symmetry is easily accommodated within analogical and explanatory coherence. When a theory gives analogous explanations to similar phenomena, it achieves two kinds of coherence simultaneously and can therefore be perceived as beautiful.

But symmetry is broader than bilateral perceptual symmetry or internal explanatory analogy. In general, a structure is said to be symmetric under a transformation if and only if the transformation leaves the structure unchanged. Many kinds of transformations establish symmetries, including spatial ones like flipping and rotation, but also conceptual ones like substitution of terms in parallel verbal constructions.

Can symmetry in general be brought within the scope of coherence theory so that its contribution to beauty can be explained in terms of emotional coherence? Following Rosen (1995), we can quantify the degree of symmetry of an object or system as the number of transformations that operate on it and preserve structure. A square, for example, is more symmetric than a triangle, because there more ways of transforming it that preserve its basic structure. Each transformation can be thought of as a kind of analogical mapping of the system to itself, with the transformed system required to be at least approximately equivalent to itself. But symmetry is not a matter of just one transformation, so it cannot be understood in terms of a single internal analogy. Rather,

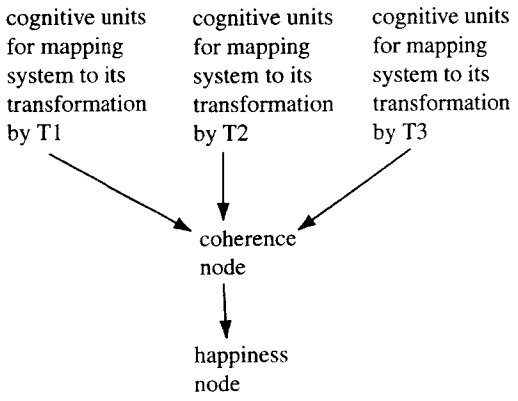

Figure 6.11
Symmetry as coherence of multiple self-analogies.

symmetry is a matter of a system having *multiple* internal analogies: many transformations of a system are analogous to it. The more such internal analogies, the greater the degree of symmetry. Symmetry, then, turns out to be a kind of metacoherence, in that it involves a summary of various judgments of analogical coherence. A fractal picture, for example, is highly symmetric, in that there are many ways of transforming it that do not change its appearance. Figure 6.11 schematizes the general relation of symmetry to analogical coherence. The more transformations that generate coherent analogies, the more activation is passed to the coherence node and the more positive is the emotional response. On the other hand, if transformations fail to produce good analogies, many constraints will be unsatisfied and the incoherence node will be activated, which produces a negative emotional reaction. Thus symmetry, like uniformity and simplicity, is an aspect of beauty that can be understood in terms of emotional coherence.

10 HUMOR

The last section gave an emotional-coherence account of the cognitive processes involved in finding something to be beautiful, and related processes can be involved in finding something to be funny. The relevance of emotional-coherence theory to humor is summarized in the following theses:

- Humor involves a shift from one coherent interpretation of an utterance or situation to a different coherent interpretation.
- This coherence shift generates the emotional state of surprise, using metacoherence mechanisms that attend to shifts in activation levels of units in a neural network representing components of the interpretations.
- Aspects of the utterance or situation generate other emotions, such as happiness, that interact with surprise to produce the overall emotional state of mirth.

Consider, for example, the following definition: "A drug is a substance that, when injected into a laboratory rat, produces a scientific paper." Until the reader or listener gets to the last two words, this sentence generates a coherent interpretation involving the expectation of a biochemical account of what a drug is. But the last two words shift to another, unexpected and surprising interpretation that defines "drug" in terms of the activities and motivations of scientists; we have production by a scientist rather than production in a rat. But surprise is not the only emotion involved: it would be even more surprising but not particularly funny if the sentence ended "produces a shower curtain." For the joke to be funny, the new interpretation must be coherent on its own terms and must generate other emotions, in this case glee at the thought that the whole purpose of drugs is to generate scientific research, which makes fun of scientific researchers. Puns are another form of humor that combine coherence and incoherence. If someone remarks "That's very punny," there is both a fit and an incompatibility between the usual interpretations of "pun" and "funny."

Emotional coherence is also evident in the following example of a humorous analogy:

> The juvenile sea squirt wanders through the sea searching for a suitable rock or hunk of coral to cling to and make its home for life. For this task, it has a rudimentary nervous system. When it finds its spot and takes root, it doesn't need its brain anymore, so it eats it! (It's rather like getting tenure.) (Dennett 1991, 177)

This story initially generates a coherent biological interpretation, with some surprise and amusement generated by learning the unusual fact that there is an organism that eats its own brain, which is incoherent with what we know about animals' eating behavior. But the real surprise comes with the parenthetical comparison with getting tenure. If the comparison were surprising but unconnected, it would not be funny: there is no point in saying "It's rather like getting a sun tan." Rather, humor arises because of coherence and emotion. First, there is a coherent analogical mapping that we can generate between the sea squirts eating their brains and tenured professors ceasing to use theirs (Shelley, Donaldson, and Parsons 1996). Second, in addition to surprise, this analogy generates emotions such as glee directed at brainless professors. Humor is thus emotional coherence, with surprise and other emotions, arising from two coherent interpretations. (Many other examples of humorous analogies are discussed in Thagard and Shelley, forthcoming.)

This account of humor subsumes other theories of humor that have been historically influential (for reviews,

see Keith-Spiegel 1972 and Lefcourt and Martin 1986). According to the incongruity theory, humor arises because an utterance or situation brings together disparate ideas in a surprising manner. On my account, incongruity is the incoherence between the initial coherent expectations in an utterance or situation and the final coherent interpretation after something surprising occurs. Another theory of humor is the superiority account, according to which humor functions to disparage someone or something. Not all humor aims at superiority, but my drug and sea squirt examples show how superiority-related emotions such as glee and gloating can be part of a humorous reaction based on emotional coherence. Such emotions tied to feelings of superiority increase the emotional intensity of the coherence-surprise reaction, and humor is the interactive sum of the cognitive/emotional response.

Finally, emotional coherence theory can accommodate the emotional-release theory of humor. Humor sometimes arises in tense and anxious situations and provides a welcome release. For example, in a difficult social situation, breaking the ice with an amusing comment can shift from one interpretation of the situation attended with negative emotions such as fear of failure and to another coherent interpretation with more positive emotional content. Nervous novice public speakers are sometimes advised to imagine they are talking to a naked audience—a surprising shift that reduces the anxiety of the situation. Many jokes start with taboo subjects such as sex and other bodily functions, then shift them to a less threatening interpretation. (Did you hear about the man with five penises? His pants fit like a glove.) Emotional release comes through a shift that is emotional as well as cognitive, producing a particularly emotionally intense kind of surprise, since it involves shifts in activation not only of cognitive nodes but also in nodes that carry the overall

emotional interpretation of the situation. Hence from the perspective of the theory of emotional coherence we can see why emotional release is an important part of humor.

My account of humor is somewhat similar to the catastrophe theory of jokes proposed by Paulos (1980), but it is less metaphorical. Humor involves a sudden shift from one state of the cognitive/emotional network to another and is therefore like a catastrophe in the mathematical sense. It is more concrete, however, to think of humor in terms of the mathematics of dynamic systems such as HOTCO networks. The initial coherent interpretation establishes a particular state of the dynamic system defined by the activation and valence values of the various cognitive and emotional nodes of the integrated cognitive/emotional network. But the punch line of the joke or the humorous event of the situation shifts the system into another stable state distant from the original one. Humor, like other emotional changes, involves a shift from one region in the state space of the system to another. Implicit in this account is the conception of an emotional state as a region in the state space of a dynamic system constituted by the activation and valence values of the nodes. The relevant dynamic system should be construed even more broadly to include a wide range of physiological states of the organism in which the neural network resides. Thus an emotion is a region of state space of a system that includes not only the cognitive/emotional neural network, but also the somatic states that influence and are influenced by the neural network. Emotional changes are then shifts from one region of the state space to another region with different cognitive, metacoherence, and somatic states. Cognitive therapy, which can be used for producing positive emotional shifts, can similarly be understood in terms of emotional coherence.

11 COGNITIVE THERAPY

Cognitive therapy is an effective method for trea... variety of emotional disorders, including depres... Unlike psychoanalysis, it does not require detailed deliv... into a patient's past, but instead concentrates on helpi... the patient to replace unrealistic beliefs and goals wit... more reasonable ones (Ellis 1962, 1971; Beck 1976). Beck writes,

> In order to understand the cognitive approach to the treatment of depression, it is necessary to formulate the problems of the depressed patient in cognitive terms. These characteristics of depression can be views as expressions of an underlying shift in the depressed patient's cognitive organization. Because of the dominance of certain cognitive schemas, he tends to regard himself, his experiences, and his future in a negative way. These negative concepts are apparent in the way the patient systematically misconstrues his experiences and in the content of his ruminations. Specifically, he regards himself as a "loser." . . . The cognitive approach for counteracting depression consists of using techniques that enable the patient to see himself as a "winner" rather than a "loser," as masterful rather than helpless. (1976, 264 ff.)

The cognitive therapist works with patients to revise beliefs and goals in ways that produce more positive appraisals of themselves and their situations.

Cognitive therapy is not merely a matter of pointing out to patients the unreasonableness of some of their beliefs and goals. Beck describes a depressed woman who was convinced that she had been a failure as a mother and concluded that she should kill herself and her children. He says, "This kind of depressive thinking may strike us as highly irrational, but it makes sense within the patient's conceptual framework" (Beck 1976, 16). Her negative views of herself suggested to her that she should commit

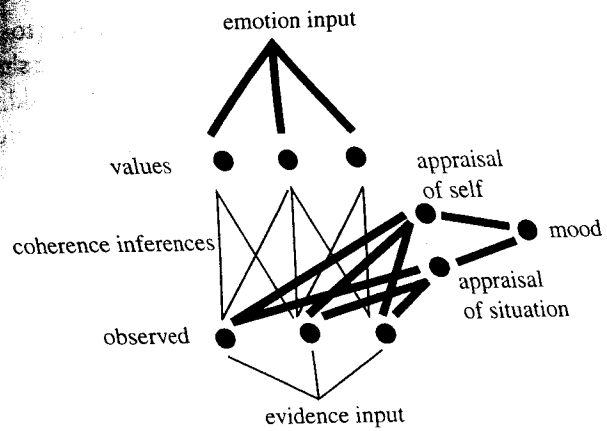

Figure 6.12
Mood changes affected by emotional coherence, expanded from figure 6.3.

suicide, and she felt she had to kill her children too to prevent them from experiencing comparable misery.

The theory of emotional coherence explains both why cognitive therapy can be difficult and why it can be successful. Figure 6.12 is an expanded version of figure 6.3, which showed how numerous coherence inferences can generate an emotional appraisal. For depressives, a coherent set of inferences imply a negative evaluation of the self as well as of his or her situation. Activations spread up from evidence input and valences spread down from emotion input, interacting to produce negative appraisals of self and situation. These negative appraisals produce a negative mood node, which then tends to keep the appraisal nodes negative. Cognitive therapy requires introducing new evidence and reforming coherence relations in ways that produce a change in the emotional appraisal of the self and the situation. For example, the therapist could help the depressed woman considering suicide to recall

times when she had been a good mother to her ch
and thereby help her to revise her belief that she is a fa
Revision of this belief along with others could then ch
her overall emotional appraisal of herself and her si
tion, which could lead to a dramatic improvement in
mood. (Moods are ongoing affective states that are bo
modes of appraisal and states of action readiness; se
Frijda 1993.) Mood changes are a kind of emotional
Gestalt shift produced by a change in emotional coherence,
driven in part by a shift in the inferences made in the belief
network but crucially accompanied by shifts in valences
attached to various nodes that affect the overall valence of
the nodes representing the patient and her situation. Cog-
nitive therapy assumes that inferential changes can affect
emotional reactions, but it lacks a theory of how inference
works and how it interconnects with emotional changes.
These gaps are filled by my account of coherence-based
inference as constraint satisfaction linked with valence
adjustments.

In contrast to cognitive therapy, psychodynamic therapy based on Freudian ideas places more emphasis on unconscious motivational processes. Westen (2000) argues that optimal treatment of patients may require integration of cognitive and psychodynamic methods, along with delving into problematic and conflicting motives and irrational beliefs. From the perspective of HOTCO, cognitive therapy is aimed at altering the cognitive constraints based on explanatory and other kinds of coherence, whereas psychodynamic therapy pays more attention to the fundamental emotional constraints implemented as valence links. Whereas cognitive therapy often achieves short-term success in alleviating depression by helping patients to readjust their belief systems, long-term therapy may be required to overcome fundamental emotional conflicts.

6.2 EVIDENCE FOR EMOTIONAL-COHERENCE THEORY

What reason is there to believe the account of emotional coherence presented above? First, the theory of emotional coherence provides a unified explanation of numerous diverse psychological phenomena of great theoretical and practical importance. This chapter has provided a qualitative account of emotional coherence and has also shown how the theory can be implemented in a computational model with well-defined structures and processes that illuminate phenomena ranging from trust to cognitive therapy. Second, introspections and anecdotes support the view that trust, distrust, and nationalism have an emotional component, and empathy is by definition emotional. It would be desirable to go beyond introspection to show that the theory of emotional coherence can explain more quantitatively the results of psychological experiments, but the relevant experiments concerning the emotional impact of coherence have not been done. Independent of issues of emotion, the theories of explanatory, conceptual, and analogical coherence have had substantial empirical applications, serving to explain a wide variety of results of psychological experiments. Hence there is psychological evidence that inference is coherence-based, but the theory of emotional coherence awaits experimental test. The third support for the theory of emotional coherence comes from recent results in neuroscience, which were in part its inspiration.

In his provocative book *Descartes' Error* (1994), Damasio describes a group of patients with damage to the ventromedial region of the brain's frontal lobe. Such patients are typically physically capable and have most of their mental capacities intact, but their behavior, the result

of severely flawed decision making, can be very odd. Elliott, for example, had been a good husband and father and successful in business. Then surgery to remove a tumor in the ventromedial area had left him apparently intellectually intact, but prone to decisions that proved disastrous for both his career and his marriage. Damasio argues that the importance of the ventromedial prefrontal cortices derives from their role in linking cognitive information about particular situations with signals that he calls "somatic markers," which are body states mediated by the emotional centers of the brain: the hypothalamus and amygdala. To put it briefly, the problem with ventromedial-damaged patients is that their decisions are cut off from their emotions, with the result that they have lost touch with what really matters to them.

Damasio's views map nicely onto my theory of emotional coherence. Valence inputs can be interpreted as based on somatic markers that the amygdala associates positively or negatively with particular things or situations. Interactions between the amygdala and the frontal cortex, where coherence-based inferences are presumably made, generate somatic markers that correspond to positive or negative valence outputs. In the case of trusting a babysitter, these somatic markers correspond to my "gut feeling" that a taxi driver was not to be trusted with my son Adam but that Christine was. In terms of HOTCO, the problem with Damasio's patients with ventromedial damage is that their coherence calculations have become severed from valence inputs and outputs.

According to LeDoux (1996), the amygdala has projections to many cortical areas, and the amygdala has a greater influence on the cortex than the cortex has on the amygdala. These influences support the assumption of the HOTCO model that coherence-based activations and emotion-generating valences are intertwined. Inference and

appraisal go hand in hand, with emotional appraisal of a situation evolving in parallel with inferences about it. The HOTCO model is also consistent with the thorough review of affective neuroscience by Panksepp (1998), who observes, "The emotional systems are centrally placed to coordinate many higher and lower brain activities" (1998, 27), and notes, "Affective and cognitive processes are inextricably intertwined in higher brain areas, such as the frontal and temporal cortices" (1998, 315). LeDoux also reports, however, that not all emotional reactions require cortical processing, for there is a direct connection from the sensory thalamus to the amygdala. Hence for some visual stimuli, preferences need no inferences (Zajonc 1980). My account of emotional coherence applies only when appraisal is based on complex inferences, not to more direct emotional reactions to salient perceptual stimuli.

Psychological experiments are required to evaluate the plausibility of the following theoretical claims concerning judgments involving representational elements:

- The valence of an element depends on both the valences and the acceptability of the elements connected to it by coherence relations.
- Explanatory, conceptual, and analogical coherence all contribute to the resulting valence of an element.
- Judgments of trust are inherently emotional and are affected by input valences.
- Judgments of trust are also affected by explanatory, conceptual, and analogical coherence considerations that contribute to output valences.

My colleagues and I are currently planning experiments that manipulate valences and coherence to determine their effect on emotional judgments about people.

13 NORMATIVE CONSIDERATIONS

Since Plato and Aristotle, philosophical and popular thought have generally assumed a contrast between rationality on the one hand and emotion on the other. This divide, however, has been challenged by such writers as de Sousa (1987), Frank (1988), Oatley (1992), and Stocker and Hegeman (1996). My concern in this chapter has largely been to give a descriptive theory of trust and other applications of emotional coherence, but in naturalistic philosophy of the sort I practice, the descriptive and the normative are closely intertwined (Thagard 1988, 1992b).

I see three reasons for considering emotional coherence as being prescriptive as well as descriptive of trust, telling us generally when we *should* trust people as well as when we do. First, the standard models of rationality in decision making have little application to real-life situations. It is usually not possible to perform expected-value calculations based on probabilities and utilities, because we rarely know the relevant probabilities and utilities, which are dubious psychological constructs in comparison with goals and emotions. Second, it is a standard normative principle that ought implies can, so that no one can be held responsible for not doing the impossible. You cannot turn off your amygdala: removing emotions from decisions is psychologically impossible, although there are undoubtedly steps that can be taken to dampen the effects of destructive emotions. So normative principles ought not to require that we eliminate emotions from decisions. Third, if Damasio is right, you may not want to turn off your amygdala, because to do so would cut your analytical decision making off from crucial emotional information about what really matters to you. For human beings, emotion-free decision making is likely to be highly defective decision making, contrary to what you might believe from the Star Trek characters Mr. Spock and Data, who purport to possess cold rationality.

I am not, of course, romantically espousing uncritical guidance by emotional intuitions, which may be of dubious quality. Explanatory, analogical, and conceptual coherence can all be viewed normatively as well as descriptively, and there are better and worse ways of performing inferences based on them. For example, explanatory inference based on neglect of alternative explanatory hypotheses is likely to lead to premature acceptance of weak hypotheses. The normative course I recommend, well within people's capabilities, is the integration of emotional inputs with coherence-based inference to yield emotionally marked and objectively desirable outcomes.

There are important cases where emotional coherence may be in conflict with other kinds of coherence. Consider, for example, the presidential candidate Jack Stanton in the novel *Primary Colors* (1996). Stanton is presented as having two major weaknesses: women and fast food. He knows that it is better for his health and appearance if he avoids doughnuts and other unhealthy foods, but frequently eats them anyway. Similarly, his womanizing is a threat to his marriage and his political ambitions, both of which he presumably values more highly than his illicit affairs, yet he seems incapable of acting in his best interests. It is easy to see from the theory of emotional coherence and the computational model HOTCO how weakness of will can arise. Valences are affected not only by permanent, reasoned valences attached to goals such as being healthy, slim, faithful, and politically successful, but also by the activation of the relevant nodes. Emotion input can be of two kinds, the first arising from reasoned judgments of the value of a goal, such as eating doughnuts, the second arising from physiological reactions to a stimulus, such as a box of

doughnuts. When faced with the doughnuts, or perhaps just the thought of the doughnuts, Stanton's doughnut node becomes strongly activated physiologically, so that the node representing the action *eat doughnuts* receives a strong valence. Deliberative coherence is swamped by emotional coherence, and this results in normatively inappropriate weakness of will. Similarly, as in my Gypsy example, social prejudice based on negative stereotypes may lead to irrational actions. Thus emotional coherence may generate normatively inappropriate judgments and behavior, although it may also be an important component in integrative reactions to complex situations.

Religious belief may survive because of the comfort and hope that it provides, despite the lack of evidence for it (chapter 4). Belief in God can be a great consolation, bringing assurance that everything will work out in one's life and that existence continues after death. Hence theism survives because of its emotional appeal, as well as because of the transmission of religious traditions from parents to children. Normatively, however, metaphysical hypotheses such as the existence of God should be evaluated on the basis of their coherence with evidence, not on the basis of desirability or tradition. HOTCO currently allows activations to influence valences, but does not allow valences to influence activations, so that the desirability of a conclusion does not have an effect on its acceptability. It is therefore incapable of modeling wishful thinking or the kind of motivated inference discussed by Kunda (1987, 1990).

Currently, HOTCO allows the activation of units, which represents the acceptance or believability of elements, to influence valences, which represent emotional attitudes toward the elements. Influence in the other direction is obviously dangerous: we should not believe something just because it makes us happy to do so. But recent

experiments by Ziva Kunda, Drew Westen, and their colleagues show that emotional attitudes can have a strong influence on factual inferences, and HOTCO can be extended to allow valences to influence activations.

Sinclair and Kunda (1999) have found that the motivation to form a particular impression of an individual can prompt the inhibition of applicable stereotypes that contradict one's desired impression and the activation and application of those that support it. For example, experimental participants who were prejudiced against Blacks inhibited the negative Black stereotype when motivated to esteem a Black individual because he had praised them. In contrast, participants motivated to disparage a Black individual because he had criticized them did apply the Black stereotype, rating the individual as relatively incompetent. Thus inference about a person's competence can be affected by whether it is in one's self-interest to view him as competent or incompetent.

In terms of HOTCO, the experimental results of Sinclair and Kunda can be interpreted as showing that valences measuring the desirability of an inference can influence the activations measuring the plausibility of the inference. The simplest way to allow valences to influence activations would be to rewrite the equation for updating activations to take valences into account. Then the activation of a unit would be a function both of input activations and of the valence of the unit. Belief would then depend directly on positive feeling. However, in Kunda's (1990) work on motivated reasoning, people do not just believe something because it makes them happy: they have to do extra cognitive work to retrieve memories that support their desired beliefs. The results of Sinclair and Kunda (1999) suggest that in addition to a motivated memory search, there is a more direct process whereby valences can sometimes influence activations.

To model this process, I plan to add to HOTCO a special class of units, called "evaluation units," which correspond to representations that have both a cognitive and an affective dimension. For example, the proposition *Frank is good* has a degree of belief, but it also has an intimate connection with the affect attached to the representation *Frank*. The unit representing *Frank is good* should thus have its activation influenced both by other activations (e.g., of the unit *Frank is a criminal*) and by the valence of the associated unit representing *Frank*. I propose that the activations of evaluation units should be a function both of input activations and of input valences from associated units. Conversely, the valences of units such as *Frank* should be influenced by the activations of evaluation units such as *Frank is good*.

Similarly, I conjecture that the participants in the experiments of Sinclair and Kunda have an evaluation unit for *I am good* that has an activation that depends in part on the valence of the correlative *I* unit. The ongoing positive valence of the *I* unit will tend to keep the *I am good* unit active, which in turn will tend to support a unit representing the belief that the Black individual's praise of the participant was accurate. This *Praise is accurate* unit is positively linked to a unit asserting the competence of the individual, which is negatively linked to the participant's negative Black stereotype of incompetence. Thus the positive valence attached to the *I am good* node will tend to inhibit application of the negative aspects of the Black stereotype. On the other hand, if the participant is criticized by the Black professional, then maintaining a positive valence for the *I* unit will encourage judging the professional to be incompetent.

A similar mechanism should be able to account for experimental results of Westen and Feit (forthcoming). They studied people's inferences during the Clinton-Lewinsky scandal of 1998, and they found that political judgments bore minimal relation to knowledge of relevant data, but were strongly predicted by people's feelings about Democrats and Republicans, Clinton, feminism, and infidelity. Factual hypotheses concerning what Clinton did or did not do should have been evaluated solely on the basis of their fit with the available evidence. But Westen and Feit's data suggest that the evidence had a small influence on people's inferences in comparison with their positive or negative feelings about Clinton and the two political parties involved. I propose to account for these results by giving HOTCO a *Clinton* unit with positive or negative valence that influences the activation of a unit for *Clinton is good*. Then the positive activation of this unit will tend to suppress the activation of a unit representing the hypothesis that Clinton is guilty, and hence to support alternative explanations of why witnesses said what they did about Clinton, for example, that they were encouraged by the Republicans. The explanatory coherence of the hypothesis about Clinton's guilt will thus be directly affected by the emotional coherence of that conclusion, just as people's confidence in the existence of God is determined by its emotional desirability. Similarly, in the O. J. Simpson trial, some jurors may have been influenced in their assessment of the evidence by their motivation to view Simpson as a good person and their emotional attitudes toward the Los Angeles police. I plan to develop a computational model of motivated inference that will apply to biased reasoning in law and science as well as to the experiments of Kunda, Westen, and their colleagues.

Trust, empathy, and the other topics of this chapter are by no means the only psychological phenomena that the theory of emotional coherence might help to explain. As chapter 5 described, judgments of right and wrong are based on interrelated explanatory, analogical, deductive,

conceptual, and deliberative considerations. It is evident from both personal introspection and the behavior of others that ethical judgments are also often highly emotional. The emotive and cognitive aspects of ethical judgment are usually treated by philosophers as orthogonal to each other, but the theory of emotional coherence shows how they can be brought back together. In the 1940s, philosophers influenced by logical positivism espoused emotivism, the doctrine that value judgments in ethics and aesthetics merely express emotions. The theory of emotional coherence shows how ethical and other value judgments can be simultaneously emotionally and cognitively coherent. The theory of emotional coherence and the computational model HOTCO are limited in that they do not deal with the full variety of human emotional responses, but they serve to show how inference can at least sometimes be both emotional and rational. Cognitive naturalism can thus take into account the affective side of human thinking as well as the cold, inferential side. The next chapter will show that cognitive naturalism can also take into account some of the social dimensions of knowledge.

14 SUMMARY

Inference often involves not only accepting or rejecting mental representations, but also adjusting positive and negative emotional attitudes towards what is represented. Trust is based on explanatory and other kinds of coherence, but it also involves acquiring an emotional attitude or valence associated with the object to be trusted. Acquiring a valence is a parallel constraint-satisfaction process much like the process of accepting or rejecting representations based on their coherence with other representations. The HOTCO model shows how emotional assessment can be integrated with explanatory and other kinds of coherence to produce judgments of trust and other value-laden decisions, such as those involved in empathy and nationalism. Emotions can also involve more general kinds of evaluations that require an overall assessment of how much coherence is achieved. Such metacoherence assessments are relevant to understanding beauty, symmetry, humor, and the mood changes that occur as the result of cognitive therapy.