

Chapter 17

Values in Science: Cognitive-Affective Maps

Paul Thagard

University of Waterloo

From P. Thagard (2012). *The cognitive science of science: Explanation, discovery, and conceptual change*. Cambridge, MA: MIT Press.

Introduction

The cognitive science of science has largely concentrated on the nature of the concepts, theories, and explanations that scientists develop and students learn. There has been a general neglect of the *values* that are also part of scientific practice. The development of science has depended on valuing observations, experiments, and objectivity in the assessment of scientific hypotheses. The adoption of these values was a crucial part of the Scientific Revolution and the Enlightenment of the 16th through 18th centuries, and the values have remained important in the current practice of science. Today, however, they are challenged from various directions, ranging from the religious right who prefer faith over evidence, to the postmodernist left who decline to “privilege” science and rationality over obscurity and irony.

Whereas the role of values has largely been neglected in psychological and computational studies of science, philosophers of science have become increasingly concerned with issues about values (e.g. Douglas 2009, Lacey 1999, Longino 1990). The major debate concerns whether there should be a role for social and ethical values in the assessment of scientific theories. The traditional view is that science should be “value free”, which means that it is restricted to epistemic values such as evidence and truth, ignoring social values that might lead to biases in scientific research. A common worry is that allowing an incursion of social values into science will produce corruptions like

those witnessed frequently in the history of science, for example the racist distortions in Nazi science and the sexist distortions in nineteenth century neuroscience. In response, philosophers such as Douglas (2009) argue that values inescapably and legitimately play a role in scientific deliberations, and the key to improving science is to make application of these values rationally appropriate as opposed to irrationally biased.

This chapter explores what cognitive science can contribute to this normative enterprise of finding an appropriate place for values in science by addressing questions that are largely ignored in philosophical discussions:

1. What are values?
2. How can values bias scientific deliberations?
3. How can values legitimately contribute to scientific inferences?

I will argue first that values are emotionally valenced mental representations, and show how interconnected groups of values can be represented by cognitive-affective maps, a new technique for displaying concepts and their emotional connotations. Second, I will use cognitive-affective maps to show how values can distort scientific deliberations in ways that diminish rationality. Third, more positively, I will describe how values can legitimately affect scientific developments at all three stages of research: pursuit of discoveries, evaluation of results, and practical applications.

What are Values

Most scientists operate with epistemic (knowledge-related) values such as evidence and truth, and also with social values such as human welfare. We can contrast three opposed views of what values are: ideal structures, social constructions, and mental representations. On the first view, values are ideal structures such as Platonic forms or

abstract linguistic structures. On the second view, values are mere complexes of behavior that result from social interactions. On the third view that I want to defend here, values are mental representations with emotionally valenced meanings. Here “valence” is the term psychologists use to indicate emotionally positive associations (e. g. to concepts such as *baby*) or negative associations (e. g. to concepts such as *death*).

Consider, for example, the value *truth*. On the ideal structure view, truth is an abstract concept that exists independently of human minds, and the value attached to it is an abstract relation between a person (also an ideal entity) and the abstract concept. Philosophers often write of propositional attitudes which are supposed to be relations between two kinds of abstract entities: persons and propositions, which are the meanings of sentences. Analogously, the ideal structure view of values can understand them as “conceptual attitudes”, that is, relations between persons and the abstract meanings of concepts.

In contrast, the social construction understanding of truth views it as a way of talking that emerges from interactions among people. Shapin (1994) writes of the “social history of truth”, taking truth to be a product of collective action rather than correspondence to reality. From this perspective, truth and other scientific values are just preferences that people happen to have as the result of their social circumstances.

Cognitive science offers an alternative to ideal structure and social construction accounts of the nature of values. As chapter 1 described, the standard assumption in cognitive science is that inferences such as decisions result from mental representations and processes. A mental representation is a structure or process in the mind that stands for something. Typical examples are concepts which stand for classes of things,

propositions which stand for states of affairs, and rules which stand for regularities in the world. In contrast to the ideal structure position, concepts and other mental representations are not abstract entities, but rather parts of the world as brain processes in individual humans (Thagard 2010a). As chapter 18 argues, the structure of a concept can be highly complex, with pointers to visual and emotional information.

The emotional component of concepts and rules shows how values can be mental representations. Consider, for example, the concept of a car. Brain representations of cars include verbal information such as that a car is a kind of vehicle and that there are many kinds of cars, for example, sedans and convertibles. As theorists such as Barsalou (et al. 2003) have argued, the representation of the concept of a car also includes sensory information, such as images of the visual appearance of a car, as well as its sound and perhaps even its smell. For some people, cars may be emotionally neutral, but for many people the mental representation of a car will include either a positive or negative emotional association. For example, a race car driver probably thinks of cars very positively, but an environmentalist concerned with pollution may well think of them very negatively. In psychological terms, we can say that cars have a positive valence for the racer, but a negative valence for the environmentalist. There are various ways in which valences can be modeled computationally, including the attachment of a valence to the units in a localist neural network (ch. 5, Thagard 2006), or as convolution between neural populations (ch. 8).

Because part of the representation of a concept is an emotional valence, values can be easily be understood as valenced representations. My valuing cars consists in having a mental representation of cars that includes a positive emotional valence arising

from activity in brain areas relevant to assessing gains, such as the nucleus accumbens. My disvaluing of pollution consists in my having a mental representation that includes a negative emotional valence arising from activity in brain areas relevant to assessing losses, such as the amygdala. For a neurocomputational model of assessment of gains and losses, see Litt, Eliasmith, and Thagard (2008).

The claim that values are emotionally valenced mental representations does not entail that they are purely subjective, just a matter of personal tastes. Concepts can vary in how well they represent the world: *tree*, *electron*, and *virus* are good examples of successful representations; *demon*, *phlogiston*, and *aether* are examples of representations that the march of science has discarded. Similarly, people can be wrong in their values, for example mistakenly believing that cheeseburgers are good for people. In line with Thagard (2010a), I hold that positive values are objectively right if they promote the satisfaction of real human needs, including both basic biological ones and also psychological ones such as relatedness to other people.

Mapping Values

Assuming that individual values consist of emotionally valenced mental representations, we can now consider the nature of systems of them. Values do not occur in isolation: scientists typically have connections among values such as truth, evidence, and objectivity, whereas a religious thinker might prefer a different set of values such as God, faith, and obedience.

To display scientific value systems, we can use a technique that I developed for understanding conflict resolution (Thagard 2010c, forthcoming; Findlay and Thagard forthcoming). Cognitive-affective maps are graphical diagrams that show both cognitive

structures (concepts, goals, beliefs) and their emotional valences; “affect” is the term used by psychologists and neuroscientists for matters related to emotion, mood, and motivation.

The following conventions are used in these maps. Elements (mental representations) are depicted by shapes:

- Ovals represent emotionally positive (pleasurable) elements.
- Hexagons represent emotionally negative (painful) elements.
- Rectangles represent elements that are neutral or carry both positive and negative aspects.

Straight lines depict relations between elements:

- Solid lines represent the relations between elements that are mutually supportive.
- Dashed lines represent the relations between elements that are incompatible with each other.

Moreover, the thickness of the lines in the shape represents the relative strength of the positive or negative value associated with it. Figure 17.1 illustrates the conventions used in representing values.

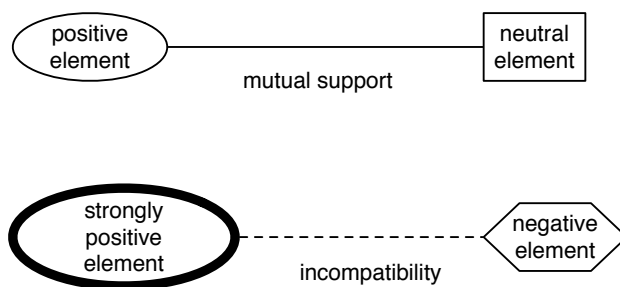


Figure 17.1 Conventions for cognitive-affective mapping.

Figure 17.2 is a cognitive-affective map of my own set of epistemic values, indicating positive values such as truth and objectivity, as well as conflicting negative

values such as error and bias. Other values not displayed include simplicity and analogy, both of which are part of my theory of explanatory coherence which I take to be the main route to truth and objectivity (ch. 5, 7). The structure in figure 17.2 is not just an abstract diagram, but also represents a neural network of the HOTCO sort (ch. 5) that can be used as a computational model of why scientists make the decisions that they do. For example, if religious people advocate ignoring the lack of solid evidence for a miraculous event, then the structure displayed in figure 17.2 will reject their claims, because of the positive value attached to evidence and the negative value attached to faith.

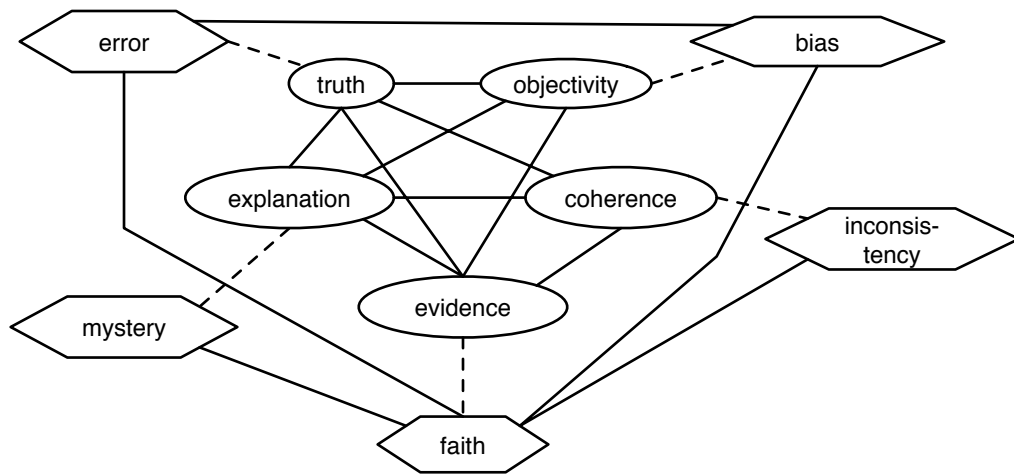


Figure 17.2 Epistemic values relevant to science. Concepts in ovals are positive values, whereas concepts in hexagons are negative values. Solid lines indicate mutual support, and dotted lines indicate incompatibility.

Not every scientist or philosopher has the set of values shown in figure 17.2, but I view this group as defensible for reasons set out in many publications (e.g. Thagard 1988, 1992, 1999, 2000). Others might choose to emphasize other values such as prediction,

falsification, or empirical adequacy; I invite readers to draw their own cognitive-affective maps expressing their positive and negative values.

The same technique can be used to map social and ethical values, as shown in figure 17.3. This particular set of values is my own, reflecting an ethical view (needs-based consequentialism) defended elsewhere (Thagard 2010a). These values differ from those of people who base their ethics on religion or a priori arguments. On my view, social values arise from the empirical fact that people universally have a set of biological and psychological needs. The biological needs are easiest to establish, as people obviously suffer severe harm if deprived of food, water, shelter, and health care. More contentious are the primary psychological needs of relatedness, competence, and autonomy, but there is evidence that these are also vitally important for human wellbeing (Deci and Ryan 2002, Thagard 2010a). Competence is the need to engage optimal challenges and experience physical and social mastery. Autonomy is the need to self-organize and regulate one's own behavior and avoid control by others. Relatedness is the need for social attachments and feelings of security, belongingness, and intimacy with others. Figure 17.3 shows the internal coherence among vital human needs and their incompatibility with a host of human conditions I disvalue, such as malnutrition and oppression.

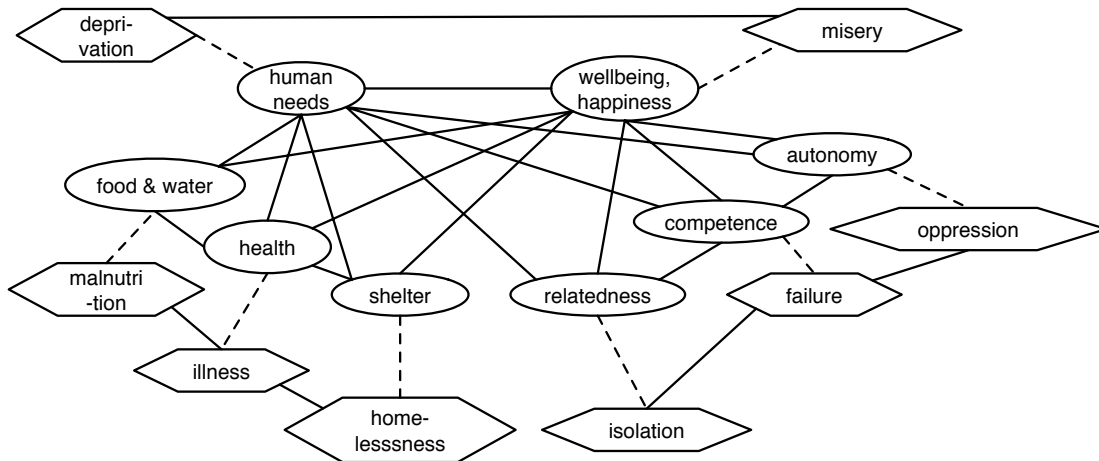


Figure 17.3 Social and ethical values related to satisfaction of human needs. Ovals indicate positive values, and hexagons indicate negative values.

Displaying sets of epistemic values (figure 17.2) and social values (figure 17.3), makes it possible to explore how they might interact in decisions concerning science. In practice, however, the decisions of particular scientists are not driven only by epistemic and social values, but also by personal goals such as success, fame, and prosperity. Figure 17.4 displays some of the personal values that can influence scientific decisions. This list is intended to be descriptive (what people often do value) rather than normative (what they ought to value). Normative assessment would require evaluation of the personal values shown in figure 17.4 with respect to the ethical values represented in figure 17.3. There are some obvious connections: wealth helps to satisfy the biological needs for food and shelter, and family helps to satisfy the psychological need for relatedness. The big difference, however, is that the social values in figure 17.3 are intended to take into account all the people affected by a decision, whereas the personal values in figure 17.4 operate just at the individual level.

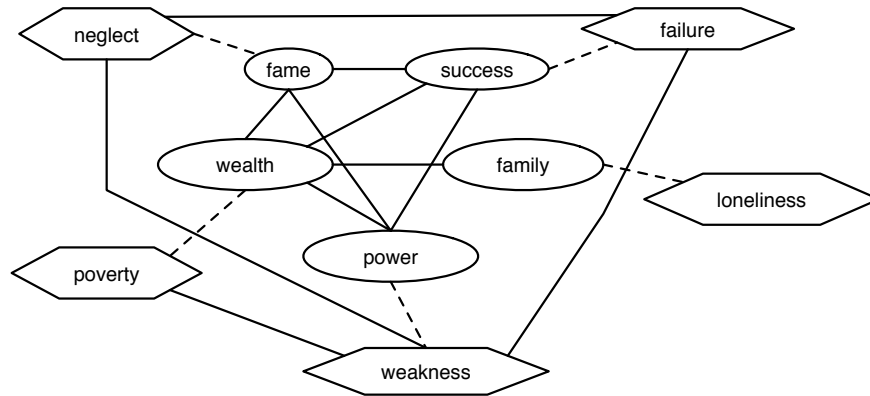


Figure 17.4 Personal values of some scientists. Values that are positive for a particular scientist are represented by ovals, and negative values are represented by hexagons. The different priority that different individuals place on these values could be shown by thickening the lines in the ovals and hexagons.

Appropriate Uses of Values in Science

I will try now to develop a general model of how the three kinds of values – epistemic, social, and personal – do and should interact. Then it will be possible to identify cases where science becomes biased because epistemic values are slighted in favor of social or personal values.

Different kinds of values are relevant to different contexts in scientific research and development. Philosophers of science commonly distinguish among the contexts of pursuit, evaluation, and application of scientific ideas. The context of pursuit includes deciding what research projects to pursue, setting goals, carrying out experiments, and generating new explanatory hypotheses. The context of pursuit includes discovery, but is much broader because it includes all the planning that needs to go into figuring out where discoveries might be achieved. The context of evaluation (also called the context

of justification) assumes that evidence has been collected and hypotheses have been generated. According to the view presented in Part II, evaluation consists of inferences to the best explanation in which hypotheses are accepted or rejected depending on their explanatory coherence with respect to all the available evidence and competing hypotheses. Finally, the context of application includes decisions about how to use validated scientific ideas for practical purposes such as medicine, engineering, and social policy.

The context of pursuit legitimately mingles epistemic, social, and personal values. Scientists who are putting together a research plan can appropriately take into account *all* of the following questions concerning a potential project:

1. To what extent is the project likely to contribute to epistemic goals such as accumulating evidence and leading to the acceptance of explanatory theories?
2. To what extent is the project likely to contribute to human wellbeing as the result of satisfaction of human needs?
3. To what extent will the project further the professional careers of the scientists involved?

Different kinds of research will vary with respect to the importance of epistemic and social goals. To take examples from cognitive science, we can distinguish between basic research about fundamental psychological and neural processes, on the one hand, and applied research aimed at improving practices in education, mental health, and other areas relevant to human wellbeing.

Should scientists take into account their own professional success in deciding what research projects to pursue? There are at least two reasons for allowing personal

values to be appropriate in the context of pursuit. First, to disallow them would be to insist that scientists be moral and epistemic saints, always ignoring their own interests in favor of general ethical and scientific goals. Few scientists are psychologically capable of either moral or epistemic sainthood, so a normative stance that requires perfection is unreasonable.

Second, eliminating the pursuit of personal interests in science may inhibit rather than enhance the accomplishment of the epistemic and social goals. Scientific work is often arduous and frustrating, and the personal motivations that come from dreams of personal success may be part of the motivation that keeps researchers toiling in the lab and stewing up new ideas. Moreover, the goal of personal success has the advantage of promoting diversity, encouraging researchers to undertake novel and risky research projects: there is more glory from work that is recognized as highly original.

The same considerations legitimize the pursuit of personal goals in the context of application. Turning ideas from physics, chemistry, biology, or cognitive science into socially valuable projects or commercial products naturally involves both epistemic values such as working with true theories and social values such as promoting wellbeing. But people applying science may also be driven by personal goals such as wealth, for example technologists who apply scientific ideas to become billionaires. Without sinking to the egoistic view that only greed is good, we can still recognize that personal values can contribute to the overall success of applied projects because of their contributions to motivation, diversity, and novelty. The prospect of personal success can strongly encourage applied scientists such as technologists to work extraordinarily hard on risky projects that may have large payoffs, including ones that promote social wellbeing and

may even stretch scientific theory in valuable directions. Of course, personal and social goals can also distort or interfere with epistemic ones, in the pursuit and application of science; the next section will consider psychological reasons why biases occur.

If the view I have been developing is correct, then decisions about the pursuit and application of science are multifactorial, involving consideration of the extent to which different projects promote the satisfaction of a full range of epistemic, social, and personal values. Descriptively and normatively, the process of decision is not reducible to some straightforward algorithm such as maximizing expected utility, but requires rather a complex process of constraint satisfaction that can include revising goals as well as choosing actions (Thagard 2010a, ch. 6). Figure 17.5 shows a decision problem which is neutral between two projects, but which could be settled if there were stronger links of support between one project and the various kinds of goals.

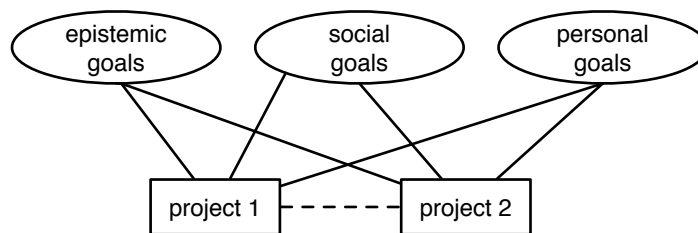


Figure 17.5. Map of decision between two competing projects, taking into account three kinds of goals. Ovals indicate positive values, and rectangles indicate neutrality. Solid lines indicate support, and the dotted line indicates incompatibility.

I have argued that social and personal values legitimately contribute to scientific decisions in the contexts of pursuit and application, but the role of values in the context of evaluation is much more contentious. The claim that social and personal values

should have *no* impact on evaluation and acceptance of theories has substantial appeal. Normative models for scientific inference (including explanatory coherence, Bayesian probability, and the confirmation theory of the logical positivists) all legislate that the evaluation of hypotheses should be done on the basis of evidence and reasoning alone, with no value component. We should accept theories that fit best with the evidence, and reject the alternatives. In all these schemes there is no room for the social and personal contributions to decision making shown in figure 17.5. Instead, the decision situation is limited to the sparse arrangement mapped in figure 17.6.

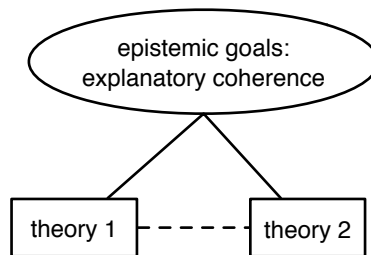


Figure 17.6 Theory acceptance as a decision problem concerning only epistemic goals.

In contrast, Richard Rudner (1961, pp. 32-33) argued that social goals can legitimately play an indirect role in the assessment of hypotheses:

Since no scientific hypothesis is ever completely verified, in accepting a hypothesis on the basis of evidence, the scientist must make the decision that the evidence is sufficiently strong or that the probability is sufficiently high to warrant the acceptance of the hypothesis. Obviously, our decision with regard to the evidence and how strong is "strong enough" is going to be a function of the importance, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis. Thus, to take a crude but

easily manageable example, if the hypothesis under consideration stated that a toxic ingredient of a drug was not present in lethal quantity, then we would require a relatively high degree of confirmation or confidence before accepting the hypothesis-for the consequences of making a mistake here are exceedingly grave by our moral standards. In contrast, if our hypothesis stated that, on the basis of some sample, a certain lot of machine-stamped belt buckles was not defective, the degree of confidence we would require would be relatively lower. How sure we must be before we accept a hypothesis depends on how serious a mistake would be.

Here social goals do not contribute directly to the evaluation of the evidence for and against scientific theories, but do contribute indirectly by setting a threshold for acceptance, making it higher in cases where acceptance might have harmful consequences. Douglas (2009) defends at greater length a similar position. Neither Rudner nor Douglas proposes the opposite of this harm principle - a benefit principle that would allow hypotheses that would be beneficial if true to be accepted with a lower standard of evidence.

A similar issue arises in the law (Thagard 2006, ch. 9). In the British legal system that also operates in the U.S., Canada, and some other countries, people accused of crimes are not supposed to be convicted unless the prosecution shows that they are guilty *beyond a reasonable doubt*. In legal discussions, there is no accepted definition of what constitutes a reasonable doubt, but my preferred interpretation is shown in figure 17.7. In civil trials, decisions are based on *a preponderance of evidence*, which allows a hypothesis concerning responsibility to be evaluated merely on whether the evidence is

maximally coherent with that hypothesis or its negation. In criminal trials, however, there is a presumption that convicting an innocent person is a harm to be avoided, even at the risk of letting a guilty person go. There are other ways in which reasonable doubt might be understood, for example as requiring a higher probability, but problems abound with applying probability theory in legal contexts (Thagard 2004). As shown in figure 17.7, the hypotheses that the accused is innocent and that the accused is guilty are both evaluated with respect to how well they explain the evidence presented to the court, but the conclusion of guilt is somewhat inhibited by the connection to the negative value of convicting the innocent. Hence conviction requires stronger evidence than acquittal.

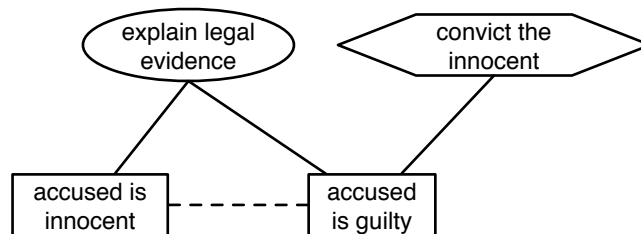


Figure 17.7 Map of reasonable doubt in criminal trials: acceptance of the hypothesis that the accused is guilty is inhibited by the negative value attached to convicting the innocent.

By analogy, we could understand scientific inference employing the harm principle of Rudner and Douglas as discouraging acceptance of a hypothesis if there is a risk of harm associated with it, as in the judgment that a chemical is not toxic. Figure 17.8 shows how this might work, introducing a presumption against accepting a hypothesis where the acceptance might lead to actions that cause harm. At this point, we have two thirds of the values shown in figure 17.5, as social values have once more been introduced into what figure 17.6 tries to mark as a purely epistemic exercise.

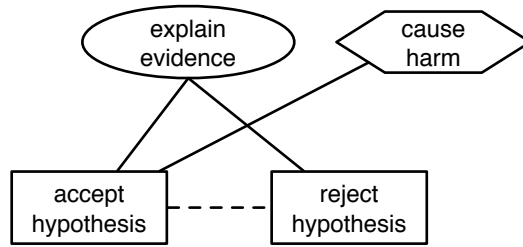


Figure 17.8 Map of caution in scientific inference: acceptance of the hypothesis is inhibited by the harm it might lead to.

Disturbingly, however, the legal analogy seems to break down, because the British legal principles assume that convicting the innocent is a special harm, and does not worry explicitly about the risks associated with acquitting the guilty. In contrast, science seems to lack the asymmetry between accepting a hypothesis that causes harm and not accepting a hypothesis that might bring good. In some cases, rejecting a hypothesis that is potentially relevant to human needs can also cause harm, for example if a drug is not approved that might have saved lives. Or consider climate change as discussed in chapter 5. Accepting the hypothesis that human behavior is a major cause of climate change has potentially negative consequences if it leads to actions such as reducing climate emissions that limit economic growth and so increase poverty. However, the status quo may be even more dangerous, if greater carbon emissions contribute to still more global warming, leading to disasters such as drought and flooding. So where do we end up? Perhaps figure 17.9 portrays the most appropriate values, allowing that social goals can play a small role, subordinate to epistemic goals, in determining what gets accepted and rejected.

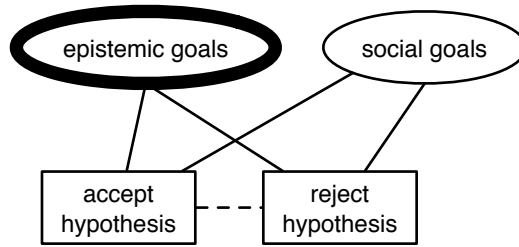


Figure 17.9. Social values affect the acceptance and rejection of hypotheses, but to a lesser extent than epistemic goals. The links from social goals to the nodes for accepting and rejecting the hypothesis may be positive or negative, depending on the risks and benefits of the hypothesis. Thicker lines indicate higher valence.

What about personal values, whose incorporation into figure 17.9 would restore it to the same structure as figure 17.5? I can think of no good arguments why personal goals should play a role in hypothesis evaluation, which does not need the motivational and diversity concerns that support the inclusion of personal goals in the contexts of pursuit and application. The objective social aims of science are legitimate values relevant to establishing thresholds for acceptance and rejection of hypotheses, but personal values have no such legitimacy and can produce the biases discussed in the last section.

It is dangerous, however, to allow even a small role for social goals as in figure 17.9, because it opens the context of evaluation to serious biases. Examination of the ways in which social and personal goals have historically led to severe corruptions of scientific values may shed light on whether the value map in figure 17.9 is indeed the most appropriate one for the context of evaluation.

Biases in Scientific Inference

A defender of the ideal that evaluation of scientific hypotheses should concern only epistemic values can point to many disastrous incursions of the social into the scientific, such as the Nazi rejection of Einstein’s relativity theory as Jewish science. Cognitive science can provide insights into the mental processes by which such biasing occurs, and thereby help to provide guidance about how to encourage people to follow normatively superior practices.

First we need a breakdown of ways in which epistemic and social values can interact. Table 17.1 displays four possibilities of ways in which good and bad values can interact. By good epistemic values I mean concepts such as truth and objectivity shown in ovals in figure 17.2. By good social values I mean concepts such as needs and wellbeing shown in ovals in figure 17.3. A great deal of science fits in the optimal quadrant 1, which combines good epistemic values and good social values. For example, much of modern medical science belongs here, as exemplified by the dramatic progress in understanding and treating infectious diseases: well-validated theories about the how bacteria and viruses invade cells have contributed to dramatic lessening in deaths from infections.

	Good epistemic values	Bad epistemic values
Good social values	1. Much of science, e.g. medical theories of infection	2. Lysenko’s rejection of genetics
Bad social values	3. Nazi rocket science	4. Nazi rejection of Einstein

Table 17.1 Examples of interactions of good and bad social and epistemic values.

Quadrant 2 shows the rarer kind of case where good social values lead to the suspension of good epistemic values. The closest case I can think of is the Soviet Union's espousal of Lamarckian evolutionary theory under Lysenko, which was partly motivated by egalitarian concerns about social improvement: the hope was that intentionally altered environments could change the heritable nature of organisms ranging from plants to people. Unfortunately, this motivation led to neglect of the substantial amount of evidence supporting Mendelian genetics. The case of Lysenko was not, however, just about social goals, as it also involved the pursuit of personal goals such as power.

Other instances for quadrant 2 occur in some postmodernist critiques of science where justifiable concern with the mistreatment of people on the basis of their sex, race, and colonial status has led to the blanket rejection of Enlightenment values of truth and objectivity. Much alternative medicine also seems to fit in quadrant 2, where the desire to meet human needs for health leads people to ignore scientific procedures for evaluating medical treatments, such as controlled clinical trials.

Quadrant 2 encompasses the all-too-common cases where effective science is used for evil purposes. Nazi rocket science was scientifically strong, as shown by the recruitment of leading German scientists such as Werner von Braun by the Americans and Soviets after World War II. But the aims of the Nazi investment in rocketry was world domination, not satisfaction of universal human needs. Similarly, many other kinds of applied science from atomic weaponry to germ warfare employ epistemically good science in the service of aims inimical to general human welfare.

By far the most cognitively interesting and socially troubling kind of situation is shown in quadrant 4, where bad social values combine with bad epistemic values to produce atrocious results. Much pseudoscience fits in this category, where unscrupulous manipulators pursue their own personal goals of wealth and fame and promulgate ideas that are not only epistemically dubious but also harmful to those who adopt the ideas. A psychological example is the “Secret” that people can get whatever they want just by wishing for it, an idea that has generated millions of book sales by beguiling people into overestimating the control they have over the universe. Climate change denial of the sort analyzed in chapter 5 also falls into this quadrant, when politicians in the pockets of oil companies distort science in order to pursue policies that in the long run hand hurt human welfare.

All of the cases in quadrants 2-4 arise from the cognitive-affective process of motivated inference that was discussed in chapter 5 concerning climate change. Recall that motivated inference is not just wishful thinking, but leads to more subtle distortions through selective use of evidence, not total abandonment. Social goal or personal goals eclipse epistemic goals without completely eradicating them, except in the most extreme cases where complete nihilism about evidence and reality is advocated. Figure 17.10 shows the general case where motivated inference allows social and personal values to eclipse epistemic ones, producing the pathologies shown in quadrants 2-4 of table 17.1.

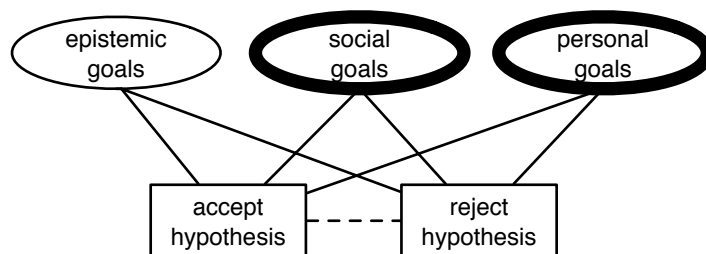


Figure 17.10 Motivated inference: Eclipse of epistemic goals by social and/or personal goals. Thick ovals indicate higher emotional valences.

Motivated inference is only one of the ways in which emotions can distort thinking; other affective afflictions include self-deception, weakness of will, conflicts of interest, depression, and manic exuberance (Thagard 2007, 2010a). There is an additional affective affliction of great importance that has received surprisingly little attention from psychologists and philosophers: countermotivated inference (Elster 2007). Whereas in motivated inference people tend to believe what they desire, in countermotivated inference people tend to believe what they fear. This way of fixing beliefs sounds ridiculous, but it is very common in everyday situations such as jealous spouses, nervous parents, political paranoids, economic doomsayers, anxious academics, and hypochondriacs. Fear leads people to focus on scant evidence for a dangerous outcome, illegitimately increasing their belief in the outcome. I am now developing a neurocomputational model of how this kind of fear-driven inference works.

Countermotivated inference can also occur in scientists and their critics. Science often leads to technologies whose consequences are hard to foresee, such as atomic power, recombinant DNA, and assisted reproduction. Sound science policy requires avoiding motivated inference in which the optimism of technologists sometimes leads them to ignore serious risks, but it also requires avoiding countermotivated inference in which people's ill-informed fears can lead them to be unduly convinced that a technology is dangerous. Examples might include excessive concerns about genetically engineered foods and about the Large Hadron Collider, which some commentators predicted would produce devastating black holes.

The susceptibility of human thinking to motivated and countermotivated inference naturally suggests swinging back to the traditional view that good reasoning should ignore emotion and instead employ normatively correct resources such as formal logic, probability theory, and utility theory. The problem with this reversion is that many important decisions in science and everyday life are hard to analyze using these tools, so that people naturally fall back on emotionally encoded values (Thagard 2006, ch. 2). There are no neat calculations that can tell us what to do about limiting climate change (ch. 5) or teaching biology (ch. 14). Hence the decisions that scientists and others need to make about what projects to pursue, what theories to accept, and what applications to enact will unavoidably have an emotional, value-laden aspect. Normatively, therefore, the best course is not to eliminate values and emotions, but to try to ensure that the best values are used in the most effective ways, avoiding affective afflictions.

One step toward effective reasoning is to make values explicit using cognitive-affective maps. Figure 17.10 shows generally how social and individual values can swamp epistemic ones, but the illumination of particular cases would require detailed identification of the specific values operating, including ones taken from figures 17.2-17.4. We could, for example, spell out what distorted social values led the Nazis to reject Einstein's theory. Cognitive-affective maps of values are only one tool useful for this task, supplementing the kind of propositional analysis of motivated inference that chapter 5 provided for climate change.

Another potential use for cognitive-affective maps is to display emotional conceptual changes like those described in Part IV. Someone who begins to appreciate the overwhelming evidence for Darwin's theory may change the valences associated with

the concept of evolution and associated concepts. It would be interesting to map emotional conceptual changes that sometimes occur in scientists as well as students. These may concern concepts specific to particular theories, but also the most general epistemic and social values such as truth and wellbeing.

Conclusion

This chapter has used cognitive-affective maps to illuminate the legitimate and sometimes illegitimate roles that values play in scientific thinking. The maps are not just a diagramming technique, but also an exercise in computational cognitive science: they derive from the HOTCO artificial neural network model of how inferences are made, including ones that can be biased by emotional influences.

I have rejected the doctrine that science should be value free, but more importantly have tried to show just how values do and should affect scientific decisions. Cognitive-affective maps have been useful for displaying the rich structure of epistemic, social, and personal values. Moreover, they show how all three kinds of values can legitimately contribute to decisions in the contexts of pursuit and application of scientific ideas. More subtly, the maps provide a way of seeing how social values can make a modest but valuable contribution in the context of evaluation, with inference taking into account the potential costs and benefits of accepting and rejecting hypotheses.

Just as important, cognitive-affective maps display how motivated inference can lead to distortions in scientific judgments when social or personal values play too large a role in the acceptance or rejection of scientific theories. Both good and bad social values can distort scientific inferences, in ways that cognitive science can explain by building

models of mental mechanisms that simulate cases where minds work well, and also cases where minds fall short of normative standards because of affective afflictions.

Values are irremovable from science because they are irremovable from the minds of scientists. Attempts to eliminate values from science are not only futile but counterproductive, because the contexts of pursuit and application need values to guide decisions, and even the context of evaluation has a modest role for values. Understanding how minds work in all these contexts is crucial for developing both descriptive and normative accounts of inference. An epistemologist or educator who ignores cognitive science is like an engineer who ignores physics, or a medical doctor who ignores biology. Similarly, administrators, politicians, and others concerned with science policy need to appreciate what the cognitive science of science can tell us about the mental nature of values in science and their effects on many kinds of decisions.

REFERENCES

- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84-91.
- Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. Rochester: University of Rochester Press.
- Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.
- Elster, J. (2007). *Explaining social behavior*. Cambridge: Cambridge University Press.
- Findlay, S. D., & Thagard, P. (forthcoming). Emotional change in international negotiation: Analyzing the Camp David accords using cognitive-affective maps. *Group Decision and Negotiation*.
- Lacey, H. (1999). *Is science value free? Values and scientific understanding*. London: Routledge.
- Litt, A., Eliasmith, C., & Thagard, P. (2008). Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research*, 9, 252-273.
- Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press.
- Shapin, S. (1994). *A social history of truth*. Chicago: University of Chicago Press.
- Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton, N.J.: Princeton University Press.
- Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.
- Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
- Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, 18, 231-249.
- Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.
- Thagard, P. (2007). The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy*, 24, 367-380.
- Thagard, P. (2010a). *The brain and the meaning of life*. Princeton, NJ: Princeton University Press.
- Thagard, P. (2010c). EMPATHICA: A computer support system with visual representations for cognitive-affective mapping. In K. McGregor (Ed.), *Proceedings of the workshop on visual reasoning and representation* (pp. 79-81). Menlo Park, CA: AAAI Press.
- Thagard, P. (forthcoming). Mapping minds across cultures. In R. Sun (Ed.), *Grounding social sciences in cognitive sciences*. Cambridge, MA: MIT Press.

The Cognitive Science of Science

Part I Introduction

1 What Is the Cognitive Science of Science?

Part II Explanation and Justification

2 Why Explanation Matters

3 Models of Scientific Explanation (with Abninder Litt)

4 How Brains Make Mental Models

5 Changing Minds about Climate Change: Belief Revision, Coherence, and Emotion
(with Scott Findlay)

6 Coherence, Truth, and the Development of Scientific Knowledge

Part III Discovery and Creativity

7 Why Discovery Matters

8 The Aha! Experience: Creativity Through Emergent Binding in Neural Networks (with Terry Stewart)

9 Creative Combination of Representations: Scientific Discovery and Technological Invention

10 Creativity in Computer Science (with Daniel Saunders)

11 Patterns of Medical Discovery

Part IV: Conceptual Change

12 Why Conceptual Change Matters

13 Conceptual Change in the History of Science: Life, Mind, and Disease

14 Getting to Darwin: Obstacles to Accepting Evolution by Natural Selection. (with Scott Findlay)

15 Acupuncture, Incommensurability, and Conceptual Change (with Jing Zhu)

16 Conceptual Change in Medicine: Explanations of Mental Illness from Demons to Epigenetics (with Scott Findlay)

Part V. New Directions in the Cognitive Science of Science

17 Values in Science: Cognitive-Affective Maps

18 Scientific Concepts as Semantic Pointers