

BELIEFS ABOUT THINKING

Jonathan Baron*

Citation: Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education*, pp. 169–186. Hillsdale, NJ: Erlbaum.

Attempts to improve thinking and reasoning assume that people do not think as well as they might. The actual conduct of thinking does not measure up to an ideal standard. For example, I and others have argued that people are often biased toward views they already hold. They often fail to consider alternative views, counter-evidence, or goals that their favorite plans will subvert. In making this argument, I hold up (and defend) a standard of *active open-mindedness* as an ideal.

Yet, people have their own standards. They are capable of thinking of themselves as careful, fair-minded, thoughtless, biased, decisive, faithful to their beliefs, or wishy-washy. They also apply these standards to others, as when they judge their friends, co-workers, or political leaders.

I shall argue here (with some preliminary data) that part of the discrepancy between people's thinking and ideal standards is that people's own standards differ from the ideal. Thus, people who think poorly by ideal standards may reject those standards. They may think they are thinking well when they are actually thinking badly. This argument implies that the teaching of thinking may involve modification of people's standards. It is not just a matter of prodding people to live up to the standards they already hold.

My argument requires a specification of good thinking. If we are to claim that people are thinking badly and don't know it, we need a clear standard that we can oppose to theirs. I shall therefore begin with a summary of the theory of good thinking explained in Baron (1985, 1988). I shall then discuss its implications for the formation of standards about thinking. Following this, I shall present some preliminary evidence concerning judgments of thinking.

1 What is thinking?

Before introducing the sketch of the theory of good thinking, we need a general way of talking about thinking itself. By 'thinking,' I mean a conscious response to doubt or ignorance. (Baron, 1985, ch. 3, elaborates this account.) It is what

*This work is supported by a grant from the National Institute of Mental Health. John Sabini is a co-investigator and has been helpful in designing the studies reported here, as has Andrea Bloomgarten. (Both are in the Psychology Department of the University of Pennsylvania.) I also thank John for his comments on the paper.

we do when we are at a loss, at least for a moment, about what to do, what to believe, or what to adopt as a personal goal. For example, I read in the newspaper the suggestion that surrogate mothering be regulated by law. What do I think about that?

We may analyze all thinking into search and inference. We search for *possibilities*, *evidence*, and *goals*. These are the elements of thinking. They are conscious representations of actions or propositions.

Possibilities are possible answers to the question that inspired the thinking: prohibition, no law at all, or something in between. Possibilities (like goals and evidence) may be in mind before thinking begins, or they may be added as a result of search or suggestion from outside. In the surrogate-mother case, I might think that it should be outlawed, or that the law should stay totally out of it except for the enforcement of contracts.

Each possibility may be seen as having a *strength*, which represents the extent to which it is seen by the thinker as satisfying the criteria that constitute the goal. When the goal of thinking is to assess the appropriateness of a belief (e.g., that surrogate mothers want to keep the babies they bear) as a basis for action (e.g., outlawing the practice), we may speak of the strength of a possibility as the thinker's *degree of belief* in it, or, in some cases, its probability. In decision making, the strength of a possibility corresponds to its overall desirability of an act, taking into account all relevant goals. In such cases, we might sometimes imagine that a possibility is evaluated on several dimensions, each corresponding to a goal, and the overall strength of the possibility is some sort of combination of these separate evaluations. For example, the possibility that surrogates be outlawed is good for those who might change their mind but bad for those who would not and for those who employ them. These dimensional strengths can be used to guide the search for new possibilities, e.g., we might require more stringent consent procedures rather than a total ban.

Evidence is anything that can be used to decide among the possibilities: reports of actual cases (e.g., of surrogate mothers who were happy with their action or those who were not), imagined cases, moral principles and arguments (freedom of contracts, freedom from exploitation), etc. Evidence may be sought or made available. One possibility can serve as evidence against another, as when we challenge a scientific hypothesis by giving an alternative and incompatible explanation of the data.

Goals are the criteria used to weigh the evidence. How much do I care about the welfare of children, the freedom and sanctity of contracts, the feelings of biological mothers? Goals are not all given. I have to search for them and sometimes discover them just as I search for and discover evidence. The goal determines what evidence is sought and how it is used. For example, the goal of protecting the feelings of surrogate mothers leads to a search for evidence about those feelings. Goals and evidence together affect the strength of possibilities, but we do not speak of a piece of evidence being satisfied or reached, although we may say this of a goal.

One type of goal is a subgoal, a goal whose achievement will help us achieve our main goal. For example, the idea of obtaining informed consent might help

to protect surrogate mothers. As Duncker (1945) pointed out in a different context, a subgoal also a also partial possibility. The idea of informed consent is a partial solution, and it also sets up a new goal (how to obtain consent).

The use of evidence, in the light of the goals, to increase or decrease the strengths of the possibilities may be called *inference*. It is useful to think of each piece of evidence as having a *weight* with respect to a given possibility and goal. The weight determines how much it strengthens or weakens the possibility in the light of the goal. This weight depends on the thinker's knowledge and beliefs. A weight by itself does not determine how much the strength of a possibility is revised; rather, the thinker controls this revision. Thus a thinker may err by revising the strength of a possibility too much or too little.

Inference is only part of thinking. The rest is search. This is why logic is incomplete (at best) as a normative or prescriptive theory of thinking.¹

The relationship among the elements of thinking may be illustrated as follows:

G G

P E
E
P
E

* The evidence affects the strengths of the possibilities, but the weight of the evidence is affected by the goals. Different goals can even reverse the weight of a piece of evidence. For example, if I am trying to decide between two cars (possibilities), one of which is heavy (evidence), concern with safety (a goal) might make the size a virtue (positive weight) but concern with mileage (another goal) might make it a detriment (negative weight).

Why just these phases: search for possibilities, evidence, and goals, and inference? The idea is that thinking is a method of choosing among (or otherwise evaluating) potential possibilities, that is, possible beliefs, actions, or personal goals. For any choice, there must be a purpose or goal, and the goal is subject to change. I can search for (or be open to) new goals; hence, search for goals is always possible. There must also be elements that can be brought to bear on the choice among possibilities; hence, there must be evidence, and it can always be sought. Finally, the evidence must be used, or it might as well not have been gathered. These phases are necessary in this sense.

¹Another problem with logic as a standard is the practical problem of applying it. If we are charitable in granting the right unstated premises to a thinker (as is Henle, 1962), we may find almost any thinking to be logical. If we are not charitable, insisting that premises be stated or at least conscious, we might find good thinking to be nonexistent outside of the logic classroom. On the other hand, the fallacies studied by 'informal' logicians might be useful heuristics for detecting poor thinking.

2 Now, what is good thinking?

Baron (1985, chs. 1-4) has argued that good thinking involves *optimal search* for possibilities, evidence, and goals, and *fairness* in the search for evidence and in inference. These criteria are designed to maximize the expected desirability (utility) of the outcome of thinking in terms of the thinker's goals - not just the immediate goals but all the goals that are affected by the thinking in question. In other words, people who conform to these criteria will, on the average, do best at achieving their own goals.

Thinking goes wrong for three reasons. First, our search misses something that it should have discovered. For example, if I argue for surrogate motherhood on the basis of respect for contracts, I may ignore the argument that contracts made under various forms of duress or ignorance are already invalid. Second, we make see evidence and goals, and inferences, according to criteria irrelevant to our goals. For example, we may be unfair to some of the possibilities at hand. I may ignore evidence when it goes against a possibility I initially favor. The same favoritism for a possibility may cause me to cut off my search prematurely for alternatives to my first idea (e.g., regulation) or for reasons why it might be wrong. Third, we may think too much. Like any other activity, thinking has a cost, and after some amount of thinking, its cost exceeds its expected benefits (Baron et al., 1986).²

Poor thinking is usually characterized by too little search and - most importantly - by biases in favor of possibilities that are favored initially. By contrast, good thinking is actively open minded. It consists of search that is thorough in proportion to the importance of the question, and fairness to possibilities other than the one we initially favor. Good thinking is a virtue, like thrift, that is best practiced in moderation. We call it a virtue because most people do not have enough of it. There are more spendthrifts than penny-pinchers, and there are more people who are resistant to new possibilities, goals, and evidence than people who are open (as argued by Baron, 1985, ch. 3). This may be particularly true in the domain of citizenship, where the benefits of our thinking for others may have little effect on the amount of thinking we do.

These two principles - optimal search and fairness - are also the standards we apply in academic settings. When I read a student's paper, or a colleague's, or sometimes even my own, the things I look for are omissions of relevant evidence, omissions of statements about goals or purposes, and omissions of alternative possibilities, other answers to the question at issue. I also look for partiality to the thesis of the paper, partiality that may itself cause the omissions just mentioned. When students take these kinds of criticisms to heart and try to become more thorough and more impartial, they are becoming more intelligent thinkers. They are acquiring habits and values that will increase their effectiveness just as surely as would an improvement in their memory or their mental

²There are of course other reasons why thinking may go wrong, such as thinking that occurred prior to the episode in question, mental capacities, or opportunities for acquisition of relevant knowledge (Baron, 1985, ch. 5). The three reasons given refer to the conduct of thinking at the time.

speed.

People who follow these standards would be those who seek the truth, not those who feel their first intellectual obligation is to defend a certain belief, to make it be true despite the evidence. They would be those who try to make the best decisions for themselves and others, not those who want only to say, ‘I told you so.’ They will be people who want *to be* right, not those who want *to have been* right.

So far, I have not said how to implement these standards. That is part of the *prescriptive* theory that specifies how thinking should actually be conducted. The model sketched so far is *normative*. It specifies standards without saying how to achieve them in the real world.

2.1 What kind of theory is this?

I have argued (Baron, 1985, chs. 1-4) that a normative model of thinking may be justified in terms of expected-utility theory, a normative model of decision making. By this theory, the best we could do with the knowledge available to us at the time of a decision is to imagine all possible consequences of each possible action. We should determine the probability of each consequence and its desirability. If we could measure subjective probability and desirability, we could multiply them for each consequence and sum across consequences to arrive at a subjective expected desirability, or expected utility.

We may apply this model to the general description of thinking as a decision process, and thus arrive at a general normative model for thinking. We thus apply expected-utility theory to the question of whether we ought to search for an additional possibility, an additional goal, or an additional piece of evidence. We should carry out these searches when the expected utility of doing so is positive, all things considered. Finally, we may apply the same theory to the question of how evidence ought to be weighed. Specifically, we ought to weigh evidence in a way that is fair to all possibilities, that is, in a way that assigns strengths to beliefs in a way that is most helpful in making decisions that achieve our goals.³

Note that this is still a *normative* model⁴. In order to figure out whether it is worth looking for an additional possibility or not, we must step outside of the specific situation and carry out a full analysis.

A normative theory like this one is distinct from a *prescriptive* theory, which tell us actually how we ought to do something, because the former is idealized. (For further discussion of this and related distinctions, see: Baron, 1986; in press; and Hare, 1981.) When we take normative theories seriously as guides for action, they usually become self-defeating (in the sense of Parfit, 1984). For example, if we attempted to use expected-utility theory for deciding whether to continue thinking, we would spend so much time making that decision that we

³Note that this specification allows some self-deception. For example, if one has decided that knowing the time of one’s death will not affect one’s decisions and will only cause dread, one should ignore evidence for the possibility that one will die soon.

⁴There may have been a bit of ambiguity on this point in Baron (1985).

would on the whole achieve more of our goals if we followed some simple rule of thumb instead.

Nonetheless, we would want our decisions to conform to normative models if we could make them do so. Thus, the normative model of decision making may provide not only an idealized *method* of making decisions but also a *standard* by which to evaluate our actual thinking. If our rules of thumb are questioned, our ultimate standard for evaluating them is the extent to which they help us achieve our goals. Fortunately, expected-utility theory provides a measure of closeness to this end.⁵ We can therefore estimate the expected utility of different methods of decision making, including proposed prescriptive models and descriptive models of what people actually do without our advice, for a certain type of problem (as done by Johnson & Payne, 1985).

To justify a prescriptive model, we must argue that the model in question can bring people closer to this standard than they would be without it. We must argue that people depart from this standard in certain systematic ways and that the prescriptive model in question counteracts these departures or biases. The application of expected-utility to the conduct of thinking provides a way of arguing that people do not think as they ought and that they may be helped by thinking somewhat differently.⁶

More specifically, I have argued that people tend to search too little when thinking is important, and to be biased toward possibilities they already favor. The ‘too little’ and ‘biased’ refer to departures from the normative model of thinking. Anything that counteracts these departures will be prescriptively advisable.

Even if these particular claims are wrong, the framework I have sketched makes clear a major role of empirical psychology. This role is to determine just how people deviate from the normative model of thinking, and to assess the effectiveness of possible prescriptions to reduce the deviations.

The theory so far does not *dictate* prescriptions in any sense. Prescriptions are designs or inventions, the purpose of which is to achieve a certain goal, the goal of bringing our thinking more into line with this normative theory. The business of prescriptive theory is a process of invention, invention of our culture and ourselves.

In thinking and decision making, prescriptions may concern three general objects: rules of action (including informal rules of thumb), personal goals, and beliefs. These correspond to the three elements of any decision, with rules of action analogous to options or possibilities. The decision in this case is how to conduct one’s thinking. *Rules of action* include such things as heuristics and more formal methods. Polya’s heuristics are a good example of these. They are things to do when one is stuck. Other heuristics may be generally useful

⁵This is not necessarily true of any normative model. For example, formal logic provides a right answer but not a measure of degree of departure from the right answer.

⁶It might be possible to apply utility theory to heuristics of thinking directly, rather than applying it first to the general framework of search and inference. However, the search-inference framework captures all the relevant aspects of the conduct of thinking itself, so it provides a reliable guide for evaluation of the effect of heuristics on goals achievement.

in counteracting natural biases: thinking of alternative possibilities, looking for evidence against an initial idea, asking about goals. Other heuristics may be specified, and conditions may be stated for when each heuristic is most useful (e.g., when the issue is important), but these are not my concern here.

Personal goals (as elements of prescriptions) may be seen as parts of one's life plan (Baron, 1985, ch. 2). They may be adopted by a process of decision making, but they are long-term decisions that constitute the goals for other decisions. Again, some personal goals may be more conducive than others to thinking in conformity with the normative model. For example, the goal of making a certain belief true is not only impossible to achieve but also contrary to good thinking about the belief. 'Faith' is a word that defends a self-defeating goal. Similarly, the goal of being a perfect decision-maker on the first try will prevent one from revising one's decisions in the light of good argument.

I use the term *beliefs* here in a narrow sense, which refers to beliefs that form the basis of personal goals. Like other choices, personal goals are based on beliefs about how other goals are best achieved. Religious beliefs are a good example of beliefs in this sense. A person who holds these beliefs has adopted personal goals that would not have been adopted if the beliefs were in sufficient question. Many things that we call 'values' are beliefs in this sense, although some are more properly thought of as goals. The standards for good thinking that are the main topic of this paper are also beliefs of this sort. People who believe that a certain way of thinking is good (for something) will then establish (to varying degrees) personal goals of thinking in that way.

3 Prescriptions for beliefs about thinking

I want to focus now on prescriptions concerning beliefs. I do not mean to imply that heuristics and goals are unimportant. Indeed, I have suggested that these three kinds of objects are closely interrelated (Baron, 1985, ch. 7). The reason for being particularly concerned with beliefs is that efforts to influence them might be quite effective in making people better thinkers. Most of our educational system is set up to impart beliefs, facts, and habits, but not goals. Of these three, beliefs may be the most general in their effect on thinking.

The idea of a prescriptive theory of beliefs about thinking may be more plausible through an analogy with other belief systems. In many other domains, it has been suggested or shown that people hold naive theories, which must yield to more mature theories either through drastic reorganization or gradual change (e.g., Vosniadou & Brewer, 1987).

The closest parallel to the distinction between naive and sophisticated theories in the domain of beliefs is the developmental theory of Kitchener and King (1981; King, Kitchener, Davison, Parker, & Wood, 1983). Their analysis does not concern subjects' beliefs about thinking itself but rather about justification of beliefs and the nature of knowledge. They have proposed a developmental sequence (similar to those of Kohlberg, 1971, and Perry, 1971) in subjects' assumptions about reality, knowledge, and justification, as expressed in interviews

about belief dilemmas.

There are seven stages. Subjects move through the stages as they get older, very likely in part as a result of education. The first four stages involve a gradual break with the idea that truth is absolute and known to all. The breakdown of this idea is what leads to subjectivism or relativism. At Stage 4, one person's belief about anything is as good as another's. No evaluation is possible. The top three stages involve a gradual recognition of the possibility of general standards of justification such as those advocated here.

The important point about this theory, and the evidence supporting it, is that it points out the existence of naive beliefs about beliefs themselves. Although the sequence does not concern thinking, decision-making, and belief-formation themselves, it is clear that the beliefs tapped by Kitchener & King are relevant to thinking. The higher stages provide reasons for the value of thinking, where the lower stages do not.

There are many beliefs that make for good thinking in general - i.e., those that counteract the biases I have sketched - and many beliefs that tend to encourage these biases. Among the former are the belief that thinking often leads to better results, that difficulties can often be overcome through thinking (rather than, say, through luck), that good thinkers are open to new possibilities and to evidence against possibilities they favor, and that there is nothing wrong (per se) with being undecided or uncertain for a while. Among the latter are beliefs that changing one's mind is a sign of weakness, that being open to alternatives leads to confusion and despair, that quick decision-making is a sign of wisdom or expertise, that truth is determined by authority, that we cannot influence what happens to us by trying to understand things and weigh them, and that use of intuition alone is the best way to make decisions. The former beliefs act to oppose the natural biases I have described, and the latter act to support them (whatever germ of truth they might otherwise contain).

The importance of beliefs in thinking is consistent with the suggestion of Perkins et al. (1983) that poor thinking often results from a 'makes-sense epistemology,' in which the thinker believes that the way to evaluate conclusions is by asking whether they 'make sense' at first blush. It is also consistent with the claims of Dweck and Elliott (1983) that children's beliefs about the nature of intelligence influence their response to failure in problem solving. Both the belief that error is due to stupidity and the belief that success is due to effort may become self fulfilling in the long run. (See also Kreitler and Kreitler, 1976 and elsewhere, concerning the role of beliefs in general.)

If people do not believe that thinking is useful they will not think. This is perhaps the major argument one hears against thinking about things like nuclear war, religion, or morals: 'These matters are beyond me. They are best left to experts who are capable of thinking about them - if anyone.'

4 How people judge thinking

We may study people's beliefs about thinking by giving them examples of thinking - in the form of thinking-aloud protocols supposedly generated by others, for example - and asking our subjects to evaluate the thinking.

This is not only a useful research tool but also a potential educational tool. One problem in the teaching of thinking has been the measurement of success of such instruction. It is difficult to test thinking directly, because its success or failure depends on so much besides the quality of the thinking. However, the judgment of others' thinking is fairly easy to test; it can even be done objectively, in multiple-choice format. Instruction in thinking ought to improve performance on such a test. Even if students merely learn the standards without internalizing them, this may be sufficient success. Arguably, the goals of instruction in thinking should not take the form of indoctrination but rather of simply placing before the student a set of standards that the student may accept or reject. A fair test is one that insures that the student has *understood* the standards, so that they may be applied correctly to new instances of thinking. Whether students then goes on to apply them to their own thinking may be a matter for them to decide (although, for reasons I shall discuss, we ought to hope that they do this).

As an illustration of the possibility of assessing students' judgment of thinking, I want to report some preliminary results from research conducted in collaboration with John Sabini and Andrea Bloomgarten. Initially, we simply wanted to measure individual differences in students' judgment of the thinking of others and in the beliefs that supported these judgments. We examined the effects of three variables: whether evidence was one- or two-sided, the strength of the conclusion drawn (which may or may not agree with the total evidence presented), and the subjects' own beliefs (which may lead them to evaluate statements more favorably when they agree with the conclusion). We did not in these studies address other issues, such as the thoroughness of evidence search in general (aside from its two-sidedness) or the relevance of the evidence to the issue. These remain for future study.

In a first study, to assess subjects' judgment of the thinking of others, we presented the following task to 96 undergraduates at the University of Pennsylvania.

Instructions. Imagine that each of the following selections is a record of a college student's thinking while answering a questionnaire ... Give each selection a grade for the *quality of thinking* it represents: A, B, C, D, or F. You may use plusses and minuses. Briefly explain your reasons for giving different grades (if any) to different selections in the same group (1-8).

Item A: Automobile insurance companies should charge more for city dwellers than for suburbanites.

1. My first thought is that each group of people should pay for its own accidents. City dwellers surely have more accidents, and their cars get broken into and stolen a lot more. I'll say 'strongly agree.'

2. My first thought is that each group of people should pay for its own accidents. City dwellers surely have more accidents, and their cars get broken into and stolen a lot more. I'll say 'slightly agree.'
3. My first thought is that each group of people should pay for its own accidents. City dwellers surely have more accidents, and their cars get broken into and stolen a lot more. On the other hand, it doesn't seem fair to make people pay for things they can't help, and a lot of people can't help where they live. I'll say 'slightly agree.'
4. My first thought is that each group of people should pay for its own accidents. City dwellers surely have more accidents, and their cars get broken into and stolen a lot more. On the other hand, it doesn't seem fair to make people pay for things they can't help, and a lot of people can't help where they live. I'll say 'strongly agree.'

Selections 5-8 were analogous in form, but on the opposite side of the issue. The other items, each with eight analogous selections, were: 'Social Security benefits should be taxed,' and, 'The nations of the world need to make special efforts to reduce the growth of population.'

Each group of eight responses was divided into two groups of four, one group on each side of the issue. Within each group of four, the first two considered arguments (evidence) on one side, and the last two considered arguments on both sides. Within each of these pair, the items differed in the strength of the opinion expressed. A more moderate opinion would seem more appropriate when both sides had been considered.

The grades were converted to a numerical scale, from 12 for A+ to 0 for F. For the four responses in each group, the mean grades were 6.89 (s.d.=1.83), 6.38 (1.85), 8.31 (1.95), and 4.77 (2.56), respectively (i.e., B-, C+, B, and C). The best grade was given to the thinking that considered both sides and reached a moderate conclusion (e.g., #3 above). The worst grade was given to the thinking that considered both sides and reached a strong conclusion (e.g., #4). In general, then, the consistency of the conclusion with the arguments presented was more important to these subjects than whether or not both sides were considered.

However, there were substantial individual differences. We noted the following different types of justifications:

Content (given by 47% of the 89 subjects whose answers could be scored). These justifications pointed to the substance of the argument presented, often arguing back or pointing out counterarguments, e.g., 'Some people have no choice but to live in the city ...' The implicit assumption behind these justifications was that the *correctness* of the evidence and conclusions brought forward was a reliable index to the quality of thinking. Correctness, of course, was determined by consistency with the judges own beliefs. For example, in the item on taxing social security benefits many subjects said (falsely) that rich people do not collect social security benefits, so

this was not an issue. A couple of subjects used nothing but this type of justification. One of them wrote several pages taking issue with every detail. (In later studies, conclusions were expressed as either ‘agree’ or ‘disagree’, thus removing the chance to assess agreement between the *strength* of the conclusion and the arguments. In those studies a much higher proportion of subjects gave justifications in terms of content alone.)

Weight (45%). These were based on the consistency or inconsistency of the conclusion with the arguments thought of. The judgment was based on the strength or goodness of the arguments according to the subject’s own judgment, e.g., ‘Strongly disagreeing isn’t fair ...’ followed by a substantive argument. I take weight judgments to be normatively correct, and very likely prescriptively correct as well. Some arguments are indeed better than others, and it would be inappropriate to ignore this fact. However, justifications of this sort are open to bias. A judgment of the consistency of an argument and a conclusion may be affected by the judge’s agreement or disagreement with the conclusion.

Logic (61%). These pointed to the consistency of the conclusion with the arguments presented on *formal* grounds. There were three forms of such arguments. In the most common form, subjects felt that bringing up both sides of an issue should lead to a weak conclusion. Hence, the strong conclusion in case 4 (and case 8) was not justified, simply because the thinker had seen both sides. In another form, subjects felt that one-sided arguments justified strong conclusions, so case 2 was given a low grade. In a third form, however, judges felt that one-sided arguments were consistent with *weak* conclusions. These subjects suggested that a person who thought of only one side could not know how strong the argument would be on the other side, so they should not be so confident in their own side. Hence, case 1 was given a low grade. The use of these kinds of justifications may be seen as a prescriptive rule, which serves the function of avoiding the kind of bias that weight judgments are prone to. Rather than attempting to assess the true weight, one avoids the issue by simply counting pro’s and con’s, making judgments on formal grounds alone.

One-sided (8%). A judgment was positive for one-sided arguments or negative for two-sided arguments on the basis of form. That is, one-sidedness was seen as a virtue. For example, a justification for grades of D and F for two-sided answers was, ‘These don’t even make sense. They can’t make up their minds.’

Two-sided (52%). A judgment was positive for two-sided arguments or negative for one-sided arguments on the basis of form. That is, two-sidedness was seen as a virtue.

Strength (13%). Strength of the conclusion is itself seen as a virtue, e.g., ‘conviction.’

Moderation (7%). Moderation is itself seen as a virtue.

Conceptually, it may be reasonable to speak of a continuum here. At one end are purely formal judgments, e.g., moderate conclusions are appropriate to two-sided arguments and extreme conclusions are appropriate to one-sided arguments, regardless of content. Many subjects adopted this approach, giving identical grades to arguments on both sides throughout. At the other extreme is an attempt to make a judgment of the appropriate weight of each argument. The danger of this approach is that subjects will impose their own beliefs on those they judge, so that inferences are judged as good ones if the conclusions agree with their’s. This may be analogous to a kind of hindsight effect (Fischhoff, 1975), and is worthy of study in its own right. Normatively, to judge the thinking of another, we must try to put ourselves in that person’s position. This may be difficult to do, however, and the formal approach may be a good prescriptive device to avoid the bias of judging thinking by its conclusions.

The results suggest that, although many people believe that consideration of opposing arguments is a manifestation of good thinking, many other people do not notice such two-sidedness, and at least a few others find it bothersome. Many people also evaluate thinking according to its conclusions. This effect is well known in the study of logical reasoning (e.g., Morgan & Morton, 1944), but its effects may be more insidious in everyday reasoning, where the weight of arguments depends on subjective judgment. It may be a major means by which people resist evidence against views they favor. If people think that the criteria of good thinking allow them to judge evidence in this way, they will think in a biased way without knowing that they are being biased.

Aside from content, judgments were based primarily on the consistency of the conclusion and the arguments thought of. It seems as if those subjects who do attend to the form of the thinking (as opposed to its content) base their judgments predominantly on the consistency of the conclusion and the evidence *that the thinkers think of*. Roughly, this is the ideology of the logician. What matters is consistency, and less attention is paid to whether or not one is optimally thorough and fair in the search for the evidence one uses.

One difficulty with the present study was that the presence of ‘strongly agree’ versus ‘slightly agree’ (and ‘disagree’) answers may have focused subjects’ attention on the consistency of the strength of the answers with the arguments presented. In subsequent studies, we attempted to eliminate this problem by removing degrees from the conclusions drawn.

One study was done on my undergraduate class. They were given the following moral dilemma and were asked to evaluate the thinking exhibited in a series of responses to it on the assumption that each response was given by another student in a thinking-aloud task:

‘Professor Smith teaches a class with 50 students. After final grades are posted and summer vacation has begun, Jones, a student in the class, finds that he has just missed getting the B he needs in order to keep his scholarship, which

he needs quite badly. (There are 5 other students in the class who came just as close to getting a B, and they have left.) He asks Smith whether he could rewrite his paper (on which he got a C) so as to raise his grade. Should Smith let Jones rewrite the paper or not? Explain.'

Eight of the responses were two sided, e.g.:

No. On the one hand, it will help Jones, and it will set a precedent for other humane acts by Smith and anyone else who hears about it. However, it would also set a precedent for breaking other rules that people expected would be followed. Also, the number of scholarships is limited, and if Jones gets one, this will deprive someone else who is probably more deserving. These factors outweigh the others.

Yes. On the one hand, it would set a precedent for breaking other rules that people expected would be followed. Also, the number of scholarships is limited, and if Jones gets one, this will deprive someone else who is probably more deserving. However, it will help Jones, and it will also set a precedent for other humane acts by Smith and anyone else who hears about it. These factors outweigh the others.

* Nine responses were one-sided. These stated one or two of the arguments used in the two-sided arguments. The students assigned grades on a scale from A to F, as in the last study. To derive an overall score for the value accorded to two-sidedness, the (total, converted) grades assigned to one-sided arguments were subtracted from the grades assigned the two-sided arguments.

The same class was given the following dilemma in an earlier assignment. They were told to think about it and transcribe their thoughts as literally as possible as they occurred.

'It is suspected that great mineral wealth will be discovered at the sea bottom in the next few decades and that some countries will be in a technological position to mine it. The oceans are now property of no nation, and their bottoms have never before been contested. Imagine that you are attending a conference to discuss: how this wealth should be allocated among nations; how to motivate people to make the required (major) investment to begin the mining; and how future decisions (such as modifications of the scheme) should be made. What kind of arrangements do you think would be best?'

Some protocols showed actively open-minded thinking, e.g.:

Wealth must be divided among nations fairly. What does 'fairly' mean? Should allocation be based on the *size* of the country? Some nations are significantly larger than others. But some countries have more people per unit area. Should allocation be based on overall population size? It would be *very* difficult to get all nations concerned to agree their shares were fair. Wait, the United Nations has a certain number of representatives from each country. They would be the ideal group to handle this. Total wealth should be divided

by overall number of representatives, then allocated according to number of representatives per country. *But* some nations would be better able to *use* the mineral wealth. These would be nations with greater technology. Therefore, underdeveloped nations would be unable to benefit as well as nations that are financially more secure. That would be unfair. (This goes on for a couple of pages.)

* Others showed no evidence of criticism of an initial idea (other than working out its details) nor of consideration of other possibilities, e.g.:

I believe that the most logical way of allocating the mineral wealth beneath the ocean is to allocate the ocean floors by extending national borders outward along the ocean floors. In effect, this plan would treat the ocean floor in the same way as exposed land surfaces. The water above the floor should still remain international territory, except where it is already considered national property. ... Establishing boundaries in this manner is fairly simple, but it will favor nations with long coastlines. along large bodies of water, but is no less fair than the rules for establishing national air space. (This goes on as well.)

* Answers were classified into these categories, with questionable answers omitted. (The classification was blind with respect to the other data.) The two groups of subjects differed significantly ($t_{82}=1.97$, $p=.02$) in the difference between the grades for two-sided responses and the grades for one-sided responses. The 71 open-minded subjects gave mean grades of 7.3 (B-) and 3.7 (C-) to two-sided and one-sided arguments, respectively. The 13 'closed-minded' subjects gave mean grades of 6.3 (C+) and 4.1 (C-), respectively. Thus, there does seem to be a relationship between the standards used to judge the thinking of others and the kind of thinking one does on one's own.

A third study of individual differences examined 40 subjects' responses to each of four moral dilemmas, like (and including) the one about Professor Smith. In addition to answering each dilemma, subjects were asked to, 'briefly list all the relevant considerations and principles.' Their answers were scored (blindly) for whether they mentioned arguments on both sides or not. The proportion of scorable answers that were two sided correlated with a measure of *belief* in two-sidedness consisting of the difference between grades assigned to two-sided and grades assigned to one-sided answers to the Smith dilemma: $r=0.26$, $p=.05$. (This measure was like that used in the last study except that the answers used were matched so that they contained identical arguments. Thus, one-sided arguments contained *two* arguments on the same side, which were recombined for the two-sided arguments. Perhaps because answers were matched in complexity, the mean difference was not significantly different from zero; subjects did not on the whole consider two-sided arguments to be better than equally thorough one-sided arguments.)

As an additional measure of subjects' beliefs about thinking, we included three scenarios involving belief formation, such as, 'Judy had to decide which

of two candidates to vote for in a primary election. She initially favored one of them, but wondered whether this was right. What should she do? Why?' The other two scenarios dealt with a decision about which of two friends to hurt and a decision about which bank to use. Written responses to these scenarios were scored (blindly to other information) with respect to whether further thinking was mentioned or not. For example, some subjects said that Judy should gather more information, while others said that she should go with her initial feeling. The number of scenarios in which additional thinking was recommended correlated with use of two-sided thinking in the dilemmas themselves: $r=0.28$, $p<.05$. (A third measure of belief in two-sided thinking, based on a questionnaire about the scenarios, did not correlate significantly with anything except the scenario measure just described, although all correlations were in the predicted direction.)

There seem to be competing beliefs about thinking in our culture. Where do these beliefs come from? Why doesn't everyone think that two-sided thinking is better than one-sided thinking? It would be easy to argue that beliefs in one-sided thinking are the result of the evolution of institutions, such as organized religions and nations. To survive, that is, to keep its adherents from one generation to the next, each of these institutions must convince its adherents that its views are correct even though many outsiders will argue otherwise. Those institutions that inculcate an ideology in which defense of one's belief is a virtue and questioning is a vice are the ones most likely to overcome challenges from outside. By this argument, enough of these institutions still survive as to have a substantial influence on our culture.

There may be some truth in this, but I think there may be another answer to the question, which will be the subject of future research. It is possible that people are simply confused about two different standards concerning thinking, which we might call the 'good thinker' (the standard I have been advocating) and the 'expert.' In many ways, experts appear to be the opposite of good thinkers. Because they *know* the answer to most questions, they do not have to think very often, compared to novices. Thus, when a news commentator criticizes a political candidate for waffling and being unsure (as might befit a good thinker faced with many of the issues that politicians must face), the implication is that the candidate is not expert enough to have figured out the right answers yet. Similarly, a person who adopts a 'know it all' tone of voice - speaking without qualification or doubt - is giving a sign of expertise in the matter at hand. Some parents (perhaps because they *are* experts about the matter under discussion) may talk this way to their children, who come to think of it as a 'grown up' way to talk.

This confusion of expertise and good thinking may reinforce the institutional pressures mentioned earlier (if they exist). Those who are considered wise and respected members of the institution may talk like experts, encouraging their followers to know rather than to think. And how are they supposed to know? Without thinking, there is only one way: listen to the experts.

Although expertise and good thinking both contribute to success in achieving goals (Baron, 1985, ch. 5), they are not the same thing. We must not assume

(or allow others to assume) that an understanding of expertise will solve the problem of providing standards for thinking.

5 Conclusion

The view of thinking I have presented may be seen in an interpersonal, moral context (Baron, 1985, ch. 6). The standards of thinking are analogous to other standards of interpersonal conduct, such as those of business competition (or even international relations). In all these cases, we may adopt an aggressive, uncooperative stance, or a cooperative one. In the domain of thinking, the aggressive stance is the belief that one should defend one's own beliefs. The cooperative stance is the belief that one should be open to the arguments of others, evaluating them on the basis of their form rather than their conclusions, and letting them influence one's own beliefs to the extent that they are good arguments.

If most people took the cooperative stance (and if enough people could distinguish good arguments from poor ones), then the best arguments would usually prevail (because they would be fairly considered by all), and we would all benefit (Baron, 1985, ch. 6). The aggressive stance has no comparable justification. One might argue that one must defend one's beliefs, for if one does not, truth might not prevail. However, those who think this way cannot condemn their opponents for thinking the same way. If nobody were open to persuasion, we might as well not talk at all. Those who take advantage of the openness of others without being open themselves are free riders, like those who watch public TV without contributing. In sum, good thinking as I have defined it is not just good for those who do it. It is good for us all.

I have argued that the way we carry out our thinking is influenced by our beliefs about how we ought to think. Few of us think we are thinking badly, especially on matters of morals or public affairs. If I am right, this is an optimistic conclusion. We are in a position to improve our mutual thinking through influence on beliefs about thinking. This is promising because arguments may be brought to bear - arguments of the sort I have sketched here - to discuss, determine, and persuade people of the right kind of thinking. It is more difficult to change desires. It is also convenient for the practice of education, for it is easy to measure beliefs, and easier still to measure the extent to which certain standards are understood. I hope that the line of work sketched here will ultimately lead to the development of methods for assessing beliefs about thinking, and therefore to the improvement of thinking itself.

References

Baron, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.

- Baron, J. (1986). Tradeoffs among reasons for action. *Journal for the Theory of Social Behavior*, 16, 173-195.
- Baron, J. (1988). *Thinking and deciding*. New York: Cambridge University Press.
- Baron, J. (in press). Utility, exchange, and commensurability. (Invited paper.) *Journal of Thought*.
- Baron, J., Badgio, P., & Gaskins, I.W. (1986). Cognitive style and its improvement: A normative approach. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, Vol. 3. Hillsdale, NJ: Erlbaum.
- Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58 (Whole No. 270).
- Dweck, C.S., & Elliott, E.S. (1983). Achievement motivation. In P.H. Mussen (Ed.), *Carmichael's manual of child psychology*, Vol. 2. New York: Wiley.
- Fischhoff, B. (1975). Hindsight \neq foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Hare, R.M. (1981). *Moral thinking: its levels, method and point*. Oxford: Clarendon Press.
- Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, 69, 366-378.
- Johnson, E.J., & Payne, J.W. (1985). Effort and accuracy in choice. *Management Science*, 31, 395-414.
- King, P.M., Kitchener, K.S., Davison, M.L., Parker, C.A., & Wood, P.K. (1983). The justification of beliefs in young adults: a longitudinal study. *Human development*, 26, 106-116.
- Kitchener, K.S., & King, P.M. (1981). Reflective judgment: concepts of justification and their relationship to age and education. *Journal of applied developmental psychology*, 2, 89-116.
- Kohlberg, L. (1971). Stages of moral development as a basis for moral education. In C. Beck & E. Sullivan (Eds.), *Moral education*, pp. 23-92. University of Toronto Press.
- Kreitler, H., & Kreitler, S. (1976). *Cognitive orientation and behavior*. New York: Springer.
- Morgan, J.J.B., & Morton, J.T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, 20, 39-59.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.
- Perkins, D.N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The expanding frontier*, pp. 177-189. Philadelphia: Franklin Institute.
- Perry, W.G., Jr. (1971). *Forms of intellectual and ethical development in the college years: a scheme*. New York: Holt, Rinehart & Winston.
- Vosniadou, S., & Brewer, W.F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research*, 57, 51-67.