Chapter 5

Confirmation bias – Myside bias

Hugo Mercier

Published in Pohl (Ed) *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory*

The confirmation bias—"the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" (Nickerson, 1998, p. 175)—has been described by contemporary researchers as "ubiquitous" (Nickerson, 1998), "perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning" (Evans, 1989, p. 41). Earlier scholars had similar insights. Here's for instance Francis Bacon:

> The human understanding when it has once adopted an opinion … draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects; in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate. (Bacon, Novum Organum, Book 1, Aphorism 46, 1620)

Indeed, the existence of the confirmation bias might seem obvious to anyone who has talked politics with someone from the other side of the spectrum. As the journalist Jon Ronson put it "ever since I learnt about confirmation bias I've started seeing it everywhere."[i]

As it turns out, this quip raises interesting questions: Why wouldn't the researchers who have claimed to prove the existence of the confirmation bias also be biased? Couldn't their experiments be designed in such ways that they are more likely to make participants look biased? Couldn't these scholars have mistaken rational behavior for proof of a confirmation bias?

In this chapter I will argue that most of the conventional wisdom about the confirmation bias is wrong—starting with its name. As we'll see, there is no such thing as a general tendency to confirm whatever one thinks about, only a tendency to find arguments that support one's own views—a *myside bias* (see Perkins, 1989). Moreover, this bias is not as prevalent as is often thought: The best-known experiments that purport to demonstrate the wide scope of the confirmation bias are flawed. Finally, I will suggest that the myside bias is not an undesirable bug that we must strive to get rid off, but an adaptive feature that we can use to our advantage.

DEMONSTRATIONS OF MYSIDE BIAS

Hypothesis testing

The main source of evidence for the confirmation bias used to be two unrelated series of experiments on hypothesis testing. These experiments study what people do when they want to know whether one of their ideas is true or not. Both sets of experiments played a large role in shaping beliefs about the confirmation bias. It is thus worth going in some details over their logic and their results.

In the nineteen fifties, the English psychologist Peter Wason wanted to test whether people were able to test hypotheses the way scientists were supposed to do it, that is, through falsification—a good hypothesis being one that

can withstand repeated attempts to show it is false. To this end, he developed a simple task that emulates the scientific process: the rule discovery task (Wason, 1960). As the name indicates, the goal of the participants is to discover a rule. The rule is the equivalent of a law of nature, except that in this case it is the experimenter who determined what the correct rule is (the arrangement of numbers in a triple). The experimenter starts by providing participants with a triple that conforms to the rule: 2, 4, 6 (which gave the task its moniker: the '2, 4, 6'). Participants can then form a hypothesis about the rule—for instance "series of number to which 2 is added at each step." To test this hypothesis they provide a triple, and the experimenter tells them whether it conforms to the rule or not. This is the equivalent of a scientist conducting an experiment to test her hypothesis. Participants can test as many triples as they like. When they think they know what the rule is, they lay out their hypothesis, and the experimenter tells them whether it is correct or not. If the participants got it wrong, they can test more triples to evaluate another hypothesis until they get it right or give up.

In most cases, participants tested triples that were compatible with the hypothesis they had in mind. For instance, if a participant had made the hypothesis "series of number to which 2 is added at each step," she would be more likely to test it with "10, 12, 14" than with, say, "2, 4, 7." The difference between these two tests is that if the experimenter says that "10, 12, 14" conforms to the actual rule, the hypothesis is confirmed, whereas if he says that "2, 4, 7" conforms to the rule, the hypothesis is falsified. Because of this, Wason described the participants as having a tendency to confirm their hypotheses. Moreover, he found that this tendency either slowed down the participants'

progression towards the correct rule, or precluded them from reaching it altogether.

It was soon pointed out, however, that Wason's reasoning was faulty (see Poletiek, 1996). In fact both types of triples—those that conform to the hypothesis and those that don't—can either confirm the hypothesis or falsify it. If "10, 12, 14" does not conform to the actual rule, then the participant's hypothesis is falsified. And if "2, 4, 7" does not conform, then the hypothesis is (weakly) confirmed. To tell whether a participant really wants to confirm her hypothesis, one would have to know how she expects her test to turn out. All we can infer, when she generates a triple that conforms to the hypothesis she has in mind, is that she engages in what has been called a 'positive test strategy.'

Whatever the participants' strategy is called, there seems to be a problem with it, since it hinders the discovery of the correct answer. As it turns out, positive test strategy is in fact the most rational course of action in the vast majority of the cases (Klayman & Ha, 1987). If it misfires in the case of Wason's rule discovery task, it is because the triple provided by the experimenter is deceptive: It leads participants to think of a very specific hypothesis—such as "series of number to which 2 is added at each step"—when in fact the correct rule is very general—"two ascending numbers." An analogy can explain why this is misleading, and why the positive test strategy is usually efficient.

Imagine that you suspect a mountain to hold a gold deposit. You have an idea of where the gold might be—in the north flank, say—and you can choose where to prospect first. The positive test strategy is to dig in the north flank. Using this strategy, not only do you find the gold if your hypothesis was correct, but you are also likely to quickly find out if your hypothesis is wrong. If the gold

is on the south flank (i.e. your hypothesis does not overlap with the truth), you'll find out immediately that your hypothesis is wrong. If the gold is only on a specific ridge of the north flank (i.e. your hypothesis is too broad), you're also likely to find out quickly (unless you get lucky and then never prospect in other places you believe gold to be, which is unlikely). It's only if there happens to be gold not only in the north flank, but also somewhere else (i.e. your hypothesis was not broad enough), that the positive test strategy might mislead you. Such a situation, however, should be rare, since people are more likely to start from a hypothesis that is overly general rather than one that is overly specific.

Thus, Wason's discovery rule task does not show that participants have a confirmation bias. Instead, participants rely on a rational strategy in an environment designed to trick them. The same conclusion applies to the other standard demonstration of confirmation bias in hypothesis testing.

This second paradigm, developed by Mark Snyder and his colleagues, aimed at understanding how people conduct interviews—more specifically, whether interviewers suffer from a confirmation bias in selecting the questions they ask interviewees. Participants were put in the position of interviewers, and they were given a hypothesis about an interviewee: That she was either introverted or extraverted, depending on the condition the participants were in (Snyder & Swann, 1978). The interviewers had to pick from a list some questions to ask the interviewee. The relevant questions on the list were designed so that they would be likely to confirm one of the hypotheses. For instance, when asked "In what situations do you wish you could be more outgoing?" most people, including most extroverts, should provide an answer: Only an extreme extrovert would be unable to find a situation in which they do not wish they were more

outgoing. As a result, these questions are not diagnostic when it comes to differentiating mild introverts from mild extroverts, since both of these groups are likely to answer them positively. The issue was that the interviewers tended to select questions that were likely to confirm their hypothesis. For instance, interviewers who had been given the hypothesis that the interviewee was introverted were more likely to ask questions such as "In what situations do you wish you could be more outgoing?" They would then receive answers suggesting that there are situations in which the interviewee wishes she were more outgoing, leading the interviewer to infer that the interviewee is indeed an introvert, when in fact she could also have been a mild extrovert.

This experiment suffers from flaws similar to those of Wason's rule discovery task. First, even though the questions might be more likely to confirm the participants' hypothesis, they still have the potential to falsify it. For instance someone who is extremely outgoing could simply answer "None" to "In what situations do you wish you could be more outgoing?"—with such an answer, the hypothesis that the interviewee is an introvert would be quite falsified. If we do not know what answers participants expect to get, we cannot say that they are bent on confirming their hypothesis. Second, the participants' strategy might be quite rational: The experimenter might have led them to form the hypothesis that the interviewee was either an extreme introvert or an extreme extrovert, in which case the questions they were asking were quite appropriate. An experiment published a few years later showed that in fact participants are quite good at selecting diagnostic questions, and that they do not show much trace of a confirmation bias (Trope & Bassok, 1983).

The experiment Yaacov Trope and Miriam Bassok (1983) conducted was similar to that of Snyder and Swann. They asked participants (interviewers) to select questions so they could tell whether someone (the interviewee) was an introvert or an extrovert. But they made a couple of changes. First, they varied how diagnostic the questions were. Some questions were highly diagnostic ("Are you usually the initiator in forming new relationships?"). These questions could discriminate between mild introverts (who would typically say no) and mild extroverts (who would typically say yes). Other questions were less diagnostic ("After a test, do you compare your answers to those of other students?"). For these questions, only individuals from one extreme of the spectrum (here, extreme introverts) would be likely to answer differently from everyone else, while mild introverts and mild extroverts would give similar answers. The second change introduced by Trope and Bassok was that interviewers believed that the interviewee was either an extreme introvert, an extreme extrovert, a mild introvert, or a mild extrovert.

If the participants were keen on confirming their hypothesis, they could easily do so. They just had to pick questions that had low diagnosticity. This is not what the participants did: In every case they preferred to pick highly diagnostic questions. Moreover, when the participants picked the low diagnosticity questions, it was often appropriately, in order to test an extreme hypothesis. For instance, asking participants who they thought might be extreme introverts "After a test, do you compare your answers to those of other students?" Such a question, which does not discriminate between mild introverts and mild extroverts, can discriminate extreme introverts, the only ones liable to answer "No." This strongly suggests that the participants in the original Snyder

and Swann experiment only picked low diagnosticity questions because they had no alternative high diagnosticity questions and they had formed extreme hypotheses about the interviewees.

Two of the most important paradigms used to support the existence of the confirmation bias turn out to only reflect rational behavior. I now turn to another task—also designed by Peter Wason, and even more famous—which paints a more complex picture.

--- insert Text box 1 about here ---

Wason selection task

The Wason selection task—or four cards selection task—is the most studied task in the psychology of reasoning (Wason, 1966) (see Text box 1). In the standard version of the task, most participants answer either that the A card, or the A and the 4 cards should be turned over. The correct answer, however, is to turn over the A and the 7. The 4 card is not necessary, since the rule doesn't say what should be on the other side of even numbers; by contrast, the 7 is necessary, since it can falsify the rule—if a vowel is revealed on the other side, the rule is false.

At first, it was thought that the preference for the 4 over the 7 reflected a confirmation bias, a failure to pick a card that might falsify the rule. This interpretation, however, was promptly belied by two observations. First, the A card, picked by most participants, can falsify the rule, so that its choice cannot be explained by a conformation bias. Second, it's enough to add a negation to the consequent in the rule (i.e. "If there is a vowel on one side, then there is *not* an even number on the other side) for participants to provide the correct answer—in this case A and 4 (Evans & Lynch, 1973). Does the addition of a negation allow

participants to shed their confirmation bias? No. Whether there is a negation or not, participants tackle the task in exactly the same way. When they read the rule, the participants' attention is directed towards some cards by mechanisms of pragmatic inference (Sperber, Cara, & Girotto, 1995). These mechanisms allow us to understand utterances. They enrich the literal meaning so that it makes sense in the current context (for instance going from "Can you pass me the salt?" to "Please pass me the salt"). In the present case, the cards that are made most relevant by the rule are simply those that the rule mentions—the vowel (A) and the even number (4). This does not change when a negation is introduced in the consequent. Thus, although participants use the same mechanisms both with the standard and with the negated rule, only in the latter case do they reach the correct answer. In neither case do they reach this initial answer because of a confirmation bias.

Mechanisms of pragmatic inference act rapidly: Participants' attention turns towards the relevant cards quickly after reading the rule (Ball, Lucas, Miles, & Gale, 2003). However, most participants take several minutes to complete the task. What happens in this lapse of time? Usually, not much. Most participants stick to their intuitive answer until the end. Yet they haven't been idle. Think aloud protocols reveal that they have considered many reasons… but they overwhelmingly thought of reasons why they should pick the cards they had intuitively selected. If they think about others cards at all, it is only to gather reasons why they are not relevant (Lucas & Ball, 2005). This looks like a confirmation bias: Participants seem bent on confirming their initial intuition. We'll see presently that this is more properly describe as a myside bias, but a few observations are already worth making. First, that there is no confirmation bias

in the mechanisms that guide participants' initial, intuitive reaction. Instead the bias affects reasoning, it makes them consider only a biased sample of reasons. Second, the bias can affect reasoning even in the absence of emotions or commitment (in the standard Wason selection task the rule is abstract, and the participants do not publicly commit to their intuitive answer before reasoning about it).

Thought listing

The last demonstration of the confirmation bias we'll discuss here is also the simplest: asking people to list their thoughts. For instance, list your thoughts on a controversial issue—the death penalty, immigration, tax rates. Take your time, list as many thoughts as you'd like. Now go over your list and count how many thoughts mention arguments going against your own position on the issue (assuming you do have a position). Say, if you have listed thoughts on the death penalty, and you think it should be abolished, how many arguments did you list supporting its use? Even arguments that you went on to refute? A good bet is that you listed very few, if any, such arguments (see, e.g., Kuhn, 1991; Perkins, 1989). The ease with which we find arguments in support of our positions, and the difficulty of finding arguments against our positions has been interpreted as proof of a confirmation bias (e.g. Edwards & Smith, 1996).

Some readers, however, might have noticed a flaw in this demonstration. Are people who give few arguments against their own position necessarily biased? On the contrary, it might seem like the sane thing to do, by contrast with mostly providing arguments against the opinions we hold. A rational thinker might be expected to consult the arguments relevant to a given topic and, on this basis, form a position. If our thinker happens to mostly know of arguments

supporting one position, then she should list these arguments and endorse the position they support. As a result, her list of arguments would be hard to distinguish from that of a biased thinker who starts with a set position and then looks for arguments in its favor. This is a sound objection, but more sophisticated experiments have shown beyond doubt that something like the confirmation bias affects how people gather their thoughts. But before moving on to these more sophisticated experiments, we should briefly review a simple one that makes an important point.

In the thought listing exercise above, you had been asked to list thoughts and, if you are like most people, most of these thoughts could be said to 'confirm' your position—hence the inclusion under the 'confirmation bias' label. What would have happened if you had been asked instead to reflect on a position you disagree with? Would you have also been biased to find 'confirming' arguments, that is, arguments that confirm the position you disagree with? No, instead you would have been biased to find disconfirming arguments, arguments that refute the position you disagree with (Edwards & Smith, 1996). We do not have a general confirmation bias—a bias to confirm anything we think about—but a myside bias—a bias to find arguments that support our position (see Text box 2).

--- insert Text box 2 about here ---

Edwards and Smith (1996) also showed that the myside bias was more forcefully expressed when the participants reflected on an argument they disagreed with. Participants were more motivated to refute such an argument than to support an argument they agreed with. This asymmetry makes sense in a social context. When we interact with people, we are only expected to offer justifications when we disagree with them, not when we agree. The myside bias

might thus be helping us find arguments in anticipation of having to justify our position—a suggestion I will explore further below.

The most conclusive evidence for the myside bias was produced thanks to an original experimental paradigm called 'choice blindness.' In choice blindness experiments, participants answer a question, and are then tricked into believing that they have answered something different from their actual answer. In the most relevant choice blindness experiment, participant were asked to give their opinion, on a scale from "completely disagree" to "completely agree," on a series of moral statements such as "If an action might harm the innocent, it is morally reprehensible to perform it" (Hall, Johansson, & Strandberg, 2012). After she handed back her questionnaire, the experimenter performed a sleight of hand which inverted the meaning of some of the statements, so that the question above, for instance, would now read "If an action might harm the innocent, it is morally permissible to perform it." The answer scales, however, were not inverted. As a result, if a participant had agreed to the first statement, she now agreed with a statement meaning the exact opposite. Approximately half of the participants did not detect the manipulation. These participants then justified positions they thought they held, even though they had taken an exactly opposite stance a few minutes before. Indeed, the participants were as biased when they were justifying 'fake', attributed positions than when they were justifying their genuine positions, overwhelmingly producing supporting arguments in both cases.

This is a perfect demonstration of the myside bias. It is impossible to claim that the participants were looking for reasons and then building a position on this basis. Instead, they were looking for reasons to justify whatever position

they thought they held. Other experiments—for instance in the large literature on motivated reasoning (Kunda, 1990)—offer evidence of a myside bias, but none are quite as persuasive as these choice blindness experiments.

LIMITS OF THE MYSIDE BIAS

I have argued that there is no evidence of a general confirmation bias. Instead, the evidence points to a myside bias—a tendency to look for reasons that support one's opinions. This definition suggests two important limits to the myside bias. First, it only affects a process that deals with *reasons*—that is, reasoning. Second, it only affects how people *look* for reasons, not how they evaluate reasons. It's time to flesh out both of these claims.

Reasoning is sometimes understood as a general process of inference, including unconscious inferences. It is useful, however, to restrict the term 'reasoning' to more specific cognitive mechanisms, mechanisms that deal with reasons. The vast majority of the inferences we perform—for instance going from "can you pass me the salt?" to "please pass me the salt"—are performed without paying attention to the reasons why we perform them—we do not have to go through a chain of reasoning such as 'it is obvious that I can pass the salt, and this information would be of no relevance to my interlocutor, therefore she must mean something else, probably that she wants me to pass her the salt' (Mercier & Sperber, 2011). In this perspective, reasoning is a cognitive mechanism that allows us to find reasons and to evaluate the support relationships between a reason and a conclusion. For instance, reasoning allows a speaker to find a reason—"you can easily reach it from where you're sitting"—to support a request—"please pass me the salt"—and it allows her interlocutor to tell how much support the reason lends to the request.

Cognitive mechanisms—including reasoning—can have a variety of biases. For instance, it makes sense for food selection mechanisms to be biased towards judging that a substance is toxic, given the relative costs of ingesting a toxic substance (high) compared to passing over an edible one (low). However, there is no evidence of a confirmation or myside bias in cognitive mechanisms besides reasoning. This should not be surprising given that such a bias would be widely maladaptive. An animal eager to confirm mistaken beliefs—that there are no predators around for instance—would not survive very long.

Even as a specific feature of reasoning, the myside bias is often believed to affect equally how people find reasons and how they evaluate them. In particular, people are supposed to evaluate other people's arguments very critically when they disagree with their conclusion. I mentioned earlier research by Edwards and Smith (1996) showing that participants were particularly keen to find thoughts refuting an argument whose conclusion they disagreed with. As a result, participants rated these arguments as being very poor, which suggests that argument evaluation is indeed biased. The issue with such experiments, however, is that they conflate the way people rate arguments after they have had time to gather counter-arguments, and the way they evaluate arguments as they read (or hear) them.

For instance, imagine a participant who believes that the death penalty should be abolished, and who reads the following argument:

> Sentencing a person to death ensures that he/she will never commit another crime. Therefore, the death penalty should not be abolished. (Edwards & Smith, 1996, p. 9)

Does she reject the argument as soon as she realizes that it challenges her beliefs? Or does she do so because she has been able to find counter-arguments, for instance that there are more moral ways of ensuring that someone does not commit another crime? This is an important distinction. If there is a strong bias when people immediately evaluate the argument, such that arguments with unpalatable conclusion are likely to be immediately rejected, then argumentation should be largely pointless: People would not change their minds when exposed to challenging arguments. By contrast, if the bias mostly stems from the production of arguments that follows immediate evaluation, then there is hope: The counter-arguments raised at this stage can be addressed, in the course of a discussion for instance.

Although we cannot rule out that evaluation is intrinsically biased, two pieces of evidence suggest that most of the bias stems from the production of counter-arguments that takes place after people have been confronted to a challenging argument. The first piece of evidence is that there is a relationship between how many counter-arguments people produce and how negatively they rate the argument (e.g. Edwards & Smith, 1996). The second is that when participants are allowed to discuss issues at length, then they can be convinced to change their minds (see Mercier & Sperber, 2011). For instance, in a series of experiments participants were confronted with arguments defending the good answer to a logical task (the disjunctive reasoning task from the end of Text box 1) (Trouche, Sander, & Mercier, 2014). When participants who had given the wrong answer were only exposed to one good argument, less than half of them changed their mind. By contrast, when participants were able to discuss the problem at length, so that their counter-arguments could be addressed, all of

those exposed to the correct answer changed their mind—even if they had been extremely confident before the discussion.

These results suggest that when we read or hear an argument, we might not be biased in our evaluation of the argument. By contrast, we are biased in looking for arguments after this initial evaluation: If we are not swayed by the argument, we find arguments that counter it. The myside bias could thus affect argument evaluation indirectly only, through the biased production of counter-arguments.

EXPLAINING THE MYSIDE BIAS

When told about the myside bias, most people are not surprised. On the contrary, the myside bias seems to explain why they sometimes fail to convince others (not, obviously, why they sometimes fail to be convinced…). Some manifestations of the myside bias, however, should be quite puzzling. When people reason on their own, the myside bias often has dire epistemic consequences—piling up reasons that support our preconceived views is not the best way to correct them. Not only does the myside bias stop people from fixing mistaken beliefs—as in the Wason selection task—but it can even make things worse. When they reason on their own, people can become overconfident and strengthen their preexisting beliefs (for references, see, Kunda, 1990; Mercier & Sperber, 2011). A similar phenomenon takes place when people reason with people who agree with them: Arguments for the side everyone agrees on pile up without being criticized, and the average opinion can become more extreme (Isenberg, 1986).

In light of its consequences, we are entitled to ask: Why do we have a myside bias? Two broad classes of explanations are usually offered, cognitive

and motivational. The former attempts to explain the myside bias through other, more fundamental biases in the operation of cognitive mechanisms. For instance, here is the explanation ventured by Jonathan Evans, which is fairly representative:

> Subjects confirm, not because they want to, but because they cannot think of the way to falsify. The cognitive failure is caused by a form of selective processing which is very fundamental indeed in cognition—a bias to think about positive rather than negative information. (Evans, 1989, p. 42)

The issue with such an explanation is that it only applies to a putative confirmation bias, not to the myside bias. But we do not have a confirmation bias—when it comes to ideas we disagree with, falsification comes easily. Thus there does not seem to be any hard cognitive constraints that would favor confirmation over falsification. And no cognitive constraint has been offered that would account for the myside bias.

Motivational explanation suggest that people are motivated to hold some beliefs, and that reasoning serves to prop up these beliefs, to insulate them against attacks (Kunda, 1990). But why would we want to hold some beliefs in particular? Beliefs are there to guide our actions, and mistaken beliefs lead to harmful behavior. Why wouldn't we want our beliefs to be the best approximation of reality we can come up with? Holding some beliefs can make us happy, but cognitive mechanisms do not evolve because they make us happy, they evolve because they increase our fitness. In any case, people are often biased to find arguments supporting beliefs that make them miserable—that their partner is cheating on them, that they are about to be fired, etc. It seems

that neither cognitive nor motivational explanations can account for all the observed features of the myside bias.

Another way of thinking about the myside bias is in terms of its potential usefulness. What is the myside bias good for? Clearly, it doesn't help us reach sounder beliefs on our own. But when we have to defend our beliefs, to convince someone else of their worth, then a myside bias is desirable. It is easier to convince an audience with arguments that support our position than with arguments that challenge our views. This fits well with the argumentative theory of reasoning, which suggests that argumentation is the main function of reasoning (Mercier & Sperber, 2011). In a nutshell: Humans cooperate on an unprecedented scale among primates. To do so, they rely on equally unprecedented communication skills. Successful communication, however, raises an evolutionary challenge. For communication to be stable, senders must not be able to abuse receivers to the point at which communication stops being beneficial. In humans, receivers defend themselves against harmful communication by filtering messages. They are more likely to accept messages that fit with their prior beliefs, and messages sent by people they trust. Argumentation enables the transmission of messages when trust isn't sufficient, thereby enlarging the scope of what can be successfully communicated. For instance, if I had peremptorily asserted the position held here on the myside bias, few readers would have been taken my word for it. By contrast, laying out arguments to support this position is (hopefully) more likely to convince. According to the argumentative theory, argumentation would be the main function of reasoning.

From the perspective of the argumentative theory of reasoning, the myside bias should be specific to reasoning—as it seems to be. Crucially, the myside bias should only apply to how we find arguments. When we evaluate others' arguments—at least when we disagree with their conclusion, the most common case—the main task of reasoning is to recognize good arguments and to make us change our mind accordingly, so that we end up with better beliefs on average. A strong myside bias in evaluation would make argumentation pointless. As argued above, a good case can be made that argumentation evaluation is either unbiased, or at the very least not so biased as to preclude effective argumentation. Finally, as might be expected of an evolved, adaptive trait, no one seems to be exempt from the myside bias—intelligence, motivation, and open-mindedness do not protect against the myside bias (see, e.g., Stanovich, West, & Toplak, 2013).

If the argumentative theory of reasoning can account for the main features of the myside bias, one could still wonder why it also affects solitary reasoning, or even why people reason on their own altogether. Two lines of explanations can be sketched here. The first is that there are likely social benefits to reasoning on one's own. Solitary reasoning allows discarding beliefs that we can't justify (Kunda, 1990), and it makes us ready to defend those we can. The second is that the extent of solitary reasoning we observe in our cultures might be due to the huge differences between the environment in which we evolved—the environment for which reasoning and its myside bias were shaped—and the current environment. Denizens of modern societies are exposed to an unprecedented variety of opinions, along with the many disagreements this leads them to anticipate.

Chapter 5

DEBIASING THE MYSIDE BIAS OR MAKING THE BEST OF IT?

Many debiasing techniques have been tried to reduce the impact of the myside bias (for review, see Lilienfeld, Ammirati, & Landfield, 2009). They have met with limited and temporary success. One of the techniques that sometimes work is to ask participants to consider reasons for answers different from their own. However, in other cases participants have had trouble coming up with reasons why they might be wrong (Kuhn, 1991; Perkins, 1989), so the scope for the efficacy of this technique might be limited.

A more promising approach is to make the best of the myside bias. When people who disagree exchange arguments, the myside bias, from a potential flaw, turns into an efficient way to divide cognitive labor. Instead of having to consider the pros and cons of every answer, each discussant only has to look for the pros of his answer, and the cons of the others' answers. To the extent that people feel free to voice contrary opinions, and that they have no outstanding pressure to stick to their original ideas, the myside bias should not stop the best ideas from spreading in the group (for review see, Mercier & Sperber, 2011). For this to work, however, people must have different opinions to start with. If this is not the case, then substitutes can be used—devil's advocates, red teams—but they are not quite as efficient as genuine dissent.

Performance on the Wason selection task illustrates the stark contrast between the two approaches. Debiasing techniques that target individual performance—teaching people about conditionals, or explaining the good answer to another version of the task—have had very limited effects. By contrast, asking participants to solve the task in small groups creates, on average, a fivefold increase in performance (e.g. Moshman & Geil, 1998). Such

results strongly suggest that psychologists and educators should focus less on trying to improve solitary reasoning at all costs, and teach instead how to make the best of interactions with others.

SUMMARY

• The confirmation bias is a tendency to seek information that confirms a thought we currently consider.

• Famous experiments—the rule discovery task, experiments on interrogation behavior, the four cards selection task—purported to demonstrate the existence of a confirmation bias. Properly interpreted, these experiments in fact reveal rational behavior, not a confirmation bias.

• However, other experiments reveal a myside bias: A tendency to find arguments that defend our beliefs, whether they are supportive (if we agree with something) or refutational (if we disagree with something).

• The myside bias can have dire consequences when people reason on their own or with like-minded peers, leading to overconfidence and polarization.

• However, the myside bias does not preclude people from accepting arguments that challenge their position, provided their counter-arguments can be addressed.

• Standard cognitive and motivational accounts of the myside bias are unconvincing. Instead, the myside bias might play an adaptive role in making people better at finding arguments in order to convince others.

• The best way to deal with the myside bias might not be to fight it by trying to improve on solitary reasoning, but to make the best of it by making people exchange arguments with each other.

FURTHER READING

Nickerson (1998) provides a good overview with many intriguing examples from outside the lab. Oswald and Grosjean (2004) provide an earlier review which details how to administer the interrogation behavior task. For specific paradigms, see Klayman and Ha (1987), and Poletiek (1996) for hypothesis testing; Sperber *et al*. (1995) for the Wason selection task; Hall *et al*. (2012) for choice blindness; Kunda (1990) for motivated reasoning; and Kuhn (1991) for informal arguments. Mercier and Sperber (2011) provide an introduction to the argumentative theory of reasoning, with a lengthy discussion of the role the confirmation bias (which should have been referred to as myside bias) plays in it. For a review of the debiasing literature, see Lilienfield *et al*. (2009).

REFERENCES

Ball, L. J., Lucas, E. J., Miles, J. N., & Gale, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *The Quarterly Journal of Experimental Psychology*, *56*(6), 1053–1077.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*, 5–24.

Evans, J. St. B. T. (1989). *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: Lawrence Erlbaum.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*(3), 391–397.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PloS One*, *7*(9), e45457.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141–1151.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Kuhn, D. (1991). *The skills of arguments*. Cambridge: Cambridge University Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*(4), 390–398.

Lucas, E. J., & Ball, L. J. (2005). Think-aloud protocols and the selection task: Evidence for relevance effects and rationalisation processes. *Thinking and Reasoning*, *11*, 35–66.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, *4*(3), 231–248.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology*, *2*, 175–220.

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Hove, UK: Psychology Press.

Perkins, D. N. (1989). Reasoning as it is and could be: An empirical perspective. In D. M. Topping, D. C. Crowell, & V. N. Kobayashi (Eds.), *Thinking Across*

Cultures: The Third International Conference on Thinking (pp. 175–194). Hillsdale, NJ: Erlbaum.

Poletiek, F. H. (1996). Paradoxes of falsification. *Quarterly Journal of Experimental Psychology*, *49A*, 447–462.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202–1212.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, *22*(4), 259–264.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168.

Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology*, *19*(6), 560–576.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, *143*(5), 1958–1971.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–137.

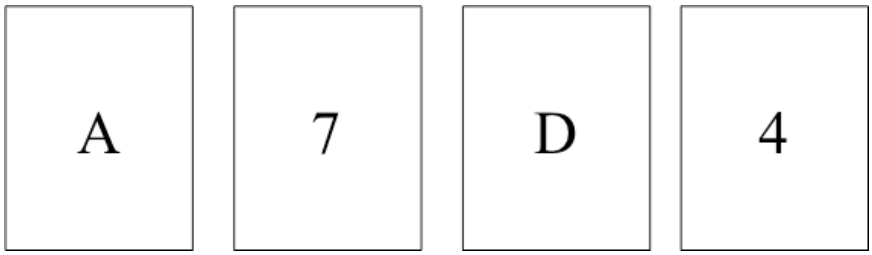
Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology: I* (pp. 106–137). Harmandsworth, England: Penguin.

Text box 5.1 Details of the classroom demonstration

The experiment has two phases: an individual phase and a group phase.

Individual phase.

Distribute to the students sheets with a standard Wason selection task such as the following:

> Here's a rule regarding the four cards below: "If there is a vowel on one side, then there is an even number on the other side."
>
> 
>
>
> What card(s) need(s) to be turned over to determine whether the rule is true of these four cards? You can select as many cards as you want.

Give them 5 minutes to complete the task, then ask them to make, at the back of the sheet, a list of thoughts about the correct answer. Each thought should be put on a separate line (this should take another 5 minutes).

Group phase.

Form groups of students, ideally at random and of 4 or 5 students. Distribute a new sheet with the same problem, one per group, and ask the students to reach a consensus regarding the correct answer. Give them at least 15 minutes to do so.

Gathering the results.

First, give and explain the correct answer. Ask how many students gave the correct answer during the individual phase, and how many did so during the group phase (with a show of hands for instance). Then ask the students to count the total number of thoughts they listed during the individual phase, as well as how many of these thoughts comprise arguments going against their answer, even if they are refuted (for instance, if they have not selected the 4 card: "the 4 card seemed relevant because the rule mentions even numbers, but in fact it doesn't matter"). The students announce their counts turn by turn while you make a running total for both numbers.

Analyzing the results.

A binomial test can be used to show that there are fewer thoughts containing arguments that attack the students' answers than expected at chance. This suggests that participants suffered from a myside bias, which might explain why most of them stuck to the wrong answer. A chi-square can compare the number of correct and incorrect responses in the individual and group phases. The improvement between the individual and the group phases shows that in spite of their myside bias, students who had the wrong answer and were exposed to the correct answer were able to change their mind and to adopt a better answer.

FAQ.

<u>Could the Wason selection task be too difficult, or be already known by the students?</u>

The experiment is much more likely to succeed if a few students are able to find the correct answer on their own. If there is a danger that this might not

happen (e.g. the class is very small), then using a marginally easier logical task might be preferable. Also, many psychology students might already know of the Wason selection task. Several alternative problems might be considered, such as those of the Cognitive Reflection Test or, better, its recent extension by Toplak *et al*. (2014).

Aren't the statistics wrong?

Comparing the raw results of the individual and group phases is indeed not quite correct. When the N is large enough, other solutions can be implemented, such as considering each group as a single data point, or even comparing the scores of the real groups to that of nominal groups during the individual phase. Similarly, the tests demonstrating the existence of the myside bias should be performed on individual averages. In both cases however, the effects tends to be very strong, so that all statistical methods yield similar results.

Does failing to give arguments against one's own answer really represent a myside bias?

Not necessarily. In the 'though listing' subsection of the main text, I explain the limits of this method and give arguments suggesting that in spite of these limits, it is likely to reflect the operation of a myside bias in most cases.

Is the improvement in the group phase necessarily due to people accepting challenging arguments in spite of their myside bias?

Yes. Unfortunately there is no space here to cover this ground, but see for instance section 2.3 of Mercier and Sperber (2011), and Trouche *et al*. (2014). For instance, experiments have shown that the improvement is not simply due to the fact that the participants are asked the same question a second time, and

thus given more time to think: if they are put in groups straight away, they also perform better than individuals (Moshman & Geil, 1998).

Text box 5.2 Confirmation bias and myside bias.

Strictly speaking, a *confirmation bias* is a tendency to confirm whatever one thinks about. This should apply even when people disagree with what they are thinking about, which is implausible and has been proven to be false (Edwards & Smith, 1996). When people consider a thought they disagree with, they are biased towards falsification: they find exceptions, counter-arguments, counter-examples. It is thus wrong to speak of a general confirmation bias. Instead, people have a *myside bias*: a tendency to find arguments that defend their position, whether this entails supporting a position they agree with, or refuting a position they disagree with.

FOOTNOTE

[i] http://jonronson.posterous.com/julie-burchill-and-psychopaths retrieved February 25 2013.