Biased argumentation and critical thinking

Vasco CORREIA, Universidade Nova de Lisboa

*A man with a conviction is a hard man to change. Tell him you disagree and he turns away. Show him facts or figures and he questions your sources. Appeal to logic and he fails to see your point.*

Leon Festinger (2008 [1956]: 3)

1. Introduction

Although the problem of biased argumentation is sometimes reduced to the problem of intentional biases (sophistry, propaganda, deceptive persuasion)\(^1\), empirical data on human inference is consistent with the view that people are often biased not because they want to, but because their emotions and interests insidiously affect their reasoning (Kunda 1990, Baron 1988, Gilovich 1991). Walton (2011: 380) highlights this aspect:

Many fallacies are committed because the proponent has such strong interests at stake in putting forward a particular argument, or is so fanatically committed to the position advocated by the argument, that she is blind to weaknesses in it that would be apparent to others not so committed.

This phenomenon is known as «motivated reasoning» and typically occurs unintentionally, without the arguer’s awareness (Mercier & Sperber 2011: 58, Pohl 2004: 2). For example, a lawyer may be biased in the defense of a client because (s)he deliberately intends to manipulate the jury, to be sure, but also because the desire to win the case (or the sympathy toward the client, etc.) unconsciously distorts the way (s)he reasons and processes the relevant evidence. In such cases, the arguer is sincerely convinced that his or her

arguments are fair and reasonable, while in fact they are tendentious and fallacious.

In the past decades, empirical research in psychology and neurosciences not only confirmed that emotions greatly affect human reasoning, but also revealed to what extent and in which variety of ways this seems to happen. Although the significance of these studies has been questioned by some authors, the dominant view is that people tend to fall prey to a host of cognitive and motivational biases that affect their inferential and judgmental reasoning. As Larrick (2004: 316) observes, «the existence of systematic biases is now largely accepted by decision researchers, and, increasingly, by researchers in other disciplines». At any rate, psychologists now investigate dozens of types of cognitive illusions (Pohl 2004: 1, Thagard 2011: 164).

The phenomenon of motivated reasoning poses a considerable challenge for normative theories of argumentation, which tend to assume that the rules of logic and dialectic are sufficient to ensure the reasonableness of people’s arguments. Insofar as motivational biases tend to occur unconsciously, it appears that even well-intended arguers, who genuinely wish to reason in fair terms, may end up putting forward arguments that are skewed and tendentious. To that extent, the intentional effort to observe the rules of argumentation may not suffice to ensure the rationality of debates. As Thagard (2011: 157) points out, «it would be pointless to try to capture these [motivated] inferences by obviously fallacious arguments, because people are rarely consciously aware of the biases that result from their motivations». Moreover, this difficulty is aggravated by the fact that arguers often tend to

---

2 Some researchers argue that the discrepancies between normative models of rationality and people’s reasoning are not indicative of human irrationality, but rather the result of (1) random performance errors, (2) computational limitations of the human brain, (3) a misconception of the relevant normative standards of rationality, and (4) a different interpretation of the task by the subject (for a review, see Stanovich & West 2000). Gigerenzer (2008: 13), in particular, suggests that cognitive illusions are in fact adaptive forms of reasoning which promote the achievement of goals under constrains of time and information. Cohen (1981: 317), on the other hand, insists on the shortcomings of some empirical studies, which, in his view, neglect the differences of interpretation between the experimenters and the subjects, and involve mental tasks that are not representative of the normal conditions of reasoning. Other researchers have also questioned the presumed universality of such results, which are based almost exclusively on samples drawn from western, educated populations (Heinrich et al. 2010).
rationalize their biases; in other words, come up with good ‘reasons’ to justify *post-factum* beliefs initially acquired under the influence of motives (desires, goals, emotions).

This chapter has two purposes. The first is to elucidate some of the ways in which motivational biases lead arguers to commit unintentional fallacies. Drawing on recent work in psychology and argumentation theory, I explore the hypothesis that there are privileged links between specific motivational biases and specific forms of fallacious reasoning. To make this point clearer, I propose to categorize motivational biases in three different classes, according to the type of motive that underlies them: (1) *wishful thinking*, in which people are led to believe that p because they desire that p, (2) *aversive thinking*, in which people are led to believe that p because of the anxiety that not-p, and (3) *fretful thinking*, in which, as Thagard (2011: 159) explains, «people believe something, not just despite the fact that they fear it to be true, but partly because they fear it to be true».

The second purpose of this chapter is to argue that, even though motivational biases are in principle unintentional, there are certain control procedures that arguers can adopt if they wish to counteract their error tendencies. I identify several ‘debiasing strategies’ and argue that they can significantly contribute to promote the rationality of people’s argumentative reasoning, both at a dialogical and at an individual level. While most normative theories of argumentation focus on the problem of establishing the ideal rules of how people ought to argue, the last section of this article focuses on the problem of what discussants can do to effectively adjust their behavior to those rules.

The proposed categorization of motivational biases provides a structure for the three first sections of this chapter. In section 2, I examine the effects of wishful thinking in everyday discourse and suggest that it underlies our tendency both to commit the *argumentum ad consequentiam* fallacy and to fall prey to the ‘confirmation bias’, which in turn tends to accentuate the problem of polarization of opinions. In section 3, I argue that there is an intimate correlation between aversive thinking and two forms of fallacious reasoning: ‘misidentifying the cause’, on the one hand, and ‘slothful induction’, on the other. Furthermore, I explore Festinger’s (1957) hypothesis that aversive thinkers tend to become ‘defensive’ and to explain away their inconsistencies through rationalizing (rather than rational) arguments. In section 4, I examine...
the puzzling phenomenon of ‘fretful thinking’ – or ‘fear-driven inference’ (Thagard 2011: 159) – in which arguers are biased in a self-defeating way, typically by ‘jumping to (negative) conclusions’ or by committing the ‘slippery slope’ fallacy. Finally, in section 5, I briefly examine several debiasing strategies and show how they can be useful to counteract the effects of biases upon people’s everyday reasoning.

2. Wishful thinking

Wishful thinking is generally described as a form of motivated reasoning in which the subject is led to conclude that p under the influence of a desire that p. Although this phenomenon is sometimes reduced to the inference ‘I wish that p, therefore p’, it seems doubtful that wishful thinkers actually commit the fallacy in those terms. As much as I may wish to be beautiful, rich and famous, for example, it is clear that the inference ‘I wish to be beautiful, rich and famous, therefore I am beautiful, rich and famous’ is very unlikely to persuade me.³

In most cases of wishful thinking, arguers are not aware that they reach the conclusion that p merely because they desire that p. Instead, desire-driven inferences seem to involve a more complex and indirect type of fallacy – namely, a fallacious version of the argument from consequences: ‘If p than q. I wish that q. Therefore p’. As Walton (2006: 106) observes, the argument from consequences may or may not be fallacious, depending essentially on three critical questions: (1) how likely it is that the consequence will follow, (2) what evidence is provided to support the claim that the consequence will follow, and (3) whether there are consequences of the opposite value that ought to be taken into account. If the arguer’s claim is based merely on the desirability of the consequence, regardless of its likelihood, his or her argument is probably fallacious. Consider the following example: “Of course

³ Most philosophers and psychologists agree that it is impossible to decide to believe something bi et nume (direct doxastic voluntarism), although it may be possible to control indirectly certain beliefs, for example via a selective exposure to the available evidence (see Mele 2001b, for a review).
the environment talks will succeed. Otherwise it means mankind is on the
way out» (Pirie 2006: 176). Although the mutual desire to avoid a catastrophic
situation is a valid reason to believe that the parties involved in the talks will
reach an agreement, this reasoning seems fallacious insofar as the undesirable
consequences of a failure in the environment talks cannot by themselves
guarantee that they will succeed (other factors, such as economic interests,
may prevail).

Mele (2001a: 87) observes that desires tend to induce irrational reasoning
indirectly, by affecting the subject’s attention to the available evidence:

Data that count in favor of (the truth of) a hypothesis that one would like to be true
may be rendered more vivid or salient given one’s recognition that they so count; and
vivid or salient data, given that they are more likely to be recalled, tend to be more
‘available’ than pallid counterparts.

This helps explain why wishful thinking is often associated with the well-
documented confirmation bias, which consists precisely in the tendency to
search for evidence that supports what we already believe in, or what we want
to be true (Baron 1988: 280, Oswald & Grosjean 2004: 79). Thus, for
example, the desire that my philosophical position is correct may
surreptitiously lead me to focus too much on sources that seemingly confirm
it, and not enough on sources that seemingly disconfirm it. Likewise, «people
who want to believe that they will be academically successful may recall more
of their past successes than of their failures» (Kunda 1990: 483). In a classic
experiment, Lord et al. (1979) were able to demonstrate that this bias tends
to aggravate the phenomenon of ‘attitude polarization’ even when people are
exposed to the same body of information. The researchers exposed subjects
supporting and opposing the death penalty to descriptions of two fake
studies, one confirming and one disconfirming the hypothesis that capital
punishment deters violent crime. Predictably, they found that both
proponents and opponents of the death penalty rated the study that
confirmed their own views as more convincing and probative. Less
predictably, though, they found that the ‘pro’ subjects became even more
favorable to the capital punishment after being exposed to the information,
and that the ‘anti’ subjects became even more opposed to it. In other words,
the polarization of opinions seemed to have increased after exposure to
information, despite the fact that the information was the same.
Assuming that this phenomenon is caused by unconscious biases, it becomes all the more difficult to prevent it. People who succumb to the confirmation bias do not try to find support for their preexisting beliefs by deliberately twisting or misinterpreting the available evidence. As Kunda (1990: 494) explains, the problem is rather that “cognitive processes are structured in such a way that they inevitably lead to confirmation of hypotheses”. The classic assumption to explain such tendencies is that people are generally motivated to protect their belief system from potential challenges (Albarracín & Vargas 2009, Festinger et al. 2008 [1956], Mercier & Sperber 2011). Festinger et al. (2008 [1956]: 3) write: «We are familiar with the variety of ingenious defenses with which people protect their convictions, managing to keep them unscathed through the most devastating attacks». Interestingly enough, some studies indicate that people who are confident about the resilience of their beliefs are more willing to examine evidence that contradicts them, and, conversely, that people who are doubtful about their ability to defend their beliefs from future challenges tend to prefer exposure to information consistent with them (Albarracín & Mitchell 2004). According to Mercier and Sperber (2011: 65) the confirmation bias helps arguers meet the challenges of others and even contributes to a prosperous division of cognitive labor, «given that each participant in a discussion is often in a better position to look for arguments in favor of his or her favored solution (situations of asymmetric information)».

The problem, however, is that this tendency to gravitate toward information that justifies our preexisting opinions leads to a polarization of opinions that is arguably detrimental to the purpose of debates. Given the same body of evidence, people with different opinions will tend to focus on elements that are susceptible to cause their views to move even further apart. Lord et al. (1979: 2108) vehemently stress that point: «If our study demonstrates anything, it surely demonstrates that social scientists cannot expect rationality, enlightenment, and consensus about policy to emerge from their attempts to furnish ‘objective’ data about burning social issues». In addition, this aspect seems to be enhanced by the way people tend to consume information nowadays, as Mooney (2011: 3) observes, «through the Facebook list of friends, or tweets that lack nuance or context, or narrowcast and often highly ideological media that have small, like-minded audiences».
3. Aversive thinking

Whereas in wishful thinking arguers unduly infer that \( p \) is true because of the desire that \( p \), in aversive thinking arguers unduly infer that \( p \) because of the anxiety that not-\( p \). Although the anxiety that not-\( p \) is generally accompanied by a correlative desire that \( p \), the two affects seem to be causally independent and may trigger different forms of motivated reasoning (Barnes 1997: 52, Johnston 1989: 72). Like wishful thinking, aversive thinking is considered to be a ‘positive’ illusion inasmuch as it yields a significant psychological gain, namely: the reduction of the subject’s anxiety.

Aversive thinking tends to arise when arguers are confronted with evidence suggesting that what they fear might be true. For example, a man who is diagnosed with a terminal illness may reject the doctors’ arguments and persist in believing that he will survive. His way of reasoning is presumably constrained by the anxiety of thinking that his days are numbered, which can of course be psychologically devastating. But aversive thinking need not be so extreme. In everyday debates, people often become ‘defensive’ simply because one of their convictions is being challenged. As Johnson & Blair (1983: 193) observe, this seems to happen in virtue of the arguer’s ‘egocentric commitment’ to a given standpoint, as when people are blinded by their attachment to an ideology, a group or an institution. According to Sherman & Cohen (2002: 120) such defensive responses stem more fundamentally from a motivation to protect self-worth and the integrity of the self: «Because the motivation to maintain self-worth can be so powerful, people may resist information that could ultimately improve the quality of their decisions».

One of the most effective forms of aversive thinking is rationalization, i.e., the effort to justify an irrational attitude by invoking ‘good’ reasons instead of the true reason. It is notoriously difficult to refute the arguments of a person who rationalizes, given that the reasons (s)he invokes are not necessarily false. Someone addicted to pills, for example, may be able to provide seemingly reasonable explanations for abusing medication (stress at work, domestic problems, headaches, sleeping disturbances, etc.). Yet, much like the alcoholic who claims to drink for a reason, (s)he will refuse to acknowledge that (s)he takes pills mainly because of a drug addiction. Given
the social stigma associated to that diagnosis, the motivation to deny it can be strong enough to distort the way (s)he reasons.

In many cases, rationalization leads arguers to commit the fallacy of misidentifying the cause, which according to Tindale (2007: 179) may take two forms: «In the first instance, we may falsely identify X as the cause of Y when on closer inspection a third factor, Z, is the cause of both X and Y. In the second case, we may confuse a cause and an effect: identifying X as the cause of Y when it is actually Y that causes X». The latter case is illustrated precisely by the addicted person who claims that (s)he takes drugs because of all sorts of problems (work, family, health, etc.), when in general those problems are already a consequence of the abuse of drugs (Twerski 1997: 34).

According to the theory of cognitive dissonance (Festinger 1957, Aronson 1969) such responses arise when the person holds two or more ‘cognitions’ (ideas, beliefs, opinions) that are psychologically inconsistent with each other. Inasmuch as the occurrence of dissonance admittedly produces anxiety and psychological discomfort, individuals strive toward consistency within themselves by rationalizing one of the cognitions in question. Hence, Festinger (1957: 3) writes:

The person who continues to smoke, knowing that it is bad for his health, may also feel (a) he enjoys smoking so much it is worth it; (b) the chances of his health suffering are not as serious as some would make out; (c) he can’t always avoid every dangerous contingency and still live; and (d) perhaps even if he stopped smoking he would put on weight which is equally bad for his health. So, continuing to smoke is, after all, consistent with his ideas about smoking.

When the attempt to explain away the inconsistency is successful, dissonance is reduced and so is the anxiety associated to it⁴. Several experiments confirm that people tend to rationalize their inconsistencies (for a review, see Albarracín & Vargas 2009). In a classic study, Festinger & Carlsmith’s (1959)

⁴ For one reason or another, however, attempts to achieve consistency may fail and the psychological discomfort persists. Moreover, even successful rationalizations can lead to more anxiety in the long term, as Barnes (1997: 35) points out: “The reduction of anxiety can lead to other anxieties, sometimes far greater ones. The gain, therefore, is not necessarily an all-things-considered gain, nor is it necessarily beneficial for the person”.

Thierry Herman and Steve Oswald - 9783034315470
Downloaded from PubFactory at 08/23/2019 10:19:42PM
test via free access
asked students to work for an hour on boring tasks such as turning pegs a quarter turn over and over again. Participants were then asked to convince another student that the tedious and monotonous tasks were actually enjoyable and exciting. While some of the participants were paid $20 for doing this, others were paid merely $1. Surprisingly, when the participants were asked how much they really enjoyed performing the tasks, those who were paid $1 rated the tasks as more enjoyable than those who were paid $20. The researchers speculated that all the participants experienced dissonance between the conflicting cognitions: ‘The tasks were tedious’ and ‘I told someone that the tasks were exciting’. However, those who were paid $20 had a great deal of justification for lying to the other student, and therefore experienced less dissonance. Those who were paid $1, on the other hand, experienced a greater need to justify their action and presumably persuaded themselves that they really believed what they said.

This tendency to rationalize seems to be particularly strong when the perceived inconsistencies are liable to threaten the arguer’s emotional attachment to the standpoint. In a recent study, Westen et al. (2006) used functional neuroimaging to test motivated reasoning on political partisans during the U.S. Presidential election of 2004. The subjects were shown a set of slides presenting contradictory pairs of statements either from their preferred candidate, from the opposing candidate, or from a neutral figure. In addition, one of the slides presented an exculpatory statement that explained away the apparent contradiction. Then they were asked to consider whether each candidate’s statements were inconsistent or not. Predictably, the subject’s ratings provided strong evidence of motivated reasoning. First, they were substantially more likely to evaluate as inconsistent statements made by the candidate they opposed. And second, they were much more likely to accept the exculpatory statements for their own candidate than those for the opposing candidate. In addition, the scanners revealed that the brain regions specifically involved in emotion processing were strongly activated when the subjects evaluated contradictory statements by their preferred candidate, but not when they evaluated the other figure’s contradictions. The researchers concluded that biases were due to the participants’ effort to reduce cognitive dissonance: «Consistent with prior studies of partisan biases and motivated reasoning, when confronted with information about their
candidate that would logically lead to an emotionally aversive conclusion, partisans arrived at an alternative conclusion» (Westen et al. 2006: 1955).

These results suggest that, in many cases of aversive thinking, people’s arguments are less the reason why they adhere to a given belief than the post factum justification of their preexisting belief. Thus, for example, a person may believe in the immortality of the soul in virtue of a religious education, and nonetheless invoke apparently reasonable arguments to support that conviction, as if those arguments were the reason why (s)he held that belief in the first place. To paraphrase one of Aldous Huxley’s famous quotes in *Brave New World*, it seems fair to say that people often come up with seemingly good reasons to justify beliefs that they initially acquired for bad (or unjustified) reasons. This constitutes, according to Haidt (2010: 355), the fundamental «Problem of Motivated Reasoning: The reasoning process is more like a lawyer defending a client than a judge or scientist seeking the truth».

In more extreme cases of aversive thinking, such as denial, the anxiety toward the undesired conclusion is so intolerable that the subject rejects it in the teeth of evidence. As Mele (1982) suggests, this is possible because psychological inferences are not as compulsory as logical deductions: «One may believe that p is true and that p entails q without believing that q is true; and this is the kind of thing that may be explained by a want, fear, or aversion of the person». In such cases, it all happens as though the subject accepted the premises of the reasoning but not the conclusion that follows, presumably because the anxiety somehow ‘inhibits’ the inferential step. The terminally ill patient who refuses to accept his diagnosis, for example, may acknowledge that the exams are reliable and that the doctors are competent but refuse nonetheless to adhere to what they clearly suggest. Sartre (1943: 100) puts forward a similar hypothesis in the famous interpretation of the homosexual who denies his sexual orientation: «[he] acknowledges all the elements that are imputed to him and yet refuses to draw the obvious conclusion».

Such cases of aversive thinking seem to involve a specific form of *ignoratio elenchi* that Schopenhauer (1831: 11) called «denial of conclusion, per negationem consequentiae», which some informal logicians now term slothful induction, i.e., «the mistake of underrating the degree of probability with which a conclusion follows from evidence» (Baker 2003: 264). In a sense, the fallacy of slothful
induction appears to be the symmetric opposite of the fallacy of hasty generalization, insofar as the later involves ‘jumping to conclusions’ on the basis of insufficient evidence, whereas the former involves failing to draw a conclusion in the face of sufficient evidence (Correia 2011: 120). In some cases this may occur indirectly: The subject appreciates the evidence that p but rejects the very link between p and the undesired conclusion q. When the motivation to deny that q is strong enough, Thagard (2011: 155) writes, «you need to question your belief in if p then q and p, rather than blithely inferring q». To return to the case of the terminally ill patient, it may happen for example that he challenges the predictions of standard medicine and turns to less warranted therapeutic methods in a desperate attempt to deny the imminence of his death.

That being said, it is clear that people do not necessarily engage in aversive thinking whenever they feel reluctant to accept a certain reality. Whether the subject falls prey to a motivated illusion or not seems to depend on at least two factors: on the one hand, the degree of emotional attachment to the belief in question, and, on the other hand, the degree of reliability of the subject’s habits of thinking. I will return to this question in the last section.

4. Fretful thinking

The phenomenon known as ‘fretful thinking’ (Beyer 1998: 108), ‘counterwishful thinking’ (Elster 2007: 384) and ‘twisted self-deception’ (Mele 2001b: 94) is surely the most puzzling form of motivated reasoning. Unlike wishful and aversive thinking, which both tend to create biases that are consistent with the individual’s goals, fretful thinking seems paradoxical in that it engenders self-defeating biases that yield unwelcome beliefs. This is what happens, for example, when a jealous husband is biased into thinking that his wife is having an affair, despite his not wanting it to be the case. Likewise, a pessimistic woman may underestimate her chances of getting a job despite her desire to get the job. In this type of case, it all happens as though the person’s fear that p somehow caused her to acquire the unwarranted belief that p. To that extent, as Thagard (2011: 159) points out,
Correia

«fear-driven inference is doubly irrational, from both a practical and theoretical perspective, because it gives the thinker unhappiness as well as erroneous beliefs». The author considers several plausible examples of this:

(a) My lover looks distant, so he/she must be having an affair.
(b) I haven’t heard from my teenager for a few hours, so he’s probably in trouble.
(c) This rash means I have leprosy or some other serious disease.
(d) The editor’s delay in responding to my article means he/she hates it.

As it appears from these examples, fretful thinking often leads arguers to jump to (negative) conclusions without sufficient evidence. In general, this seems to happen because the subject focuses too much on the negative aspects of the issue, presumably due to the influence of a negative emotion (fear, jealousy, anxiety, etc.) on the way she processes information. Regarding the case of the jealous man, for example, it seems reasonable to suggest that «[he] believes that his wife is unfaithful because of the effects of his jealousy on the salience of his evidence or on the focus of his attention» (Mele 2001b: 101). Rather than considering indiscriminately all the relevant evidence, he tends to focus exclusively on the elements that seem to confirm his worst suspicions, and subsequently falls into the illusion that his wife is (probably) cheating on him. This aspect is perhaps more obvious in pathological cases of ‘morbid jealousy’ (or ‘Othello Syndrome’), in which individuals incessantly accuse their partner of infidelity «based on incorrect inferences supported by small bits of ‘evidence’ (e.g., disarrayed clothing or spots on the sheets), which are collected and used to justify the delusion» (American Psychiatric Association 2000: 325). A similar explanation plausibly accounts for the case of the parents who jump to conclusions regarding their teenager’s safety or the hypochondriac who panics because of a mere rash: a negative emotion leads them to contemplate uniquely the negative aspect of things and the arguments they put forth tend to be affected by a pessimism bias.

In such cases arguers seem to fall prey to what Schkade & Kahneman (1998: 340) call the focusing illusion: «When a judgment about an entire object or category is made with attention focused on a subset of that category, a focusing illusion is likely to occur, whereby the attended subset is overweighed relative to the unattended subset». While it may not be
Fallacious, strictly speaking, to confine one’s reasoning solely and exclusively to the negative side of things, such a tendentious interpretation of the available evidence is likely to undermine both the rationality and the credibility of the resulting arguments. Walton also highlights this point:

An argument is more plausible if it is based on a consideration of all the evidence in a case, on both sides of the issue, than if it is pushing only for one side and ignoring all the evidence, even if it may be good evidence, on the other side. So if an argument is biased, that is, if it pushes only for one side, we discount that argument as being worthless. (Walton: 2006: 238)

There also seems to be a privileged link between the phenomenon of fretful thinking and the fallacy of slippery slope, given that the latter typically leads the arguer to draw a dreadful conclusion from a somewhat dubious causal association between events. As Walton (2006: 107) observes, the slippery slope leads the arguer to predict a «particularly horrible outcome [which] is the final event in the sequence and represents something that would very definitely go against goals that are important for the participant...» In fact, most slippery slopes seem to be fear-driven inferences that alert to catastrophic and exaggerated scenarios on the basis of insufficient evidence: e.g., that usage of cannabis is the first step to the use of harder drugs; that immigration leads to the loss of traditional values and eventually to the loss of a national identity; that China’s economic growth will lead to a military supremacy, which in turn will cause the decline of western powers; and so forth. It is difficult not to speculate that, in such cases, the propensity to commit the slippery slope is motivated by the arguer’s fears (or ‘fear-driven’, as Thagard says), exactly as in the case of the jealous husband and in the case of the hypochondriac. Another plausible example would be the slippery slope motivated by xenophobic fears, as in the following example (Pirie 2006: 152): «If we allow French ideas on food to influence us, we’ll soon be eating nothing but snails and garlic and teaching our children to sing the Marseillaise». Some authors hypothesize that, on such occasions, the person’s reasoning is biased by an ‘irrational emotion’, i.e., an emotion that is either based on an irrational belief or not based on any belief at all (De Sousa 1987: 197, Elster 1999: 312). This hypothesis is consistent with the claim that negative illusions may have been beneficial in the evolutionary past, given that the tendency to assume the worst seems to encourage risk avoidance (Andrews & Thomson 2009). It seems plausible, for example, that delusional
jealousy might have increased people’s vigilance against potential rivals, thereby discouraging infidelity between partners. Yet, in modern environments such biases lead to unfair and counter-productive arguments which seem to compromise the individual’s goals, particularly when they are motivated by irrational attitudes – not just jealousy, but excessive jealousy, not just distrust, but unjustified distrust; not just pessimism, but unrealistic pessimism.

5. Critical thinking and argumentative self-regulation

There has been much controversy over whether motivational biases are inherent defects of human reason that tend to undermine the rationality of people’s reasoning (Kahneman 2011, Kunda 1990, Gilovich 1991) or, on the contrary, adaptive mechanisms that tend to maximize decision-making under constraints of time and knowledge (Gigerenzer 2008, McKay and Dennett 2009, Taylor & Brown 1988). On the one hand, it seems plausible that motivational biases may turn out to be beneficial in light of the subject’s environment and goals. From this perspective, Gigerenzer (2008:13) argues, «what appears to be a fallacy can often also be seen as adaptive behavior». Thus, for example, some studies indicate that wishful thinking and self-serving biases tend to enhance people’s motivation, productivity and mood (Taylor & Brown 1988). On the other hand, however, we have seen that biases may also lead to maladaptive responses, such as denial, risk-mismanagement, prejudice, polarization of opinions, and rationalization (for a review, see Dunning et al. 2004). Furthermore, even assuming that biases may be adaptive from a utilitarian standpoint, it is clear that they often compromise the rationality of people’s arguments from a dialectical standpoint. As Johnson and Blair (2006: 191) observe, motivational biases stem from egocentric and emotional attachments which «often result in a failure to recognize another point of view, to see the possibility of an objection to one’s point of view, or to look at an issue from someone else’s point of view». 
Be that as it may, for the purpose of this section it is enough to assume that motivational biases may *sometimes* lead to irrational attitudes in everyday contexts of argumentation. The question to be asked, then, is whether arguers should do something to minimize their irrational tendencies. A number of virtue epistemologists and belief theorists have recently argued that, even though motivational biases are typically unintentional, subjects have the ‘epistemic obligation’ to try to mitigate the effects of biases upon their cognitive processes (Adler 2002, Audi 2008, Engel 2000, Mele 2001a). I have tried to defend elsewhere that this claim is also pertinent in the realm of argumentation theory (Correia 2012), and, more specifically, that the effort to counteract one’s motivational biases should be included in what Johnson (2000: 165) calls the arguers’ «dialectical obligations». For the present purpose, however, I will confine my analysis to the descriptive question that seems to be presupposed by the normative one: Assuming that arguers have the dialectical obligation to debias themselves, how can they achieve this? After all, it only makes sense to suggest that discussants are partly responsible for their irrational thinking if there is something they can do to prevent it.

Perelman and Olbrechts-Tyteca (1969: 119) maintain that biases are unavoidable flaws that are inherent in the process of argumentation: «All argumentation is selective. It chooses the elements and the method of making them present. By doing so it cannot avoid being open to accusations of incompleteness and hence of partiality and tendentiousness». To some extent, at least, the authors are probably right, for it is virtually impossible not to let emotions influence the way we reason in one way or another; and, perhaps for that reason, it is almost a commonplace to acknowledge that the ideal of impartiality is unachievable. Having said this, it is important to bear in mind that «rationality is a matter of degree», as Baron (1988: 36) points out, and that arguers may at least try to minimize the phenomenon of motivated reasoning.

As a matter of fact, it appears that arguers are not condemned to remain the helpless victims of their error tendencies. Even though motivational biases are typically unconscious, there are certain control strategies that arguers can adopt if they wish to counteract the effects of biases upon their reasoning, both at an individual and at a dialogical level. In particular, arguers may adopt a certain number of ‘debiasing strategies’ designed to promote the rationality of their attitudes in a debate.
Before examining some of these strategies, it is worth noting that the very awareness of our biases can perhaps contribute to mitigate their effects. Those who are «open-minded enough to acknowledge the limits of open-mindedness», as Tetlock (2005: 189) elegantly puts it, seem to be in a better position to overcome their cognitive weaknesses and to ensure the rationality of their arguments. For example, a scientist who is aware of the heuristic distortions induced by the confirmation bias may attempt to offset them by forcing herself to examine thoroughly sources that seem to contradict her position. Likewise, arguers who accept the notion that they may be biased without being aware of it are perhaps more likely to remain vigilant against such biases, and perhaps more willing to meet their opponents halfway in the process of solving a difference of opinion. Hence, Thagard (2011: 160) writes, «critical thinking can be improved, one hopes, by increasing awareness of the emotional roots of many inferences».

Second, arguers who wish to make sure that their arguments are fair and balanced may adopt the strategy of «playing the devil’s advocate» (Stuart Mill 1859: 35), i.e., «throw themselves into the mental position of those who think differently from them». According to Johnson (2000: 170) the effort to examine the set of standard objections to our own views constitutes a dialectical obligation that arguers must fulfill even in the absence of an actual opponent\(^5\). This normative requirement seems particularly useful to counteract the confirmation bias and the fallacy of cherry picking, since it exhorts people to contemplate alternative standpoints and sources of information which they spontaneously might tend to neglect. As Larrick (2004: 323) explains, «the strategy is effective because it directly counteracts the basic problem of association-based processes – an overly narrow sample of evidence – by expanding the sample and making it more representative».

Third, biases can more easily be detected if discussants proceed to the ‘analytic reconstruction’ of arguments (Walton 1989b: 170, Eemeren & Grootendorst 2004: 95). By analyzing their discourse into its elementary components, arguers have a better chance to detect hidden biases and to externalize their implicit commitments. Walton (2006: 227-228) stresses that

\(^5\) Johnson (2000: 165) observes that traditional approaches have focused too much on what he calls the ‘illative core’ of arguments, i.e., the set of premises that arguers advance in support of the conclusion, and not enough on the ‘dialectical tier’, i.e., the set of alternative positions and plausible objections that must be addressed.
a «bias may not only be hidden in the emotive words used to make a claim, it may also be hidden because the claim itself is not even stated, only implied by what was not said». Oftentimes, arguers themselves are unaware of their ‘dark-side commitments’ and of the extent to which these can bear on their reasoning. For example, a person’s belief in the existence of God may lead her to bring forward arguments that inadvertently beg the question with regard to matters such as morality and politics. The effort to analyze the components of argumentative discourse seems to contribute to render such commitments explicit, thereby allowing discussants to become aware of their biases.

More generally, arguers may promote critical thinking by improving their argumentative skills. After all, people who have a good understanding of the rules of logic, statistics and argumentation are presumably more likely to detect their own fallacies. Tversky and Kahneman (2008) were able to confirm this hypothesis in a recent replication of the well-known ‘Linda problem’. In the original versions of the experiment (Tversky & Kahneman 1983) the researchers submitted to undergraduates a description of Linda, a fictitious person, as a thirty-one years old activist deeply concerned with issues of discrimination and social justice. Then they asked the participants which of the following possibilities is more likely: (A) Linda is a bank teller, or (B) Linda is a bank teller and is active in the feminist movement? Surprisingly, about 85% to 90% of undergraduates at several major universities chose the second option, thereby transgressing an elementary rule of probabilities: The conjunction of two events cannot be more probable than one of the events alone. Yet, a more recent version of the experiment (Tversky & Kahneman 2008: 120), conducted with graduate students with statistic education, revealed that only 36% committed the fallacy, which seems to indicate that, at least in certain cases, the development of deductive skills can work as a safeguard against systematic errors of intuitive reasoning.\(^6\)

That is not to say that deductive skills alone suffice to ensure the rationality of the arguer’s attitudes in a debate. As Paul (1986: 379) rightly observes, «it is possible to develop extensive skills in argument analysis and construction without ever seriously applying those skills in a self-critical way».

\(^6\) Cohen (1981) would object that subjects with training in logic, probability theory and statistics only appear to be better intuitive reasoners because these are precisely the tasks that they are trained to perform.
to one’s own deepest beliefs, values, and convictions. Some authors suggest that the reasonableness of debates depends just as much, if not more, upon the discussants’ argumentational virtues, that is, on the set of dispositions and character traits that tend to promote good thinking (Aberdein 2010: 169, Cohen 2009: 49). For example, virtues such as open-mindedness, fairness, intellectual honesty, perseverance, diligence and humility seem to offset many of the biasing tendencies examined earlier. The advantage of fostering such virtues is that they tend to form a sort of ‘second nature’ (Montaigne 1967 [1580]: 407, Ryle 1949: 42) which enables people to reason in fair terms almost spontaneously, without a permanent effort to remain impartial.

Finally, discussants have the possibility of adopting what decision-theorists call ‘precommitment strategies’ of self-control, which may be described as self-imposed constraints designed to avoid irrational attitudes (Elster 2007, Loewenstein et al. 2003). In what regards argumentation contexts, such constraints aim at regulating the conditions under which the information is processed and the arguments set out. Thus, a scientist who is about to submit an article on the issue of global warming, but recognizes that her convictions are susceptible to bias her analysis, may commit in advance to several control strategies: for example, verify that she did not overlook any disconfirming evidence (confirmation bias); ask a colleague to try to detect unintentional biases; carefully examine and respond to the standard set of counterarguments; make sure that these have not been misrepresented (straw man argument); and so forth. To be sure, it may not always be easy to adopt self-regulation strategies in everyday debates, given the usual constraints of time and information, but, as Kahneman (2011: 131) points out, «the chance to avoid a costly mistake is sometimes worth the effort».

6. Conclusion

This paper sought to elucidate the problem of how goals and emotions can influence people’s reasoning in everyday debates. By distinguishing between three categories of motivational biases, we were able to see that arguers tend to engage in different forms of fallacious reasoning depending on the type of
motive that underlies their tendentiousness. We have examined some plausible connections between specific types of biases and specific types of fallacies, but many other correlations could in principle be found. Although psychology studies consistently confirm people’s propensity to be biased, motivated fallacies often appear persuasive and difficult to detect because of the arguers’ tendency to rationalize their inconsistencies (Festinger 1957) and because of the ‘illusion of objectivity’ (Kunda 1990: 483) that results from it. Given that these processes tend to occur unconsciously, people’s intentional efforts to observe the rules of argumentation are not always sufficient to prevent them from being biased.

Yet argumentational biases are not inevitable and arguers can (and perhaps ought to) counteract their irrational attitudes by submitting the process of argument-making to indirect strategies of control. The aim of argumentative self-regulation is to make sure that arguers effectively observe the rules of critical discussion in real-life contexts. In my view, this effort must be rooted in a good understanding of the very mechanisms that underlie our error tendencies. As Thagard (2011: 158, 164) suggests, «critical thinking requires a psychological understanding of motivated inference» and «a motivation to use what is known about cognitive and emotional processes to improve inferences about what to believe and what to do». The above-described strategies are mere examples of what arguers can do to promote the rationality of the way they reason, but there may be, in principle, as many debiasing strategies as there are types of motivated reasoning.

References


