Synthese (2009) 167:317–325
Knowledge, Rationality & Action 395–403
DOI 10.1007/s11229-008-9410-y

Logical fallacies as informational shortcuts

Luciano Floridi

Received: 31 January 2008 / Accepted: 11 September 2008 / Published online: 8 October 2008

Abstract The paper argues that the two best known formal logical fallacies, namely denying the antecedent (DA) and affirming the consequent (AC) are not just basic and simple errors, which prove human irrationality, but rather informational shortcuts, which may provide a quick and dirty way of extracting useful information from the environment. DA and AC are shown to be degraded versions of Bayes' theorem, once this is stripped of some of its probabilities. The less the probabilities count, the closer these fallacies become to a reasoning that is not only informationally useful but also logically valid.

Keywords Affirming the consequent · Bayes' theorem · Denying the antecedent · Formal fallacies · Rationality

1 A greener approach to logic

When it comes to managing ontic resources, philosophy may be notoriously prodigal. One only needs to recall Plato's world of ideas, Popper's Third World, or Lewis' possible worlds. But when epistemic and logical resources are in question—the nature and scope of what is knowable and the limits of what is acceptable as formally valid—the problem may rather be some excessive profligacy. All sorts of radical sceptics, rigorous empiricists and dogmatic rationalists seem to be bent on throwing away

L. Floridi (✉)
Research Chair in Philosophy of Information and GPI, School of Humanities,
University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire, AL10 9AB, UK
e-mail: l.floridi@herts.ac.uk

L. Floridi
St Cross College and IEG, University of Oxford, Oxford, UK

far too much; they waste what could be otherwise saved or recycled, both in terms of what we probably do know and in terms of what we might have a decent right to conclude. In this paper, I wish to help to redress this situation. I will argue that logic has been guilty of an "ungreen policy", by considering some formal logical fallacies as absolutely worthless rubbish, only fit for the conceptual junkyard, thus wasting their potential contribution to our information processes. I will suggest that there is a greener and much more reasonable interpretation of such fallacies, which shows that they can be rather useful, if quick and dirty, and probably riskier, ways to gain and manage one's information. Some logical fallacies are not mere mistakes of no value but informational shortcuts that can be epistemically fruitful if carefully managed.

2 What are logical fallacies?

There is no ultimate and complete agreement among logicians or philosophers on a very tight definition of what a logical fallacy is (Hansen and Pinto 1995). This is so, partly because opponents like to use the expression flexibly, in order to throw dialectical mud at each other, and partly because a large variety of errors in reasoning is often diagnosed as a logical fallacy. The result is that there is no fixed taxonomy either. However, for the purpose of this paper, the following might provide a good starting point:

(LF) a logical fallacy $=_{def.}$ any (possibly unnoticed) deductively invalid or erroneous argument with the appearance of validity or a demonstrably false conclusion from plausible reasoning.

If LF (or a working definition sufficiently similar to it) is acceptable, the next step usually consists in distinguishing between two types of LFs (Woods and Walton 1982):

(ILFs) Informal logical fallacies; and
(FLFs) Formal logical fallacies.

ILFs are really mistakes in argumentation, such as begging the question, *ignoratio elenchi*, *argumentum ad hominem*, and so forth. What is wrong with them is something in their semantics or in their strategy, not so much in their essential structure or form. ILFs will make a brief appearance later in the paper, but they are not its subject.

What concerns us here are FLFs. Their problem is entirely morphological. In particular, the two best known schemes of FLFs are denying the antecedent (DA) and affirming the consequent (AC) (Hamblin 1970), and in the rest of this paper we shall concentrate our attention on them. An old and influential tradition has denounced both as just basic and simple errors, weeds in the logical garden, to be eradicated as quickly and thoroughly as possible, wherever they occur. *Pace* Aristotle, this might not be such a good idea.

3 Do FLFs provide any information?

Answers to this question can be grouped under two headings, which are not mutually exclusive:

(a) No. Considered as plain and obvious errors, FLFs are normally dismissed as providing zero information;
(b) Yes. As errors, FLFs show, metatheoretically, that individuals make systematic mistakes and that human reasoning is faulty.

On (a), there is of course a vast literature in logic and critical thinking, but no significant advancement has been made since Aristotle condemned logical fallacies to the dustbin, in his *De Sophisticis Elenchis* (Aristotle 1938).

On (b), the literature is equally vast but less unanimous. In cognitive science, developmental psychology and neuroscience, a wealth of experiments at least since the sixties (e.g. the classic Wason selection task, in Wason 1966) has shown, allegedly, that humans are not so rational, or not as rational as they like to describe themselves. Yet, in this context, there is also a rather annoying tendency to over-simplify and reduce a variety of ways and styles of rational thinking to problem solving, the latter to logical reasoning and this to conditional reasoning. This reduction may be questionable, to say the least, and misleading, as semantic and social interpretations of the Wason selection task seem to show (Griggs and Cox 1982). On the whole, however, Stein (1996) still provides a good survey and a widely accepted conclusion about (b), so let me quote him at length: "According to experiments done over the past few decades, humans make [. . .] significant errors in various realms of reasoning: logical reasoning, probabilistic reasoning, similarity judgements, and risk-assessments to name a few. Together these [. . .] reasoning experiments are taken to show that humans are irrational. [. . .] Some philosophers and psychologists have developed creative and appealing arguments that these experiments are mistaken or misinterpreted because humans must be rational (p. 1) [. . .]. I attempt to show that these arguments fail; cognitive science can and should play a role in determining whether or not humans are rational (p. 2)". Stein might be right, but the story is probably a bit less discouraging, as I will try to show you.

4 Formal fallacies and their explanations

Formal logical fallacies are the dark side of valid inferences: DA is supposed to be a rotten variant of modus ponens (MP), while AC may be seen as a modus tollens (MT) gone wrong. This can more easily be appreciated visually, by looking at Fig. 1.

In a study now classic, Marcus and Rips 1979 presented participants with examples of MP, MT, DA and AC, and asked them to judge each argument as either valid or invalid. Results showed that participants often incorrectly judged MT as invalid, and incorrectly judged DA and AC as valid. There are two standard, positive explanations about why participants accept fallacies as valid inferences (Verschueren et al. 2001).[1]

[1] In the rest of this article I shall not be concerned with the other problem, namely why participants misjudge MT as invalid.

Fig. 1 Valid inference schemes and formal logical fallacies

MP	**DA**	**MT**	**AC**
$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$
φ	$\neg\,\varphi$	$\neg\,\psi$	ψ
ψ	$\neg\,\psi$	$\neg\,\varphi$	φ

Following a stronger informational reading, the first suggests that participants mistake "only if" for "if and only if", treating "if it is a square, then it has four sides" as the same as "if it is water, then it is H_2O". This was already Aristotle's view: "The refutation which depends upon the consequent arises because people suppose that the relation of consequence is convertible. [. . .] since after rain the ground is wet in consequence, we suppose that if the ground is wet, it has been raining; whereas that does not necessarily follow." (*De Sophisticis Elenchis*). The other, based on a weaker informational reading, holds that participants mistake necessity for probability. The two explanations may not be incompatible. To see why, we need to rely on Bayes' theorem.

5 Bayes' theorem

Bayes' theorem calculates the posterior probability of an event A given event B (that is, $\mathrm{P}(A|B)$ on the basis of the prior probability of A (that is, $\mathrm{P}(A)$)). Basically, it tells us what sort of information can be "retrodicted" (backward prediction). The well-known formula is:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|\mathrm{A}) \times \mathrm{P}(A)}{\mathrm{P}(B|A) \times \mathrm{P}(A) + \mathrm{P}(B|A^{\mathrm{c}}) \times \mathrm{P}(A^{\mathrm{c}})}$$

Let us consider a simple example. Suppose Jill receives many emails, but only a few of them (say 2%) are infected by some software virus. She uses a rather reliable antivirus software, which is successful 95% of the time. The latter does not erase her potentially infected emails, but moves them to a special quarantine folder in the e-mail client software, which Jill can check. Jill wonders how often she should check it for good emails. The question she is implicitly asking is: "what is the probability that A (=the email was infected), given the fact that B (=the email was blocked by the antivirus and placed in the quarantine folder) when, on average, 2% of all the emails I receive are actually infected and my antivirus is successful 95% of the time, that is, it provides only 5% false positives?" If the chance that some emails in the quarantine folder might not be infected is very low, then she will check it only occasionally, rather than regularly.

Suppose Jill runs an ideal test on one million emails. The result is shown in Fig. 2. The chance that an email might be infected before being blocked by the antivirus is 2%, but the chance that an email in the quarantine folder is actually infected is roughly 28%. Clearly, Jill should check her folder regularly.

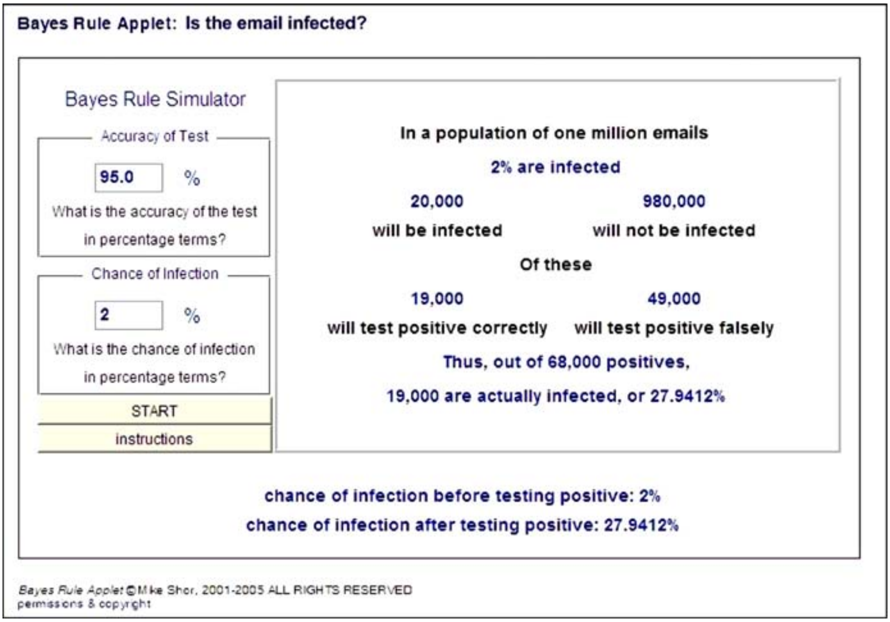

Fig. 2 A simple application of Bayes' theorem, adapted and reproduced with permission from Shor, Mikhael (2005), "Bayes Rule Applet", *Game Theory.net* http://www.gametheory.net/Mike/applets/Bayes/Bayes.html

6 Bayes' theorem and the AC fallacy

Jill is a smart girl. Maggie, a friend of hers, is not. She uses the same antivirus and receives roughly the same number of emails, with approximately the same quantity of infections but, when Jill explains to her that she should check her quarantine folder regularly, she is astonished. For she thought that, if the email was infected, then the antivirus blocked it, and since the quarantine folder contains only emails blocked by the antivirus, then all the emails in it must be infected. More formally, she reasoned that: A $\rightarrow$ B, B $\vdash$ A. Jill explains to Maggie that the previous inference is a typical fallacy (AC), but that she should not feel silly at all. For, consider Bayes' theorem once again. Look at the formula P$(B|A^c)$, where A^c (the absolute complement) is just another notation for $\neg A$. P$(B|A^c)$ indicates the probability that the antivirus blocks the email (B) when the email is not infected (A^c). Suppose we have perfect, infallible antivirus software. This will generate no false positives (no mistakes). But if there are no false positives, that is, if P$(B|A^c) = 0$, then P$(A|B) = 1$ and Bayes' theorem is degraded to a double implication: A $\leftrightarrow$ B, B $\vdash$ A, which is what Maggie perhaps had in mind. On the other hand, if there are some false positives, that is, if P$(B|A^c) > 0$, then P$(A|B) < 1$ and the formula bears a strong family resemblance to the AC fallacy: A $\rightarrow$ B, B $\vdash$ A, which is what Maggie might also have had in mind. Either way, Maggie was taking a shortcut (she disregarded the probabilities) to focus on the sort of information that she could extract from the fact that those emails were in the quarantine folder. And on the wise advice of being safe rather than sorry, she treated all its content as dangerous. The result is that Maggie is ontologically thrifty (she trusts many less items than Jill) by being logically greener (she relies on a reasoning that, although formally fallacious, can still be recycled to provide a quick and dirty way of extracting useful information from her environment). If this is unclear, the reader may try this other example.

Maggie's teacher tells her that, if she does not study enough, then she will fail her exam. Unfortunately, Maggie does fail her exam and the teacher reproaches her for not having studied enough. Maggie has learnt her Bayesian lesson, so she knows that the teacher's reasoning is fallacious. But she also knows that it is fairly accurate, as a shortcut that gets things right most of the time: on average, students who fail their exams have not studied enough. The teacher should have simply sprinkled her inference and then judgement with some "probably" and "most likely" clauses. Basically, the AC fallacy is Bayes' theorem stripped of some of its probabilities. The less the probabilities count, the closer the fallacy is to a reasoning that is logically valid.

The case for the DA fallacy is analogous. Given the same interpretations for A, B, A^c and B^c, Jill calculates, through Bayes' theorem, the probability that the email is not blocked by the antivirus given the fact that it is not infected, that is, $\mathrm{P}(B^c|A^c)$. In this case too, if there are no false positives, that is, if $\mathrm{P}(B|A^c)=0$, then $\mathrm{P}(B^c|A^c)=1$ and the formula is again degraded to a double implication $\mathrm{A} \leftrightarrow \mathrm{B}, \neg\, \mathrm{A} \vdash \neg\, \mathrm{B}$, which is perfectly valid. If there are some false positives, that is, if $\mathrm{P}(B|\neg A)\ >0$, then $\mathrm{P}(B^c|A^c)\ <1$, and the formula bears a strong family resemblance to the DA fallacy: $\mathrm{A} \rightarrow \mathrm{B}, \neg\, \mathrm{A} \vdash \neg\, \mathrm{B}$. The DA fallacy is also Bayes' theorem stripped of some of its probabilities. We are now ready for an overall interpretation.

7 Formal fallacies and their Bayesian interpretation

Consider Fig. 3. In row 1, all four schemes introduce $\varphi \rightarrow \psi$. If we interpret $\varphi \rightarrow \psi$ informationally, it means that there are no false negatives: in our example, if the email is infected, the antivirus blocks it.

MP and MT rely only on this information, so there is no need for probabilities. However, DA and AC also assume (and here is the logical mistake) that there are no false positives (double implication), or that, if there are, they are so improbable as to be disregardable (degraded Bayes' theorem). So DA and AC are Bayesian "quick and dirty" informational shortcuts. When we use them, we bet that $\mathrm{A} \rightarrow \mathrm{B}, \mathrm{B} \vdash \mathrm{A}$ or that $\mathrm{A} \rightarrow \mathrm{B}, \neg\, \mathrm{A} \vdash \neg\, \mathrm{B}$. The bet might be risky (we might be wrong), but it often pays back handsomely in terms of lower amount of informational resources needed

Fig. 3 Valid inferences and formal logical fallacies

	MP	**DA**	**MT**	**AC**
1	$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$	$\varphi \rightarrow \psi$
2	φ / ——— / ψ	$\neg\,\varphi$ / ——— / $\neg\,\psi$	$\neg\,\psi$ / ——— / $\neg\,\varphi$	ψ / ——— / Φ

to reach a conclusion (see the case of the teacher assessing whether Maggie studied enough). Moreover, it is easy to show that the information gain increases (the bets are less risky) the more the following conditions are satisfied:

1. *soundness*: A is true and A $\rightarrow$ B valid;
2. *relevance*: A and B in A $\rightarrow$ B are relevantly related and not independent: the occurrence of A affects the occurrence of B meaningfully (not as in "if Paris is the capital of France then bachelors are unmarried", but more as in "if Paris is the capital of France then Lyon is not the capital of France) (Floridi 2008);
3. *constraints*: we assume $A_1, A_2, \ldots, A_k$ to be mutually exclusive events, whose union is the whole sample space of an experiment, and B to be an event with P(B) $>$ 0.

8 Advantages of the Bayesian interpretation of formal logical fallacies

The "greener" approach to DA and AC just offered presents several advantages.

First, and quite interestingly, parallel results have been recently achieved in the Bayesian analysis of *informal* fallacies: "[...] three classic reasoning fallacies [...] *argumentum ad ignorantiam*, the circular argument or *petitio principii*, and the slippery slope argument [...] match structurally arguments which are widely accepted. This suggests that it is not the form of the arguments as such that is problematic but rather something about the content of those examples with which they are typically justified. This leads to a Bayesian reanalysis of these classic argument forms and a reformulation of the conditions under which they do or do not constitute legitimate forms of argumentation." (Hahn and Oaksford 2006, p. 207) The convergence is remarkable and probably significant. Humans reason in many different ways, only few of which are captured by mathematical logic, as cognitive science is increasingly showing (Chater and Oaksford 2008).

Second, the Bayesian interpretation is consistent with recent work on the interpretation of

1. *abduction* as *inference to the best explanation* (IBE),
2. the AC fallacy as a form of IBE and
3. IBE in Bayesian terms.

It all turns out to be a bit too consistent to be accidental. If we accept the previous Bayesian interpretation of DA and AC, the following quote acquires a very interesting meaning: "It has become common in AI to identify "abduction" with backward MP (i.e. AC), or with backward MP together with constraints on the set of conclusions that are allowable, or with backward MP with syntactic constraints. There is a burden on those who study restricted forms of backward MP to show us the virtues of their particular forms—they need to show us how they are smart. I suggest that they will find that backward MP is smart to the degree that it approximates, or when it is controlled and constrained to approximate, or when it implements, inference to the best explanation." (Josephson 2000, p. 32). In the email example, AC is read as an IBE (inference to best explanation or retrodiction) effort: the email was blocked because it was (probably) infected; whereas DA is read as an IBP (inference to best prediction) effort: the email is not infected so it will (probably) not be blocked.

Third, formal fallacies are interpreted in terms of information-gathering and gain, rather than argumentative and dialectical strength, as in post-Aristotelian logic, or mathematical utility, as in post-Fregean logic. To put it less cryptically, AC and DA are still a disaster, if our goal is to win an argument, because our opponent will not have to be too smart to provide plenty of counterarguments. They are also mathematical calamities, whenever our goal is to provide valid proofs that convey total certainty and faultless conclusions, for their morphology makes them utterly unreliable. But when it comes to extracting information from whatever is at hand and without too much fuss, they can be serviceable. This point is related to the last advantage, which is best left for the conclusion.

9 Conclusion: rationality regained

Anyone who has taught logic to undergraduates knows that logical fallacies are very common and hard to eradicate. Convincing normal, intelligent, decently educated, young minds, who have successfully negotiated very complex cognitive tasks for almost two decades, that their most basic ways of reasoning produce only worthless rubbish takes a while and some effort. It is also an unrewarding and tedious task, which is not always successful. A common result is the development of a schizophrenic attitude: life is life and logic has got little to do with it. Otherwise put: when in the logic class do as the teacher does, but then, at the pub, you can go back to your old habits and infer that if someone has had too many drinks and drives, that person will have an accident and since she did have an accident she must have been drunk. A greener approach to the way in which we treat fallacies has the advantage of making people see not only the limits of similar ways of reasoning, but also their value, provided one is aware of the sort of epistemic risks one is running. In a way, AC and DA are very powerful tools—inferentially, they are the equivalent of shoot first and ask questions later—developed by embedded and embodied agents to cope, as quickly and successfully as possible, with a hostile environment. The fact that we have had the leisure, in the past few millennia, to show how badly logical fallacies serve us when the inferential game becomes more refined and subtle, either dialectically or mathematically, should not prevent us from building a healthier relation between life and logic and recognise their potential utility. The sooner we start recycling them the better.[2]

[2] This paper was first presented at the First Workshop on the Philosophy of Information and Logic, University of Oxford 3–4 November 2007. I am indebted to Sebastian Sequoiah-Grayson for his kind encouragement and his patience with the editorial process and I am very grateful to the participants for the discussion during and after the meeting, for it generated several improvements. A second version was the subject of an invited talk at the yearly meeting, in Cumberland Lodge, of the Department of Philosophy of the University of Hertfordshire (15–17 February 2008). I wish to thank Brendan Larvor for that opportunity and all the attendees for the feedback I received on that occasion. The students in my logic class deserve more than a passing acknowledgement, for they have been the subjects of several conceptual experiments involving the topics covered in this paper, during which I am not sure nobody was mentally harmed. Finally, a special thank to Mikhael Shor, for having made the applet available online and for allowing me to use it and reproduce the picture in Fig. 2, and to Joanna Gilles for copyediting the text.

References

Aristotle, (1938). *De Sophisticis Elenchis. On sophistical refutations* (trans: Forster, E. S.). On with the Cosmos/by Furley, D. J. Cambridge, MA: Harvard University Press.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.

Floridi, L. (2008). Understanding epistemic relevance. *Erkenntnis, 69*(1), 69–92.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*, 407–420.

Hahn, U., & Oaksford, M. (2006). A Bayesian approach to informal argument fallacies. *Synthese, 152*(2), 207–236.

Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

Hansen, H. V., & Pinto, R. C. (Ed.) (1995). *Fallacies: Classical and contemporary readings*. University Park, PA: Pennsylvania State University Press.

Josephson, J. R. (2000). Smart inductive generalizations are abductions In: P. A. Flach & A. C. Kakas (Eds.), *Abduction and induction: Essays on their relation and integration*. Dordrecht, Boston: Kluwer.

Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior, 18*, 199–224.

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Clarendon Press.

Verschueren, N., Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001). Why do participants draw non-valid inferences in conditional reasoning? *Current Psychology Letters: Behaviour, Brain, & Cognition, 6*, 57–70.

Wason, P. C. (1966). Reasoning. In: B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth: Penguin.

Woods, J., & Walton, D. N. (1982). *Argument : The logic of the fallacies*. Toronto: McGraw-Hill Ryerson.