

May 18th, 9:00 AM - May 21st, 5:00 PM

The polysemy of 'fallacy' —or 'bias', for that matter

Frank Zenker
Lund University

Follow this and additional works at: <https://scholar.uwindsor.ca/ossaarchive>

 Part of the [Cognitive Psychology Commons](#), [Logic and Foundations of Mathematics Commons](#), [Other Law Commons](#), [Other Rhetoric and Composition Commons](#), [Philosophy of Language Commons](#), and the [Social Psychology Commons](#)

Zenker, Frank, "The polysemy of 'fallacy' —or 'bias', for that matter" (2016). *OSSA Conference Archive*. 145.
<https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/145>

This Paper is brought to you for free and open access by the Department of Philosophy at Scholarship at UWindsor. It has been accepted for inclusion in OSSA Conference Archive by an authorized conference organizer of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.

The Polysemy of ‘Fallacy’—Or ‘Bias’, For That Matter

FRANK ZENKER

Slovak Academy of Sciences, Philosophy, Klemensova 19, 81364 Bratislava, Slovakia

Konstanz University, Philosophy, Box D9, 78457 Konstanz, Germany

Lund University, Philosophy & Cognitive Science, Box 192, 22100 Lund, Sweden

frank.zenker@fil.lu.se

Abstract: Starting with a brief overview of current usages (Sect. 2), this paper offers some constituents of a use-based analysis of ‘fallacy’, listing 16 conditions that have, for the most part implicitly, been discussed in the literature (Sect. 3). Our thesis is that at least three *related* conceptions of ‘fallacy’ can be identified. The 16 conditions thus serve to “carve out” a semantic core and to distinguish three core-specifications. As our discussion suggests, these specifications can be related to three normative positions in the philosophy of human reasoning: the meliorist, the apologist, and the panglossian (Sect. 4). Seeking to make these conditions available for scholarly discussion, this analysis-*sketch* should not be viewed as final or exhaustive.

Keywords: bias, fallacy, meaning analysis

1. Introduction

In research areas as diverse as cognitive science and psychology, logic and mathematics, law and argumentation theory, communication studies and contemporary rhetoric, among others, research nowadays mostly proceeds systematically. At the interface of these areas, scholars as well as scientists nevertheless converge on ‘fallacy’ as the historically given descriptor for various ways in which human behavior deviates from a normative standard. Since recently, moreover, cognitive and social biases are viewed as being intimately related to fallacies, as their subject-internal partial causes. As is standard when research bridges disciplinary boundaries, however, technical terms often lack good definitions.

Starting with a brief overview of current usage (Sect. 2), this paper offers some constituents of a use-based analysis of ‘fallacy’, by listing 16 conditions that have, for the most part implicitly, been discussed in the literature (Sect. 3). Our thesis is that at least three *related* conceptions of fallacy can be identified. The 16 conditions serve to “carve out” a semantic core of fallacy, which appears to be shared across the above research areas, and to distinguish three core-specifications. Our discussion suggests that some of these conditions relate to the three main normative positions in the philosophy of human reasoning: the meliorist, the apologist, and the panglossian one (Sect. 4). Our main purpose is to make these conditions available for discussion; the analysis *sketch* that we provide is non-final, and will probably remain non-exhaustive.

2. Fallacies and their evaluation

We agree with Hans Hansen (2015) that

[s]ome general definition of ‘fallacy’ is wanted but the desire is frustrated because there is disagreement about the identity of fallacies. Are they inferential, logical, epistemic or dialectical mistakes? Some authors insist they are all of one kind

[...]. [But t]here are reasons to think that all the fallacies do not easily fit into one category. (4.1)

Indeed, we shall suggest that at least three categories are needed. This section provides an overview of three related but distinct uses of the term ‘fallacy’ that denote (i) an inferential error, (ii) a misaligned or otherwise “faulty” task-response, and (iii) a suboptimal discourse contribution (Sect. 2.1). We moreover point to the link between reasoning and argumentation, and briefly discuss evaluative standards (2.2).

2.1. Current usage of ‘fallacy’

“Fallacies are logical errors, and sophistical refutations are sophistical because they contain fallacies,” or so Hans Hansen (2007) summarizes what he offers as the Aristotelian understanding of the fallacies, echoing the view of fallacies as defective inferences. Hansen (2004) moreover cites the claim that “[a] fallacy is an argument that seems to be valid but is not so” (Hamblin, 1970) as ‘the standard definition of fallacy’—which itself appears to owe its status to Rob Grootendorst (1986) rather than Charles Hamblin (see Hansen, 2002, p. 134). In historical perspective, however, Hansen also shows that “the most common [current] conception of fallacy” (Johnson, 1987, p. 240, as cited in Hansen, 2002) has *not* in fact been a widely shared standard.

Hansen’s recent lexicon entry on the fallacies, finally, begins by distinguishing “[t]wo competing conceptions of fallacies [according to which] they are false but popular beliefs and that they are deceptively *bad* arguments” [my emphasis] (Hansen, 2015). The ‘false belief-view of fallacy’ may seem to pertain *only* to the management of well-formed doxastic states, while the ‘defective premise-view of fallacy’—by citing an inferential pattern, or structure—pertains to transitioning between doxastic states. But one need not consider these views as competing with each other. For as Jonathan Adler (1997) has pointed out:

The proper notion of structure or form is much broader than the notion of logical structure or form. Whenever we distinguish in an inference pattern between constant elements and variables, open to substitution, where the inference turns on the pattern of these constant elements, and not the substitutions for the variables, we are specifying a structure or form (Brandom, 1988). Additionally, the pattern must yield a rich set of inferences. On this conception, criticizing some arguments for the falsity of a premise, when it expresses a rich, structural pattern, does constitute the finding of a defect in form. (p. 335)

It should thus be clear that the ‘defective premise-view of fallacy’ remains parasitic on something being “wrong” with the larger cognitive structure of which coming to form a false premise is a part. We return to the notion of structure below.

Regarding the “deceptively bad argument”-view of ‘fallacy, Barth and Krabbe’s (1982) formal dialectics, for instance, locates “badness” in the inability of some argumentative moves to be generated by the production rules of good arguments. Relatedly, Pragma-dialecticians (van Eemeren & Grootendorst, 2004) find “badness” in the (metaphorical) distance that separates actual argumentation from an ideal(ized) model for the activity type of a critical discussion, and cite mishaps in the arguer’s management of the goal-conflict—between the rhetorical goal of

winning the dispute and the dialectical goal of appearing reasonable to an audience—as a functional *explanans* why fallacies arise in discourse. Again relatedly, Douglas Walton (1995) finds “badness” in illicit dialectical shifts across various dialogue types, and has more recently considered heuristic modes of reasoning as (partial) causes of fallacies (Walton, 2010).

Such broadly pragmatic accounts of fallacy have dominated the research front in argumentation studies for the last 40 or so years. By far most research efforts have gone into studying individual fallacies as well as the argument schemes whose “evil cousin” a given fallacy may be said to be; with few exceptions, little empirical work has occurred. Throughout this period, some have raised claims such as ‘there are no fallacies’ (cf. Govier, 1983), or ‘no theory of fallacies can be had’ (e.g., Massey, 1981), or ‘all fallacies are relevance problems’ (e.g., Powers, 1995).

We cannot hope to do full justice to the breadth of this theoretical discussion, but rather cite one particularly “deviant” theoretical voice: Maurice Finocchiaro (1981) had early on observed that (too many) fallacies seem to occur not so much in real-life, but in textbooks!¹ Though proceeding not on considerations of frequency alone, John Woods (2013) has developed Finocchiaro’s observation—which is problematic since hard data for the relevant frequencies are, well, hard to come by (see below)—into a full-blown dismissal of the entire fallacy tradition. Woods’ strategy is to reject, as inappropriate, the goal that is normally—i.e., under the standard scholarly interpretation of a relevant act of arguing—ascribed to an arguer when she is charged with fallacy. Instead, he postulates a goal that committing the “fallacy” *serves* for boundedly rational agents (Simon, 1982) who interact in natural contexts. This occurs in an attempt to *naturalize* the fallacies, i.e., to “arrive” at them as a partial function of non-idealized constraints on human cognition, on one hand, and on the environment, or context, on the other.

Such developments notwithstanding, most logicians, psychologists, epistemologists, cognitive scientists, and lawyers have continued² to view fallacies not primarily as pragmatic phenomena, but have more narrowly (and more traditionally) conceived of them as bad *inferences*. Perhaps rightly so in view of a recent general critique by Boudry, Pigliucci & Paglieri (2015), who broadly argue that fallacies-as-dialectical-problems “supervene,” as it were, on fallacies-as-epistemic or inferential problems. This latter view is particularly prominent among research in the heuristics and biases tradition (e.g., Tversky & Kahneman, 1974), where ‘bias’ most plausibly denotes the subjective internal (causal) antecedent to a behavioral output that is found problematic, or wanting, given a normative standard, while ‘heuristic’ denotes a search- or decision-making procedure that abbreviates the deployment of a fuller and more effortful standard of inferencing (Gigerenzer & Brighton, 2009; Hahn & Corner, 2013).

The ‘bad inference-view of fallacy’ thus views argument as the externalized and socialized aspect of otherwise internal and individual reasoning. ‘Inference’ here denotes a *transition between information states*, irrespective of how these are mentally represented, e.g.,

¹ Despite this claim, Finocchiaro distinguished six ways for arguments to be fallacious. With (1) *formal fallaciousness* the conclusion fails to follow (deductively) validly from premises; with (2) *explanatory fallaciousness* two or more rivaling conclusions follow with at most equal certainty from premises; (3) with *presuppositional fallaciousness* the argument depends on some false presupposition; (4) with *positive fallaciousness* the premises, once enriched by others that are taken as true, support a diverging conclusion; (5) *semantical fallaciousness* pivots on the ambiguity of terms; (6) with *persuasive fallaciousness* the conclusion is the same as one of the premises (see Hansen, 2015).

² The term ‘continued’ potentially misleads, of course, for the ‘bad inference-view of fallacy’ had itself come about as a modern intellectual development—seemingly as a “separatist” development, too—in response to a classic tradition that, unlike today, did not view logic and dialectic as separate fields (see Hintikka & Spade, 2012).

propositionally, model-theoretically, or otherwise (see Adler, 2008). This establishes a close link between reasoning, on one hand, and natural language argumentation, on the other, for instance by understanding argument as an invitation, directed to a hearer, to accept the speaker's corresponding episode of reasoning, to modify, or to criticize it.

2.2. Evaluative standards

On this understanding of fallacy as *incorrect* inference, its evaluation—this being what provides the grounds for raising a fallacy charge—must invoke either a deductive or an inductive standard of validity. With the deductive standard, the specific content of a given transition between information states is normally abstracted away from, only its structural aspect being evaluated. This is necessary, of course, for any two (or more) *distinct* semantic contents must be understood as saturations of the *same* inferential structure before they can be unproblematically treated alike.

Though various “new” logics have been developed particularly during the second half of the 20th century, that *seem* to offer new standards of validity, it remains a rather well-defensible view (e.g., Spohn, 2012) that properly conceiving of non-deductive inference invariably renders it as *inductive*, i.e., any such inference features a conclusion—i.e., a terminal information state—whose content reaches beyond the content offered as premises, i.e., the initial information state.³ In deductive inference, by contrast, the terminal state features *at most the same* information as the initial state. Furthermore, the validity of a deductive inference can in principle be evaluated mechanically—*pace* issues that arise from introducing infinities—and in disrespect of considerations that can render deductive inference *sound*, which pivot on standards for deploying the predicate ‘is true’ to states of information.

By contrast, analysts who evaluate the validity of inductive inferences normally cannot *solely* rely on a similar mechanical procedure, nor can they *completely* abstract away from the semantic content of the inference. To see this clearly, consider as basic a transition from α to β (i.e., ‘alpha’; ‘beta’), where the information content of β exceeds the content of α , irrespective of the modality of α or β (e.g., metaphysical, deontic, epistemic, etc.). Call this excess information Δ (‘delta’), and now ask: when has β been *validly inferred* from α , that is, under which conditions has Δ been *validly established* on the basis of α ?

Notice that invoking principles such as ‘ α must sufficiently raise the (prior) probability of β ’, or ‘ α must provide information that is relevant in order to support β ’ can at best amount to pseudo-answers. Indeed, these as well as similar principles merely reformulate the initial question in terms of inductive logical validity. Since the evaluation of content-enlarging reasoning thus cannot proceed *solely* on the basis of structural aspects, analysts must ever recur to the specific content of an inference. Moreover, validity judgments for inductive inferences retain a modicum of uncertainty, and in special cases genuine risk. Such judgments therefore remain open to revision in view of new information, or when re-evaluating the importance of old information. Put differently, a genuine inductive transition from α to β ever remains objectionable, in the sense that information *other than* β remains compatible with α , and so may *also* be inferred on the basis of α .

Upon broadening one’s understanding of ‘reasoning’ to include features of the interaction by which speakers offer invitations to infer β from α , the full range of pragmatic

³ This also holds for so-called third-way reasoning (Woods, 2013), which is a yet-to-developed logic that would be faithful to how humans in fact reason.

considerations—ranging from interpretational and politeness to participatory and process considerations—can be appealed to also in the evaluation of deductive and inductive inferences. This again comes at the cost of inviting in particulars that having abstracting away from *would* provide the ground for considering the evaluation a general one. But at times doing so simply happens to be necessary! To give but one example, it is clear that inferring α from α itself amounts to the deductively valid transition ‘ $\alpha \therefore \alpha$ ’ (with ‘ \therefore ’ denoting ‘therefore’), also known as circular reasoning. Whenever such reasoning is in fact problematic, then, its evaluation must invoke particular considerations that apply not in addition to, but are indeed *distinct from* deductive validity. We return to this below.

On this perhaps all-too-brief background, we now turn to the specifics of how fallacies are presently conceived of in the scholarly literature. The next section offers constituents of a meaning-based analysis of ‘fallacy’, on the understanding that ‘fallacy’ is polysemous, for current usage aligns with various specifications of these constituents.

3. Towards a used-based analysis of ‘fallacy’

This section presents conditions that arise in attempting to do justice to the meaning of ‘fallacy’. Section 4 offers a proposal on which conditions collect into a distinct fallacy concept.

3.1. Target usage

We seek to account for uses of the term ‘fallacy’ such as:

“By performing action A in context C, subject SUBJ commits the fallacy F, *if ...*”

Future work will need to replace ‘if’ by ‘if, and only if’ to finalize the analysis. Notice, too, that ‘subject’ can in principle be replaced by ‘subjects’, or ‘group’, to account for cases of genuine social interaction. For one need not think of the relevant action, or behavior, to which a fallacy charge is “pinned” *as a single person action*. Rather, ‘fallacy’ may refer to discursive exchanges that exceed a single turn, and so can comprise actions or behavior brought about or sustained by more than one agent.

3.2. Sixteen conditions

We go on by presenting 16 conditions, and offer a brief commentary to motivate how they are formulated. Again, this list is not claimed to be exhaustive; it should be read to state *use-conditions* rather than *truth-conditions*.

1. *Systematic*: Action A is of the non-arbitrary pattern P.

Being of some *pattern*, or having some *structure*, is a basic concern, if a fallacy shall be systematic. The ‘false belief-conception’ of fallacy was above seen to *seemingly* run counter to this idea (Sect. 2). So in retaining the idea of being patterned, we suggest that what caused a fallacy-as-false-belief (e.g., a perception or an inference) might be viewed as resulting from something that is itself patterned.

2. *Reconstructable*: P maps onto a transition between information states α and β .

If fallacies instantiate patterns, these must somehow be represented. Rather than refer to propositions, as is standard, we use ‘information state’ and so leave open details that a particular theory of representation might fill in. ‘Information state’ being a general concept, it can of course be “fleshed out” by ‘proposition’ or ‘mental model’, for instance. Similarly, the mechanics of the transition may be further described, for instance algorithmically (in some logic) or biologically (by referencing the neuronal or some other bodily basis) or functionally (by referencing the organism’s goals).

3. *Reasoned*: α stands in the ‘is a reason for-relation’ to β —in form: ‘ $\alpha \therefore \beta$ ’.

Fallacies obviously have to do with providing reasons or support for assertions or claims that are normally distinct from the reasons themselves. On the weakest reading of ‘being a reason for’, β occurs *in response to* α having occurred. So α is a temporal (and possibly also a causal) antecedent, while β is a temporal (or causal) consequent. We merely intend the weak reading here. On the strongest reading, by contrast, ‘being a reason for’ implies that α is *intendedly* also a normatively adequate reason for β . So someone who transitions from α to β would be committed that β is an *apt* response given α . We prefer to formulate aptness as a distinct condition, below.

4. *Constrained*: Context C fulfils the application-conditions of theory T.

‘Context’ here comprises what *surrounds* the relevant transition, and can without loss of meaning be replaced by ‘environment’. To give a particularly clear example, one may distinguish *risky* from *uncertain* environments (or contexts). In the former context, the set of possible future outcomes is known to be finite, as when rolling a dice. In the latter context, by contrast, the set of possible outcomes is either not known to be finite, or is known to be infinite. Consequently, whenever T is the theory of rational choice, then to be properly applied T *requires*—as one of its application-conditions—a risky context. Notice that T, which may also be saturated by ‘logic’, among others, need not be a formal account, but might be saturated by R, for ‘rhetoric’. This should become clearer below.

5. *Decidable*: β follows from α in virtue of ‘ $\alpha \rightarrow \beta$ ’ being a theorem in T.

Condition (5), as formulated here, is properly a condition for *correct* reasoning. As a condition that features in specifying the meaning of ‘fallacy’, therefore, this condition will need to be *violated* whenever a fallacy occurs. Moreover, ‘ $\alpha \rightarrow \beta$ ’ here stands for what in condition (2) was called ‘transition’, and ‘follows from’ denotes one of various consequence relations a given theory offers as valid, i.e., as a theorem. That *some* theory or a broadly theoretical account *must* be invoked when specifying ‘fallacy’ becomes clearer when turning to the next condition. After all, the relevant theory will need to deliver the normative standard relative to which an action is fallacious, if it is.

6. *Accuracy*: ‘ $\alpha \therefore \beta$ ’ is accurate *qua* T providing the normative standard S.

Also condition (6) is formulated as one for *correct* reasoning. It states that transitioning from α to β is accurate *because* the normative standard S, provided by theory T, says so. Once again, (6) must be *false* for a fallacy to in fact occur. The important insight, here, is that S must have antecedently been coordinated to a given transition from α to β .

7. *Quality*: P is “bad” in context C.

The predicate ‘is bad’, of course, should be specified to bring out the particular problem or issue arising when a fallacy occurs. But it does not go without saying that some pattern is *in fact* “bad” in some context. For example, consider that, relative to the maximization of choice utility, it would in specific contexts be considered bad to satisfize rather than optimize on one’s preferences. In other contexts, however, satisfizing would be “OK,” e.g., whenever optimization requires more resources than are available. Notice, too, that a narrow understanding of fallacy will normally not spell out (7); rather, should whatever *is* considered a fallacy (relative to a theory) occur, then this occurrence is itself considered a “bad.” Contrast this by saturating T with a *rhetorical* rather than a *logical* account, however, and it should become clearer that for instance *your* offering the logical fallacy F2, at turn T2, in response to *my* having earlier offered the logical fallacy F1, at turn T1, can in fact be “just right,” rather than “bad.”

8. *Goal-directed*: A occurs in response to task TA of reaching goal G in manner M.

Besides having previously invoked contexts, condition (8) now invokes tasks, goals, and manners, to reference what the relevant action is triggered by, and what it seeks to bring about. This condition is indispensable whenever fallacies shall be said to arise in experimental settings (e.g., in cognitive science), but also applies to those natural settings whose “raison d’etre” is to solve broadly pragmatic coordination problems (e.g., deciding how scarce resources shall be spent). Though one may understand the relevant task to consist in reaching a particular goal in a specific manner, it pays to keep these constituents analytically apart, insofar as the fallacy may consist in having reached the goal in some broadly “bad” manner.

9. *Aptness*: ‘ $\alpha \therefore \beta$ ’ is an apt response in view of T, G, and C.

Condition (9) normally arises specifically in treatments of circular reasoning (Woods, 2013), and contrasts directly with the accuracy condition (6), above. After all, circular reasoning simply cannot be captured as a problem of accuracy, since inferring α from α itself—indeed, ‘copying α ’ may be a better description—is an accurate move, of course. However, *in fact* reasoning ‘ $\alpha \therefore \alpha$ ’ is rarely, if ever, an *apt* move in contexts where reasons for α are required. So what is apt in some context becomes a (partial) function of the context, the goal, and the task.

10. *Understood*: SUBJ correctly interprets TA and G.

Once tasks and with them goals (which to achieve in some manner can be a proper part of a task) have been introduced, (10) can be offered as a condition *which to violate* can to some extent explain why a fallacy arises, if it does. This condition reflects an observation commonly made, e.g., in cognitive science, when reasoning tasks are assigned in experimental contexts. Participants here often—indeed, too often (see Evans, 2002)—fail to fully understand the task

that is put to them. So the experimenter and her subjects fail to successfully coordinate on the *same* task. The Linda problem is a prime example of this (Kahneman, 2011). In natural contexts, too, one may explain why fallacies occur by postulating failure to coordinate on a shared goal, for instance when one participant treats the interaction as a problem-solving discourse, while the other primarily intends to *share*, rather than solve, a problem. This, of course, presupposes that what is a fallacy (or is otherwise problematic) in context C need not be so in context C*. This is perhaps particularly clear for a bargaining interaction where, epistemically, dialectically, and rhetorically, almost anything goes, as long as the ultimate goal—reaching agreement—is thereby not made impossible. Notice that because *dialogue types* (e.g., Walton, 1995) and *activity types* (van Eemeren, 2010) can analytically be distinguished on the basis of their attendant tasks-manner-goal details, their specification is tantamount to specifying a dialogue type or activity. Also compare Woods’ (2013) strategy (mentioned in Sect. 2) of reinterpreting participant’s goals in some context, such that an argumentative move that would for instance instantiate an *ad hominem* fallacy relative to the goal of securing, say, the *truth* of a conclusion need not be so viewed, if the goal is in fact to secure *assurance*, or additional evidence, that one’s interlocutor is a trustworthy source of information.

11. *Within ability*: SUBJ potentially commands the normative standard S.

Condition (11) reflects a stance according to which agents who are neither factually, nor potentially, in possession of a particular normative standard cannot commit a relevant fallacy that *not* deploying this very standard consists in. This condition, of course, is open to disagreement. The condition perhaps becomes more palatable, however, in view of the proverb: “If all you got is a hammer, then everything looks like a nail.” For it would at least appear questionable were one to treat alike cases where agents *are known* to potentially command a certain normative standard (e.g., because they had been trained in it, and had already once displayed its correct application), on one hand, and cases where agents *are not known* to command that standard, or may reasonably be assumed to lack the standard, on the other. This is perhaps particularly clear if the correct deployment of a standard *requires* having received formal instruction—as is the case, for instance, with probability theory, especially with handling single case- rather than frequency-formats.

12. *Frequency*: Among all patterns, P’s occasioned frequency is significant.

We had above (Sect. 2) alluded to frequency considerations regarding patterns of reasoning that are deemed fallacious, there citing Finocchiaro’s observation according to which the fallacies are rather infrequent—outside textbooks, that is. The notion of ‘occasioned frequency’ stems from Woods’ account (2013). In the sense of a numerical ratio, it refers to the number of occasion-tokens where some fallacy *could* be committed over the number of occasion-tokens where that fallacy *in fact* occurs. Woods’ thesis—that this ratio is strikingly low for most of what the tradition has identified as fallacies—is clearly an empirical claim. But its operationalization poses challenges. For it requires that the occasion-tokens in the numerator can be reliably identified. Rather than endorse this (or any other) condition, we simply list it here.

13. *Attractiveness*: P is not typically inaccurate or inapt in C^* contexts.

Condition 13, also mentioned by Woods (2013)—though not in a formulation that mentions contexts—provides another way of explaining why fallacies arise, namely if a given pattern of inference that is problematic in context C is in fact “OK” in a different context C^* . The broader idea (on which see also condition 16, below) is this: if there *are* distinct contexts—as appears to be case—then agents will need to identify the context they are in. But insofar as agents are less than perfectly reliable in this identification—as is plausibly the case, too—then one may account for the fact that pattern P is deployed in context C , by invoking the context C^* and the fact that P is “OK” in C^* .

14. *Incorrigible*: P 's rate of recidivism is high.

Condition 14 again comes from Woods' account (2013), and appears not to be shared widely (yet). This condition is far less “innocent” than might seem. ‘Recidivism’ refers to the rate at which a relevant pattern continues to occur *although* a pattern-deploying agent has been instructed that this same pattern *is* problematic, in any context even. For Woods, the condition is in place because some (few) fallacies—namely those whose rate of recidivism is high, and which cannot be explained away by suitably reinterpreting an agent's goal in a particular context—indicate something of theoretical interest about the “flawed” aspects of our cognitive makeup, for such flaws would be shared among (almost) all humans. These flaws, and only these, Woods retains as fallacies—indeed as *errors*—since they are committed in “virtue” of species-general cognition.

15. *Size*: The distance $D(\beta, \beta^*)$ is significant, where β^* does not follow from α .

When β *does* follow from α —i.e., whenever ‘ $\alpha \therefore \beta$ ’ *is* valid given T —then ‘ $\alpha \therefore \beta^*$ ’ (where β^* differs in informational content from β) *cannot* be valid in the same sense. Clearly, however, the information represented by β and β^* need not therefore be *very* different; rather, what separates β from β^* might be minimal, i.e., *insignificant* on a relevant distance measure, whatever the measure's details. Hence, one may consider the fallaciousness of a reasoning pattern as a partial function of $D(\beta, \beta^*)$.

16. *Similarity*: Contexts C and C^* are relevantly similar.

As indicated above, as long as there is more than one single context—which to assume is plausible—it is also plausible to suppose that a context C bears out some similarity to a context C^* . If so, then this similarity can be cited to explain why C may be mistaken for C^* . So if pattern P is “OK” in context C , but “not-OK” in context C^* , it becomes natural to “blame” the similarity between C and C^* in order to motivate *why* P was deployed (also see 13).

This concludes the list of meaning-constituents. Once again, this list is probably non-exhaustive. To suggest additional conditions, therefore, is not so much an act of criticism but of cooperation. Of course, various among these conditions could have been drawn together; the next section seeks to motivate why we have kept them distinct.

4. Three fallacy concepts and three positions on human rationality

4.1. Core conditions; narrow, extended, and pragmatic-dialogical sense of ‘fallacy’

Some of the conditions presented in Sect. 3, we submit, apply to *any* fallacy concept in the sense of providing a conceptual core. Other conditions can be viewed as specifications of this core, which remain distinct to particular variants of ‘fallacy’. A plausible candidate for the conceptual core of ‘fallacy’, we submit, is the logical conjunction of conditions 1 (“Systematic”), 2 (“Reconstructable”) and 3 (“Reasoned”). This core *underspecifies* ‘fallacy’—indeed, the same core can also be specified to yield the semantics of correct reasoning. (This being so is more of an aesthetic choice.)

As for core-specification, the three fallacy-concepts (see Sect. 2) were these:

- (i) *narrow*—a transition between information states that is invalid relative to some theory;
- (ii) *extended*—a reasoning process or behavioral output that is misaligned to a normative standard that has been coordinated to a reasoning-task;
- (iii) *pragmatic-dialectical*—a contingent discourse move that fails to optimize on participants’ goals given the norms of a dialogue type and context.

(i) appears to be of classical heritage, but has been adapted to validity standards other than deductive logic, notably probability-theory and decision-theory. On the narrow understanding, of course, a fallacy need not depend on a human agent; a computer or a robot might as well be said to commit a fallacy. Extending the narrow concept, (ii) denotes external human behavior displayed in an experimental or natural context, but also its internal (partial) causes. Here, the normative standards against which fallacies are diagnosed operationalize the validity-norms of a given theory named in (i). Finally, (iii) *seemingly* breaks with the narrow and the extended concept, denoting a suboptimal discourse move given arguers’ goals and contexts. As we argue below, however, it is more fruitful to view (iii) as a specification of (ii).

Reflecting on the list of conditions in Sect. 3, consider now that, in addition to the conceptual core (i.e., conditions 1 to 3), conditions 4 (“Constrained”), 5 (“Decidable”) and 6 (“Accuracy”) together provide for the meaning of fallacy in the narrow sense (aka. ‘inferential error’). Particularly 6, if false, accounts for fallacious reasoning, or acts of arguing that invite to accept such reasoning, and, if true, 6 accounts for its accurate variant. This narrow sense of fallacy is close to—but need *not* be viewed as being identical with—a logician’s standard notion of ‘fallacy’. After all, logicians do normally not separately consider, but rather presuppose, 4. Moreover, logicians tend to infer 7 (“Quality”), or somehow take 7 for granted, provided 1-5 are instantiated and 6 is violated, rather than demand, or specify, 7 additionally.

In condition 5 (“Decidability”), as we saw, the specification of the relevant theory, T, may cite various logics (deductive and inductive ones, that is), but may also cite dialectical or rhetorical accounts. So ‘theorem’ can be replaced by ‘normative consideration’ or ‘guiding principle’, etc., depending on the theory that provides the normative standard.

If condition 7 (“Quality”) is false, e.g., because condition 16 (“Size”) is false, then “fast and frugal” heuristics may be postulated to have replaced executing the more effortful normative standard that condition 5 (“Accuracy”) cites. In condition 8 (“Goal-directed”), the task may be an experimental one, e.g., Wason’s selection task or similar (see Oaksford & Chater, 1994, 2003),

but TA, G, M (for ‘task, goal, manner’) can also refer to the resolution of a difference of opinion on the merits *sans* unnecessary impediments, e.g., in Pragma-dialectics (van Eemeren & Grootendorst, 2004), thus specifying what is “bad” about fallacies.

Condition 9 (“Aptness”) reflects the distinction ‘consequence having’ (see condition 6) vs. ‘consequence drawing’ for contexts of reasoning and argumentation (Woods, 2013). Even logicians will *need* to endorse it in order to identify *petitio principii* as a problem, for instance, for this remains an issue of consequence drawing. Aptness thus provides one way of specifying what aspect of some context might keep a transition between information states that *is* accurate from being an *apt* act of reasoning, or argumentation.

We suggest that, in addition to the core-conditions (1-3), suitable specifications of conditions 4-6, on one hand, and 7-9, on the other, suffice to account for the extended sense of fallacy that is typical for experimental research in psychology and cognitive science. To reach the pragmatic-dialectical sense of ‘fallacy’, particularly conditions 7 (“badness”), 8 (“goal-directedness”), and 9 (“aptness”) need to be suitably specified. If this is accepted, then the pragmatic-dialectical sense of fallacy can plausibly be viewed as a variant of the extended sense of fallacy.

On this background, we now relate the remaining conditions (10-16) to the three basic views on human rationality mentioned in Sect. 1, namely: meliorist, apologist, and panglossian.

4.2. Three views on human rationality

Panglossians deny any meaningful distance between actual and normatively correct human reasoning, and can seemingly avoid charges of irrationalism from the outset. In contrast, meliorists and apologists acknowledge a gap separating the *is* from the *ought* of human reasoning. Meliorists find that this gap can be closed by improving agents’ “second nature,” for instance through providing education and instilling motivation. Apologists more narrowly acknowledge the potential of improving agents’ behavioral output, for instance by aligning the presentation of information to agents’ ready-abilities, that is, their “first nature.” Apologists, as the name suggests, seek to offer explanations why a fallacy has arisen, but may also explain alleged fallacies “away.”

These three stances remain important to one’s view on the fallacies, for what would plausibly be a fallacy for a meliorist need to be so for an apologist. By contrast, it is at least initially not clear whether a panglossian in fact possesses a meaningful notion of ‘fallacy’ in the first place.

Condition 9 (“Understood”) and 10 (“Within-ability”), we submit, are typical apologist conditions, for if either of these conditions are violated one can explain why the fallacy arose, or may support the claim that the alleged fallacy is not a *genuine* fallacy after all. A meliorist will probably not endorse 9 and 10 *as conditions for fallacy*. But she will need at least condition 10 for her overall project to be reasonable. Violations of condition 9 can play a load-bearing role in pragmatic accounts of fallacy such Pragma-dialectics (van Eemeren, 2010), where goal conflicts explain why fallacies arise, or in Walton’s dialogical account (1995), where illicit shifts between discourse types, that change the task and the goals, are viewed as fallacies. Both accounts *distinctly* specify this condition.

Conditions 12, 13, 14, but also 7 come from Woods’ account (2013), on which most traditional fallacies—his “gang of eighteen”—aren’t fallacies after all. Condition 12 (“Frequency”)—being controversial, of course, since one lacks reliable data (see above)—makes

the most sense for an apologist. After all, it is hard to see why a meliorist or a panglossian would care about frequency considerations.

Condition 13 (“Attractiveness”) is known as an Aristotelian insight. In the formulation we have offered, it best fits with an apologist stance, but also a meliorist, and even a panglossian, could endorse it. Walton (1995), for instance, phrases this condition more positively as: “has a semblance of correctness about it” (p. 255), which might perhaps take center stage for a panglossian. Generally, condition 13 seeks to explain why we “have” fallacies in the first place.

Condition 14 (“Incorrigibility”) ups the ante, so to speak, by identifying as fallacies only transition-patterns that are broadly *resistant* to correction, a condition which particularly an apologist, but also a panglossian, could endorse. By contrast, meliorists might at least struggle with condition 14, since it presents the *natural* “halting point” for their larger project.

Condition 15 (“Size”), if false, is an apologist move, regularly invoked among theorists of ecological rationality (see 12). Condition 16 (“Similarity”) is another apologist move that seeks to explain why condition 13 is the case, if it is.

4.3. Summary

The following summarizes what we had offered above.

Core conditions

1. Systematic
2. Reconstructable
3. Reasoned

Narrow sense of ‘fallacy’

4. Constrained: normally affirmed *tout court*
5. Decidable: secured via developmental status of normative theory
6. Accuracy: secured via choice of normative theory

Extended sense of ‘fallacy’ and pragmatic-dialectical sense of ‘fallacy’

7. Quality: distinct specifications
8. Goal-directed: distinct specifications
9. Aptness: distinct specifications

Conditions reflecting basic stances on human rationality

10. Understood: apologist
11. Within-ability: apologist, meliorist
12. Frequency: apologist
13. Attractiveness: apologist, meliorist, panglossian
14. Incorrigible: apologist, panglossian
15. Size: apologist
16. Similarity: apologist

5. Conclusion

We have offered some constituents of a usage-based analysis of ‘fallacy’, listing 16 conditions that have, for the most part implicitly, been discussed in the literature. Three related conceptions of ‘fallacy’ have been identified: *a narrow sense, an extended sense, and a pragmatic-dialectical sense*. Of the 16 conditions, three were said to “carve out” a semantic core of fallacy, another six were said to be specifiable so as to distinguish three core-specifications. In particular, the pragmatic-dialectical sense of fallacy was claimed to be a variant of the extended sense. Our discussion has moreover suggested that at least seven conditions can be related to the three normative positions in the philosophy of human reasoning: the meliorist, the apologist, and the panglossian. By far most of these fell on the side of the apologist. Having here made these conditions available for discussion, we invite additions and criticism, stressing that this analysis-sketch is non-final.

Acknowledgements: Research was conducted while holding a European Union FP7 Marie Skłodowska Curie COFUND fellowship at the Slovak Academy of Science, Institute of Philosophy, Bratislava, under grant no: 1225/02/03.

References

- Adler, J. E. (1997). Fallacies not fallacious: Not! *Philosophy & Rhetoric* 30 (4), 333–350.
- Adler, J. E. (2008). Introduction. In J. E. Adler & L. Rips (Eds.), *Reasoning: Studies of Human Inference and Its foundations* (pp. 1–34). Cambridge: Cambridge University Press.
- Barth, E. M., & Krabbe, E. C. W. (1982). *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*. Berlin/New York, NY: de Gruyter.
- Boudry, M., Pigliucci, M., & Paglieri, F. (2015). The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation* 29 (4), 431–456.
- Brandom, R. (1988). Inference, experience, and induction. *Philosophical Studies* 54, 267–285.
- Eemeren, F. H. van, & Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge, UK: Cambridge University Press.
- Eemeren, F. H. van (2010). *Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-dialectical Theory of Argumentation*. Amsterdam: John Benjamins.
- Evans, J. S. T. B. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin* 128 (6), 978–996.
- Finocchiaro, M. A. (1981). Fallacies and the evaluation of reasoning. *American Philosophical Quarterly* 18 (1), 13–22.
- Gigerenzer, G., & Brighton, H. (2009). Why biased minds make better inferences. *Topics in Cognitive Science* 1, 107–143.
- Govier, T. (1983). Who says there are no fallacies? *Informal Logic Newsletter* 5 (1), 2–10.
- Hahn, U., & Corner, A. (2013). Normative theories of argumentation: Are some norms better than others? *Synthese* 190 (16), 3579–3610.
- Hamblin, C. (1970). *Fallacies*. London: Methuen.
- Hansen, H. (2015). Fallacies. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from <http://plato.stanford.edu/archives/sum2015/entries/fallacies/>
- Hansen, H. V. (2002). The straw thing of fallacy theory: The standard definition of ‘fallacy’. *Argumentation* 16, 133–155.

- Hansen, H. V. (2007). Aristotle, Whately, and the taxonomy of fallacies. In: D. M. Gabbay, & H. J. Ohlbach, (Eds.), *Practical Reasoning* (pp. 318–330). Berlin and Heidelberg: Springer.
- Hintikka, J. J., & Spade, P. V. (2012). History of logic. *Encyclopedia Britannica*. Retrieved from <http://www.britannica.com/EBchecked/topic/346217/history-of-logic>
- Johnson, R. H. (1987). The blaze of her splendours: Suggestions about revitalizing fallacy theory. *Argumentation* 1, 239–253.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Strauss and Giroux.
- Massey, G. J. (1981). The fallacy behind fallacies. *Midwest Studies in Philosophy* 6 (1), 489–500.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review* 101, 608–631.
- Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and re-evaluation. *Psychonomic Bulletin & Review* 10, 289–318.
- Powers, L. H. (1995). The one fallacy theory. *Informal Logic* 17 (2), 303–314.
- Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.
- Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford: Oxford University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics & biases. *Science* 185, 1124–1131.
- Walton, D. N. (1995). *A Pragmatic Theory of Fallacy*. Tuscaloosa: University of Alabama Press.
- Walton, D. N. (2010). Why fallacies appear to be better arguments than they are. *Informal Logic* 30 (2), 159–184.
- Woods, J. (2013). *Reasoning Errors: Naturalizing the Logic of Inference*. London: College Publications.