

Two-Tier Fallacy Theory: A New Approach to Assessing Argument Quality

Martin Hinton

University of Lodz
martin.hinton@uni.lodz.pl

Abstract: Conceptions of fallacies suggested by philosophers vary significantly. Often these contributions are little more than lists, only sometimes approaching a fully-developed theory of fallacy. Where there is a clear understanding of what is meant by the term fallacy, the problem of how to identify them in discourse remains, often leading to a conflation of descriptive and evaluative analyses. We present a two-tier procedure that strictly distinguishes the descriptive and normative dimensions of identifying fallacies. The combination of the descriptive Argument Type Identification Procedure (ATIP), which enables the characterization of an argument in terms of the Periodic Table of Arguments (PTA), and the evaluative Comprehensive Assessment Procedure for Natural Argumentation (CAPNA), provides a basis for systematic, repeatable, and explainable argument acceptability judgements. We explain how this two-tier procedure overcomes some of the difficulties of fallacy identification and categorization and list several other advantages that a procedural approach to fallacies brings.

Jean H. M. Wagemans

University of Amsterdam
j.h.m.wagemans@uva.nl

Resume: Les conceptions des sophismes proposées par les philosophes varient considérablement. Souvent, leurs contributions se limitent à de simples listes, n'abordant que rarement une théorie pleinement élaborée du sophisme. Même lorsque la notion de sophisme est clairement définie, la difficulté de les identifier dans le discours persiste, ainsi conduisant fréquemment à une confusion entre les analyses descriptives et évaluatives. Nous présentons une procédure à deux niveaux qui distingue clairement les dimensions descriptives et normatives de l'identification des sophismes. La combinaison de la Procédure d'Identification des Types d'Arguments (PITA), qui permet de caractériser un argument à l'aide du Tableau Périodique des Arguments (TPA), et de la Procédure d'Évaluation Globale de l'Argumentation Naturelle (PEGN), fournit une base pour des jugements d'acceptabilité des arguments systématiques, reproductibles et explicables. Nous expliquons comment cette procédure à deux niveaux surmonte certaines difficultés liées à l'identification et à la catégorisation des sophismes et énumérons plusieurs autres avantages qu'apporte une approche procédurale des sophismes.

Keywords: argument evaluation, fallacy theory, explainable argumentation theory (XAT), Periodic Table of Arguments (PTA), Comprehensive Assessment Procedure for Natural Argumentation (CAPNA).

1. Introduction

There is a constant and natural desire in everyday life as much as in philosophical deliberation to make judgements on the acceptability of arguments. Within the field of argumentation theory, reflections about how to evaluate arguments are often framed in terms of “fallacy theory”, a fact which in itself reflects an inclination to focus on what can go wrong with arguments rather than what can go right. The notion of fallacy, often vaguely defined and loosely conceptualized, is one of the foundational ideas of argumentation theory and provides the basis for a large proportion of work done in the study of argument and reasoning. This combination of great interest and often changing conceptualization has led to a multitude of classifications of individual fallacies and suggestions as to what constitutes fallaciousness, with varying degrees of theoretical justification and often little empirical evidence. In this paper, we present a new approach to fallacies. To facilitate the understanding of how it relates to the state-of-the-art in fallacy theory, we first identify and discuss four broad groups of existing approaches: the formal, the informal, the teleological, and the cognitive. Each has a different conception not just of how to categorise and describe fallacies, but of what fallacies themselves are. All of them also have serious difficulties with the questions of how exactly arguments should be evaluated and fallacies identified. At the root of this, we believe, lies the conflation of the identification of argument types and the evaluation of argument instances; and the opacity of the assessment process makes the resolution of disagreements over the acceptability of arguments awkward and frequently unproductive.

Then, in order to overcome these difficulties, we set out an approach to fallacy identification which separates the very different aspects of description and evaluation which are frequently confused in the literature. We do so by proposing a two-tier procedure for identifying fallacies consisting of a concatenation of (1) an argument type identification procedure yielding a description of the characteristics of the argument and (2) an assessment procedure for determining its acceptability based on normative criteria.

The two-tier procedure reflects our point of view that a fallacy judgement should not be made from an a priori description of a “bad” argument which in some way resembles a classic case or paradigmatic form.

Instead, the evaluation should be based on an identification of the type of argument in question, characterizing it in a purely descriptive way. The argument is only to be judged as fallacious as a result of a subsequent evaluation in which it is found to be unacceptable based on criteria derived from norms regarding the process in which the argument is put forward, its underlying reasoning, and the language it employs.

Our development in this paper of a two-tier procedure for identifying fallacies is part of a wider process-based, systematic approach to understanding argumentation and making explicit all aspects of its description and evaluation which we refer to as explainable argumentation theory (XAT). This procedural explainability makes both the division between fallacy types and the origin of the individual fallacy judgement clear. While the norms of argumentation we employ are taken from across the literature and are largely uncontroversial, the way in which they are brought together is both a drive towards innovation and unification within the field.

In Section 2 below, we describe the four main approaches to fallacies to be found in the contemporary literature mentioned above. This is followed in Section 3 by further elucidation of some of the problems from which those approaches suffer. We then present our solution in the form of a two-tier procedure for making fallacy judgements: in Section 4, we expound the descriptive tier, the Argument Type Identification Procedure (ATIP), and in Section 5, the normative tier, the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA). In Section 6, we discuss the implications of our approach to fallacy identification. We address some possible objections and state clearly what we believe the main advantages to be. Finally, in Section 7, we conclude with a brief recapitulation of our central claims and some consideration of how this work can be developed and improved in the future.

2. Four established approaches to fallacies

As the history of the concept of fallacy is long and rich, there are many ways to distinguish the different approaches taken to its study (for an overview of the history of fallacy theory, see, e.g., Hansen 2023, pp. 12-34). There has been a tendency to distinguish traditional and modern approaches, which somewhat mirrors the division of the formal and the informal. Modern fallacy theory is often taken to begin with the work

of Hamblin (1970), who rejected what he described as the “standard treatment” of fallacies, claiming that it had traditionally centred around the concept of validity. However, this claim was disproved by Hansen (2002), who showed the great diversity of approaches taken by earlier scholars and that many did not focus on validity at all. Moreover, the distinction between traditional and modern suggests a degree of unity within those two groups which simply does not exist. Contemporary theory is greatly varied and to state that Aristotle in listing his thirteen Sophisms, Whately (1857) in dividing his fallacies into the Logical and Non-logical, and Mill in making a division “according to the nature of the apparent evidence from which the conclusions are drawn” (1882, p. 903) were following one ‘traditional’ approach is clearly unhelpful and unjustifiable. After all, the list of fallacies “has grown as a hodgepodge collection of items, proposed at various times and from different perspectives” (Hansen 2023, p. 55).

Distancing ourselves from the distinction between traditional and modern approaches, we present in this section a fourfold division of established approaches: the formal, informal, teleological, and cognitive. The division does not ignore historical developments but is made rather on systematic and conceptual grounds. Given the influence of the disciplines of logic, dialectic, and rhetoric on the various conceptualizations of fallacy, we first distinguish between “formal” approaches as primarily developed in logic and “informal” approaches as developed in dialectic and rhetoric. Second, we have singled out approaches combining formal and informal insights while declaring a defined purpose for argumentation and given them the new label “teleological” approaches to fallacy. Finally, we distinguish cognitive approaches as they deviate from the other three by studying the underlying psychological mechanisms of fallacy judgements rather than argumentative behaviour or its discursive products.

In some classifications of fallacies, the above labels (and other ones) are used to indicate types and subtypes of fallacies rather than approaches to fallaciousness itself. Within this usage, for example, “affirming the consequent” is called a formal fallacy, “straw man” an informal fallacy, and “confirmation bias” a cognitive one. However, since the same fallacy can be studied from various perspectives, we have opted to reserve these labels for indicating the approaches to fallacy

rather than for categories into which individual fallacies might be placed.

Our descriptions of these four approaches are based on a wide reading of the literature, as the references in each section attest, and informed particularly by van Eemeren et al. (2014), Hansen (2023), Wagemans (2022), and Zenker et al. (2024). In our characterizations, we not only mention key definitions and conceptualizations of fallacy but also pay attention to the proposed methods for detecting fallacies or making fallacy judgements. Our accounts of these methods are guided by questions of what textual, discursive, or cognitive object is considered, which criteria are used in evaluating that object, and what other conditions should be fulfilled to enable identification of the fallacy at issue. By doing so, we prepare the ground for our critical discussion of the approaches in the next section.

2.1 Formal approaches

As the name indicates, “formal” approaches to fallacies focus their assessment on one particular aspect of an argument: its form. They trace their roots back to Aristotle’s seminal works on logic, in particular, the so-called assertoric syllogistic (*Analytica priora* I.4-7), which is aimed at determining the validity of arguments with two premises and one conclusion (for a description of this system see, e.g., van Eemeren et al. 2014, pp. 94-105; Wagemans 2018, pp. 578-582).

As their object of evaluation, formal approaches target the underlying reasoning of an argument, scrutinizing its structure and coherence while abstracting away from the specific content or subject matter at hand, the speaker or writer’s intention behind an argument, and the discourse genre in which it functions. Fallacies are thus analysed and evaluated in a context-independent manner. Their identification takes place by examining the logical form of a piece of reasoning and subsequently determining its validity using a set of predefined rules or a collection of known fallacious patterns (for different conceptualizations of logical validity see, e.g., van Eemeren et al. 2014, pp. 150-162).

According to formal approaches, a fallacy occurs when the reasoning underlying an argument is flawed due to a formal or structural error. This happens, for instance, when someone infers the truth of the antecedent from that of the consequent in a *modus ponendo ponens*, thereby

constituting the “fallacy of affirming the consequent”. Fallacy judgements are justified by demonstrating how the argument under scrutiny constitutes a violation of a logical rule or conforms to an example of a fallacy.

Consequently, according to the formal approach, the identification of fallacies is not a trivial task that can be easily accomplished by anyone without a solid grasp of logical principles and the ability to discern faulty patterns of reasoning. To detect a fallacy, the evaluator must be able to find the logical form within the content, looking for structural flaws that compromise its validity.

2.2 Informal approaches

“Informal” approaches to fallacy theory emerged in the 1970s, partially out of dissatisfaction with the formal approach and the dominant role of logic in fallacy theory (see, e.g., van Eemeren et al. 2014, pp. 373-423). Instead of focusing on structural errors in the underlying reasoning of arguments, informal approaches recognize that logical validity alone may not ensure an acceptable argument, as many other factors significantly influence their quality. Moreover, they acknowledge that, under certain circumstances, a logically invalid argument can be reasonable after all. Rather than using the notions of “validity” and “soundness”, therefore, informal approaches work with alternative notions to indicate the quality of arguments such as “cogency”, further divided into “acceptability”, “relevance”, and “sufficiency” (Johnson & Blair 1977/1993; Govier 2010).

Central to informal approaches is the concept of “defeasibility”, which makes positive or negative evaluations of arguments to some degree provisional. New information or arguments may come to light, strengthening or weakening the inference, but a judgement can be made, accepting the current state of knowledge. This is explicitly acknowledged in the discussion of argument schemes (e.g. Walton 2011). How much knowledge there is and how much is required may also be linked to the dialectical situation which implies that fallacy judgements are dependent on the broader context in which the argument is presented.

Among the contextual factors taken into consideration are the arguer’s credibility, intent, or underlying assumptions; the specific cultural or institutional circumstances of the argumentative exchange; and

the audience's preferences and background knowledge. An important consequence of including such factors in assessing an argument is that it may be perfectly reasonable within a specific context but become fallacious when transferred to another. According to informal approaches, fallaciousness is not an absolute or inherent characteristic of an argument but rather a contingent judgement influenced by contextual factors.

An example of how contextual factors are of great importance in argument evaluation comes from the argument from expert opinion. In this case, factors such as the expert's qualification, track record, personal reliability, and conformity with other experts' opinions are crucial to the assessment of the strength of the inference. These factors outside the contents of the argument itself can affect acceptability, reflecting the informal approach to fallaciousness: an argument is not fallacious per definition but is judged based on factors relevant to the situation in which it is offered.

Informal approaches usually acknowledge that the ability to make fallacy judgements is not necessarily limited to specific individuals. While expertise, be it acquired by training or experience, may enhance one's proficiency in justifying such judgements, and put one in a better position to evaluate many specialist arguments, it is not a prerequisite. Both experts and non-experts, participants and passive audiences, and analysts of argumentative discourse can make fallacy judgements, some from a position of greater authority than others.

2.3 Teleological approaches

The third approach to fallacies we characterise here is the "teleological" approach. This is not a category commonly distinguished in the literature, but we include it in our overview under this label because it takes a middle ground between the formal and informal approaches. Like the former, it employs an overarching norm for making fallacy judgements, and like the latter, it considers contextual factors to play an essential role in such judgements.

We apply the label 'teleological' because scholars working within this approach emphasize that argumentative exchanges are goal-directed activities and that certain types of arguments or behaviour promote accomplishing that goal while other types hinder or obstruct it. In

Walton's new dialectic, for instance, a fallacy is understood as "an argumentation technique that could be used rightly in one context of dialogue, but is used wrongly in the particular case in question, in a manner that actually hinders the real and legitimate goals of the type of dialogue" (1992a, p. 267). In another teleological approach, van Eemeren and Grootendorst's pragma-dialectics, argumentative discourse is reconstructed in terms of a "critical discussion", which is a discussion in which "the parties involved in a difference of opinion attempt to resolve this difference by achieving agreement on the acceptability or unacceptability of the standpoint(s) involved" (2004, p. 58). As they explain, the resolution of differences of opinion is governed by rules, and "every violation of any of these rules can make the resolution of a difference more difficult, or can even obstruct it. We shall take such a violation to be a fallacy" (2004, p. 162). So-called "epistemological" approaches to fallacies can also be labelled teleological, as they assume that "argumentative exchanges should lead to an improvement of the epistemic state or epistemic situation of the people involved. This means that at the end of the argumentative exchange, they should have acquired new knowledge or be able to better justify the beliefs they already had. Ideally, their beliefs will have been brought in accordance with - or a step closer to - the truth" (van Eemeren et al. 2014, p. 394).

According to teleological approaches, fallacies occur when argumentative behaviour does not comply with the rules regimenting the argumentative activity or when the argumentation does not match the relevant criteria for the rational justification of beliefs. While the nature and content of these rules and criteria may vary, teleological approaches all derive them from what could be called the "overarching" norm that a good argument promotes the accomplishment of the (assumed) goal of argumentative exchanges. In some theories, the goal is realistic and varies with the dialogue type within which the argumentation is produced, as in Walton's (1998) new dialectic. In other cases, it is an idealized assumption deemed generally applicable to all types of argumentative discourse, as the goal of resolving a difference of opinion in van Eemeren and Grootendorst's (2004) pragma-dialectics and that of improving the epistemic situation in the epistemological approaches described by Lumer (2005).

An example is the argumentative behaviour of evading the burden of proof. According to the new dialectic, such evasion is considered fallacious within the context of a scientific discussion but may be reasonable within certain legal settings. This difference occurs because scientific and legal discussions have different goals. From a pragma-dialectical perspective, it is always and unconditionally fallacious because it hinders the resolution of the difference of opinion (Snoeck Henkemans & Wagemans 2015). And from an epistemological perspective, it is fallacious because evading the burden of proof violates the “responsibilist” criterion of the arguer being epistemically responsible, i.e., actually transferring their knowledge to improve the epistemic state of the addressee. In all cases, the goal of the argumentative activity is decisive for judging the quality of the argumentative move. Due to this, there is no distinction made between unacceptable moves which are common problems across all types of communication, such as vagueness and amphiboly, and those pertaining only to argumentative situations, nor between certain behaviours of arguers in those situations and the characteristics inherent to their arguments.

2.4 Cognitive approaches

The fourth and final approach to fallacies which we discuss here is the “cognitive” approach. Unlike the others, this approach studies the underlying cognitive processes of faulty reasoning in order to shed light on the mechanisms that lead people to accept arguments that are fallacious - representative collections of articles within this approach are Herman and Oswald (2014) and Paglieri et al. (2016).

An important assumption of the cognitive approach is that fallacies have a certain potential to mislead people, which aligns with the widely used definition of a fallacy as an argument that seems valid but is not (Hamblin 1970). While a fallacy presents itself as a sound and reasonable argument, upon closer examination, it becomes evident that its underlying reasoning is flawed or invalid. This “appearance condition” (Hansen 2023, pp. 55-56) explains why fallacies can be used for persuasive purposes.

Scholars working within the cognitive approach aim to find explanations of this phenomenon, trying to understand how fallacies exploit how our brains work and tap into the psychological mechanisms that shape human thinking and decision-making.

A great many of these psychological mechanisms are mental shortcuts, or heuristics, which serve as time-saving mechanisms that help humans make quick decisions but can also lead to errors in judgement and reasoning (see, e.g., Gilovich et al. 2002; Mercier & Sperber 2011). An example is the availability heuristic, which explains the so-called “anecdotal fallacy” in terms of the mental shortcut of relying on readily available information to evaluate the likelihood of an event.

Unlike the other three approaches, the cognitive approach to fallacies does not usually develop its own norms for making fallacy judgements. In most cases, the criteria for the quality of the reasoning underlying arguments come from the formal, informal, or teleological approaches, and are used as a starting point for designing experiments in which particular fallacies are associated with or mapped onto cognitive biases. Such approaches, then, are not aimed at providing norms for reasonable arguments. Rather, the very notion of fallacy and a method for identifying them is presupposed in any cognitive approach, which is aimed at helping us understand why people accept fallacies.

3. Problems of the established approaches

In this section, we discuss problems with the described approaches to fallacies, thus preparing the ground for our own proposal of a two-tier fallacy theory. To a degree, the criticism of one approach is inherent in the assumptions of the others. The development of informal approaches took place within the context of the Informal Logic movement and was an explicit reaction to the inadequacies of formal logic in evaluating real world discourse. At the same time, teleological approaches stress that argumentation cannot be judged outside the framework of a purpose of the exchange and thus see the labelling of arguments as good or bad in themselves as unjustified. This is an explicit rejection of evaluations based on argument structure alone.

Cognitive approaches are certainly relevant to the study of real world discourse and they can give us important insights into the workings of the human mind. Nevertheless, they represent a kind of second-order approach to fallacy theory. In studying how biases occur and how our own processes of cognition lead us into erroneous reasoning, we are already assuming an understanding of what would constitute such reasoning. The cognitive approach can tell us how fallacies work to lead

us into false beliefs and poor decisions, but it cannot say why they are fallacies. If we try to define a fallacy cognitively as a form of argument that leads to poor reasoning, we are locked into a circular track where poor reasoning must be defined as that which comes about because of fallacious arguments. What fallacies are, therefore, must be defined externally, using a different approach. In this way, cognitive approaches can be said to build on conceptions of fallacy developed by other approaches, not to challenge them.

In what follows, then, we concentrate on the other categories: formal approaches, associated with Logic; informal approaches, often associated with Informal Logic and Critical Thinking; and teleological approaches, in particular, Dialectics. For reasons of space, we cannot, of course, discuss all the comments and criticisms made by researchers on these fields. Instead, we concentrate on the area in which our two-tier approach is intended to make its most significant contribution: the conflation of the descriptive and normative dimensions of identifying fallacies and the problems which follow from it.

The importance of separating these two dimensions is an issue of which we believe argumentation scholars across the spectrum of approaches are very well aware. However, the lack of systematic procedures with which to do so leads to the distinction between characterizing an argument and evaluating it becoming blurred, and acts as an invitation to analysts to conflate the two aspects. Below, we discuss several problems we believe to reflect this conflation, many of which are related to the use of lists of fallacy names, some “traditional”, some of more recent coinage, which is a major feature of work within the informal approaches.

3.1 The ambiguity problem

Firstly, the names of many fallacies are ambiguous, both in terms of their descriptive and evaluative content. For example, the “slippery slope” fallacy is well-known, but there is no agreement amongst scholars as to its characteristics, whether it can also be a respectable form of argument, or what actually makes it fallacious (cf. Rizzo & Whitman 2003; Walton 2015; de Swart 2018; and Hinton 2018). Even when a fallacy is clearly defined by its ‘discoverer’, as in the case of Locke’s “ad” fallacies, and several of Aristotle’s 13 sophisms, there is still a tendency for the usage to slip over time: Locke’s *ad hominem* did not

refer to a personal attack, and his *ad verecundiam* was not any form of argument from expert opinion (see Davies 2023). It is also true, of course, that Locke did not use the word fallacy to describe these patterns of reasoning.

This problem is particularly clear in the differing uses which scholars make of the names: some retaining the traditional Latin names for fallacious uses only, others discussing non-fallacious examples under the same label. For instance, Copi and Cohen (1990) consider the *ad ignorantiam* a fallacy, but accept that some arguments from ignorance may be acceptable; Walton (1992b) and Wreen (2003), on the other hand, when discussing Copi and Cohen's work, use the terms interchangeably. Another case where scholars attempt to separate the two concerns the use of the label 'personal attack' or '*argumentum ad personam*' to describe an argumentative move and the term '*argumentum ad hominem*' for a fallacious personal attack. Again, disagreement is just around the corner, since other authors are happy to discuss non-fallacious uses of the *ad hominem* (e.g. Johnson 2009), and clearly regard it as a descriptive not evaluative label. Even where the structure referred to by the name is clear, then, which is far from always the case, the use of the same name to describe a structure and to condemn an argument as a fallacy is deeply problematic.

This ambiguity in naming is not confined to informal approaches to fallacy. Within formal logic, the fallacy of affirming the consequent is actually an invalid example of the *modus ponendo ponens* form of reasoning. Nevertheless, when someone labels an argument as a '*modus ponendo ponens*', they usually not only intend to say it is an argument that affirms the antecedent but also that they take the argument to be valid. The descriptive term for the form doubles as the evaluative term for a valid example of that form. The reason for this conflation is historical: originally, *modus ponendo ponens* meant to posit something based on positing something else, and this happens in both the valid and invalid variants. The valid variant posits *q* based on positing *p*, and the invalid one posits *p* based on positing *q*. In the later tradition, the terminology has evolved to the one used in present-day introductions to logic. The same goes for the descriptive label '*modus tollendo tollens*' and the associated evaluative label 'denying the antecedent'.

A related issue is that the evaluative nature of formal descriptions makes it hard to apply them to the realities of human communication

where validity may not be the relevant standard. Some argument forms are valid but nonsensical in dialogue, such as tautologies and irrelevant disjunctions, while other ‘logical fallacies’ are often quite acceptable in everyday contexts. Davies (2020) discusses three examples of the phenomenon of *epomenon* given by Aristotle, all of which look like the type of abductive inferences we make all the time in normal life, despite the fact that they are, technically speaking, guilty of affirming the consequent, which is usually considered a type of fallacy. One is the conclusion that a man with a high temperature has a fever, another that wet ground outside means that there has been rain, and the third, more speculatively, that a finely dressed man going out in the evening is having an affair. As Davies notes, ‘without such inferences, we would be quite at a loss to go about our everyday business’ (2020, pp. 35-36).

3.2 *The assignation problem*

The ambiguity problem just described is the direct result of the lack of separate procedures for first identifying the characteristics of an argument and then assessing its acceptability. Those working with named fallacies may point to the existence of well-developed collections of argument schemes as tools for evaluating arguments, but each of those schemes bears the name of an argument type. This brings us to the second major difficulty, how are those names to be assigned to individual arguments? This is a step on which the extensive work on schemes by Walton et al. (2008) stays silent across its 400 pages.

The methodology for labelling arguments and thus subjecting them to the correct set of Critical Questions linked to the appropriate argument scheme appears to rely on a process of matching real examples with paradigmatic templates. In many cases, of course, this procedure works perfectly well: we can recognise an appeal to popular opinion when we see one. However, it is less obvious that we can tell which of the 11 sub-types of that scheme listed by Walton et al. (2008, pp. 311-313) we should apply.

In general, as Wagemans (2020, p. 2) has pointed out, the method of comparing the elements of an argument to those of an argument scheme in order to identify its type can lead to problems of justification. In case of discrepancies between the real and the ideal, in the absence of a step-by-step procedure, it is not clear how many differences there can be or how far the argument may diverge from the model to still be considered

a match. Also, given the incompleteness of the list of argument schemes, we have no method to decide whether we have found a deviant example of an argument type already included in the list or a new one that should be added. Due to these problems, there is a risk of confusion and disagreement between analysts using this comparative method. Most seriously, they are likely to be drawn to a particular scheme on the basis of where they think the argument may be faulty, thus basing their identification on their evaluative inclination.

One attempt to solve this problem has been made by Johnson and Blair who allow the use of the name of a category of fallacy, such as ‘hasty conclusion’ as a broad fallacy name where it isn’t clear exactly how one instance should be matched. In this way, the category names “fill in any gaps between specific labels” (1993, p. 52). This move, however, invites more problems with ambiguity as one term is being used to mean a group of fallacies at one moment and a particular fallacy at another.

3.3 The underdetermination problem

A third problem is that a negative evaluation of an argument results in a fallacy name, but not a reason why the argument is fallacious. Stating that any argument is a fallacious argument of a particular type, without any further analysis or explanation of what is problematic and why, tells us little about what exactly has gone wrong and how it might be corrected. We are left with an ambiguous description “fallacious example of this type” rather than a proper evaluation.

The named fallacies, then, are a curious mix of the descriptive and the evaluative. They are named for their structure but known for their value. This same mixture, although not identified as such by the authors, is what leads to both Woods’s suggestion that “the traditional concept of fallacy is indeed empty” (2007, p. 19) and to what Boudry et al. (2015) describe as the “fallacy fork”. They claim that the entire concept of the named fallacy flounders on the prongs of a dilemma, but only belatedly note the cause: “Our very labeling practices suggest that reasoning defects can be gleaned from the respective labels” (2015, p. 452). This absorption of the error into the name of the practice is the result of the conflation which we are endeavouring to highlight.

A related issue is that the labelling of an argument with a tag which represents a negative evaluation has the effect of dismissing it from

consideration rather than simply weakening its impact. We suggest below that the more precise and carefully defined identification of argument flaws still allows an analyst to make a decision as to their strength and acceptability, acknowledging that some flaws are more serious than others.

3.4 The diverse reference problem

The fourth problem with the named fallacies is that the names themselves mask the fact that the phenomena they refer to are diverse. Some fallacies are supposed to be structural flaws in arguments, others occur because of the way an argument is used in a particular dialogue, and still others relate to moves in argumentation which are not actually arguments, but rather behaviours. Woods (2007, p. 3) discusses what he calls the traditional ‘gang of eighteen’ fallacies, of which some are structures (e.g. confirming the consequent), some are dialogical errors (e.g. *ignoratio elenchi*) and others are problems of expression (amphiboly). These names are all reflections of a negative evaluation, but what process of description could be followed to discover them when they refer to such very different things?

3.5 The idealization problem

It may be thought that much of what was described in the previous sections does not apply to teleological approaches, but this is not quite true. As we explained above, taking the teleological perspective means establishing a purpose for an argumentative dialogue and thus encouraging behaviours which aid in the accomplishment of that goal and evaluating negatively any which do not. In describing what is happening within the discourse, teleological approaches make use of idealizations regarding the assumed aims of the participants or the dialogue type in which they occur.

Boudry et al. (2015) claim that their fallacy fork dilemma applies as much to the teleological approach of pragma-dialectics as to other approaches. This is so because that entire system is based on the conception of the Critical Discussion, which is an idealisation: “Even if we assume, for argument’s sake, that the pragma-dialectical approach successfully demarcates epistemically suspect moves, these will be limited to concrete situations where the conditions of a critical discussion are

fulfilled” (2015, p. 447). These conditions, they suggest, are rarely met, and there is no doubt that they are frequently not met. That means that, just as with the named fallacies from the traditional list, the pragma-dialectical fallacies are simply not found in real discourse in their pure, easily-identifiable form. They go further, and state that even within argumentation contexts which do meet the conditions of the Critical Discussion, it is not obvious that it is possible to discriminate good reasoning from bad reasoning.

The point here is that either fallacious behaviour is obvious, in which case we leap straight to a negative evaluation, or it is not, in which case we need a nuanced description of it which is not so easily evaluated. When an argument move is described as rule-breaking the two steps of description and of evaluation have occurred in one moment. Even where a clear procedure for the description of argument moves exists there is a danger that a move which is unwelcome from the viewpoint of one protagonist will be described in such a way as to identify it as a rule-breaker. This is particularly relevant where the system depends on intersubjective validity, which means an agreement among the participants about the description or evaluation dictates the reasonableness of a move.

We are well aware that proponents of teleological approaches will claim that their systems do have robust procedures in place for describing moves before reaching an evaluation. However, our contention is that these procedures are not fully systematised and not sufficiently distinct from the evaluative framework within which they are working for general use. It should be noted that in the proposals we make below, we do not seek to deny the value of these approaches – or indeed any of those we have mentioned – rather we aim to build upon their insights by constructing a clear two-tier procedure for fallacy detection and, thus, a solid ground for the justification of fallacy claims.

4. The descriptive tier: The Argument Type Identification Procedure (ATIP)

Our two-tier procedure for identifying fallacies is based on the assumption that different types of arguments require different sets of criteria for their assessment and that, therefore, to evaluate an argument it must first be clear what type it instantiates. To establish this, we propose to

use the so-called Argument Type Identification Procedure (ATIP), which functions as a manual for identifying the type of argument in terms of the categorization framework of the Periodic Table of Arguments (PTA) (Wagemans 2016, 2023, 2025). Below, after having explained the theoretical framework of the PTA, we briefly describe the main steps of this procedure.

The PTA conceptualizes argument as a triadic structure consisting of a conclusion, a premise, and a lever. Each of these components is expressed in a statement in which a predicate is attributed to a subject, which allows the concept of a “lever” to provide a more detailed account of the inference underlying the argument than, for instance, a conditional or a Toulminian warrant. This conceptualization of argument differs from that in other approaches, which usually take an argument to consist of a conclusion and one or more premises. The PTA is “atomistic” because it posits that complex argumentation, which consists of multiple premises in various constellations (serial, convergent, linked), is built from these minimal arguments.

To classify arguments, the PTA takes a “parametric” approach in that it distinguishes between different types of argument by determining the value of three basic characteristics: the argument form, the argument substance, and the argument lever. An argument type is defined as the combination of the values of these parameters, which implies that arguments are instantiations of the same type when they share the same values and of different types if there is no or just a partial overlap of values.

The first parameter, the *argument form*, refers to the specific configuration of the subjects and predicates contained in the conclusion and the premise of the argument. The PTA distinguishes four argument forms, denoted by the Greek letters *alpha*, *beta*, *gamma*, and *delta*. In the following, we mention the configurations of subjects and predicates associated with these forms, using a, b, and c to indicate the subjects and X, Y, and Z to indicate the predicates.

If the conclusion and premise have the same subject and a different predicate, the argument takes the alpha form: “a is X because a is Y”. An example is “This book (a) is great (X) because it (a) was written by Javier Marías (Y)”. In arguments with the beta form, the conclusion and the premise have the same predicate but a different subject: “a is X because b is X”. An example is “Mars (a) is inhabited (X) because Earth (b) is inhabited (X)”. If the subjects and predicates of the conclusion

and the premise are both different, the argument takes the gamma form: “a is X because b is Y”. An example of such an argument is “Self-control (a) is good (X) because a hot temper (b) is bad (Y)”. The last form distinguished in the PTA, the delta form, has the entire conclusion reappearing as the subject of the premise. If we indicate that property by ‘Z’, the conclusion by ‘q’, and the acceptability of the conclusion by ‘A’, we can represent the delta form as “q [is A], because q is Z”. An example is “The economy has grown (q) [is acceptable (A)], because that (q) was said by the Prime Minister (Z)”.

The value of the second parameter, the *argument substance*, provides information about the content of the conclusion and premise. Which information that is, depends on the form. In arguments of the alpha form, for instance, the substance is expressed in terms of a tripartite typology of statements distinguishing between statements of fact (F), statements of value (V), and statements of policy (P). In short, statements of fact (F) describe a state of affairs that can be empirically observed or imagined, such as “The Dutch economy will grow” and “This unicorn has three wings”. Statements of value (V) are evaluative judgments based on a definition or evaluation criteria. To this category belong statements such as “*The Corrections* is a great book”, “Downloading something without permission is not theft”. Finally, statements of policy (P) advocate performing an action. Examples of this type of statement are “Give me your phone” and “Let’s bring classical rhetoric back into the curriculum”.

For alpha arguments, the argument substance itself is defined as the combination of statement types expressed in the conclusion and the premise of the argument. The argument “This book is great because it was written by Javier Marías”, for example, has substance VF, because it combines a statement of value (V) in the conclusion with a statement of fact (F) in the premise.

As mentioned above, the set of values the parameter argument substance can take depends on the argument form. For beta arguments, it is based on another tripartite typology of statements consisting of singular statements (s), particular statements (p), and universal statements (u); for gamma arguments, on the identity (i) or difference (d) of the relationships between the subjects and the predicates; for delta arguments, on whether the premise is written from a first-person (I), a second-person (II), or a third-person (III) perspective.

The third and final parameter, the *argument lever*, is a verbal expression of how the premise renders the conclusion (more) acceptable. Although the lever often remains implicit in the discourse, it is verbalised in the identification procedure. When it is not explicitly expressed, the reader or listener must choose the best-fitting candidate, using the levers of the argument types distinguished in the PTA as a heuristic device. The argument “He must have been driving too fast because he left a trail of rubber on the road”, for example, has form *alpha* and substance FF, leaving the analyst with a choice from the levers with keywords ‘cause’, ‘effect’, ‘sign’, ‘correlation’, and ‘motive’. Of these possibilities, the lever “leaving a trail of rubber on the road is an effect of having driven too fast” is most consistent with our knowledge of the world. It is, therefore, probable that this is what the author or speaker intended, which allows the interpreter to add it to the discourse.

Since in most cases, the names of the argument types are derived from the keyword of the lever, the formulation of the lever usually also gives us a label for identifying the argument type. In the cases just mentioned, these could be the “argument from cause”, “argument from effect”, or “argument from sign”, for example.

To determine the type of any given argument in terms of the argument categorization framework of the PTA, the analyst can use the Argument Type Identification Procedure (ATIP) (Wagemans 2025). The procedure consists of three parts, which are subdivided into six steps, each of which contains detailed descriptions of how to find and interpret argumentative elements in discourse. In Part I – Standardization, the analyst performs an argumentative function analysis of the elements of the statements (Step 1) and rephrases the argument in its canonical form (Step 2). Subsequently, in Part II – Characterization, the analyst determines the three basic characteristics mentioned above: argument form (Step 3), argument substance (Step 4), and argument lever (Step 5). Finally, in Part III – Naming, the analyst labels the argument with a name that indicates its type in terms of the PTA and functions as a mnemonic aid for its characteristics (Step 6).

Our choice of the ATIP as the descriptive tier of identifying fallacies is in line with the aim of being maximally explicit about the justification of the interpretation of the discourse under scrutiny. As such, it com-

plies with the general aim of developing explainable argumentation theory (XAT) as mentioned in Section 1. In the next Section, we will explain the normative tier of our method for identifying fallacies.

5. The normative tier: The Comprehensive Assessment Procedure for Natural Argumentation (CAPNA)

The output of the ATIP described in the previous section is used as input for the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA), a procedure featuring several levels of analysis of arguments. This analysis allows the user to produce an assessment covering a wide range of aspects of argumentation traditionally considered in making acceptability judgements. That assessment is the basis for a final evaluation of the argument as acceptable or fallacious. The procedure is based on normative frameworks from logic, pragmatics, informal logic, linguistics and dialectics. This is why it is referred to as “comprehensive”, but its current state of development is not regarded as final and further additions are certainly possible: the perspective is three-dimensional, but not necessarily panoptical. In order to strike a balance between the need for systematicity and repeatability and the desire for a flexibility in practical application, the procedure has a small number of clearly defined and defended (Hinton 2021) basic principles which provide a stable framework within which there is much scope for individual refinement and prioritization. It is this combination which allows for both a clear delineation of fallacy types and the possibility for the elaboration and sub-division of those types.

The first and most important foundation stone of the procedure is Hinton’s (2021, p. 48) definition of argumentation as the expression of reasoning within a process. The definition is designed to be as wide-ranging as possible and encompass all areas of arguing, whilst remaining specific and providing the necessary structure for an assessment procedure. The definition highlights the three dimensions of an argument which are relevant to its acceptability: how it is phrased, what is its underlying reasoning, and whether it complies with the requirements of the discourse situation of which it is a part. These three areas are related to the classic division of rhetoric, logic, and dialectic, and lead to three levels of analysis within the procedure where concerns from those fields can be considered. No mention is made in the definition of

the goal or purpose of argumentation. This allows the CAPNA system to embrace any expression of reasoning within any discourse process and to maintain flexibility in terms of evaluation where functions of argument are non-standard. While the process involved may be a conversational exchange, it may also be such activities as advertising, self-reflection, speech-making, and book-writing, each of which has its own set of practices and expectations.

Secondly, while the term ‘argument’ can also be used in a great variety of other ways (see, e.g., Siegel & Biro 2021), within the CAPNA, it refers to one particular combination of premises and conclusion. When used in conjunction with the ATIP, this set is realised as a single premise supporting a conclusion (with a possibly unexpressed lever). The result is that the assessment emerging from the CAPNA applies to one particular instance of argument, and a negative or positive evaluation of that argument does not aspire to settle the wider point of contention either way. Either one expressed inference from a premise to a conclusion is acceptable within the discourse circumstances, meaning that it is accepted for consideration by the parties, or it is not: the truth of any asserted standpoint is still open to attack or support by other premise/conclusion sets in any configuration.

The third and final principle constituting the CAPNA framework is that the assessment is conducted by way of Procedural Questions (PQs) and that a negative judgement can only be made as a result of an unsatisfactory answer to one or more of those questions. If all PQs are answered satisfactorily, the argument is, by default, acceptable. When an argument fails a particular PQ it is said to have committed a fallacy named for that point of assessment. This then serves as our definition of a fallacy - the failure to navigate a given PQ. This has echoes of the approach advocated by Johnson and Blair (1993), discussed above in Section 3.2, but avoids the danger of ambiguity by linking fallacy judgements directly to specific questions. One crucial consequence of applying this assessment principle is that every fallacy judgement takes place at a named point in the process, yielding both a regularity in such judgements and an explicit point at which any disagreements may be addressed. Since one argument may fail at multiple points, it may be said to have committed multiple fallacies: an argument is not labelled with one fallacy name, as is the case with pattern-matching approaches, but is marked as having certain faults. This process of linking a fallacy

judgement with a particular point of evaluation is an important aspect of the explainability stressed in XAT. There is an obvious similarity between PQs and the well-established category of Critical Questions (CQs), but also an important difference. While CQs address aspects of the reasoning highlighted in an argument scheme, PQs cover a much wider terrain in terms of the context of the argument and the way in which it is expressed.

The CAPNA, then, has three dimensions of analysis which might lead to a negative acceptability judgement. Hence, it has an inherent division into three types of fallacy corresponding to those dimensions: of process, of reasoning, and of expression (or language). Further, five principles of good process are proposed: Pertinence, Proof (burden of), Productivity, Permissibility, and Politeness and the PQs for process are organised under these headings. This means that there are five categories of fallacies of process, and these may be further subdivided in accordance with the specific question which produces the negative assessment. PQs for the reasoning analysis are categorised as concerning the premise or the lever (warrant), so there are corresponding categories of premise fallacy and lever fallacy. Finally, there are also five principles of good language for arguments: Clarity, Consistency, Coherence, Completeness and Conceptualization (Hinton 2021, 2024). The fallacies of process, of reasoning, and of expression, are, therefore, subdivided into 12 types, and can then be further divided by specific question if necessary.

Although, for reasons of space, we do not list and discuss all the possible PQs here, an illustration will be helpful. If we take the example argument given above, “This book is great because it was written by Javier Marías”, the following would be of particular importance in the evaluation:

Process:

PQ 1: Does the argument address the topic at issue?

PQ 2: Does the argument contradict what has previously been agreed?

Reasoning:

PQ 3: Is the premise true? (Was the book indeed written by Javier Marías?)

PQ 4: Does the lever provide sufficient support for the inference to the conclusion? (Is being written by Javier Marías a relevant and sufficient criterion for a book's being great?)

Expression:

PQ 5: Can the conclusion be understood in more ways than one?

PQ 6: Are there emotional or evaluative terms used?

The answers to the first two questions listed will depend on the dialogical context, if the conversation is on quite a different subject, or it has already been established that the book is of poor quality, then the argument will be found to have committed a Pertinence fallacy. Assuming that the book is, in fact, by Marías, and no Premise fallacy is present, PQ 4 brings us to the crux of the argument, and will probably require further evidence in support of the lever. Consideration of the lever may already be informed by a recognition of the ambiguity in the word 'great'. If it is not, then PQs 5 & 6 should bring that to light and may necessitate a second look at PQ 4 or even a complete reworking of the argument: "This book is of importance to the literary canon because it was written by Javier Marías" or "This book is really enjoyable because it was written by Javier Marías", for instance. The vagueness of the word 'great' may result in a fallacy of Clarity, of Consistency, or of Completeness.

Within the three principles which form the CAPNA framework, there is much scope for flexibility and even creativity. The procedural questions are presented in a particular order for reasons described in Hinton (2021, p. 172), but in practically applying them, other than for a small number of follow-on questions, there is no reason why that order cannot be changed, allowing the analyst to jump straight to what looks most likely to be problematic. It is also the case that not all PQs produce black and white answers: in particular, the question of the strength of the inference will often require a judgement from the analyst as to just how strong it needs to be. That is one of the realities of human arguing and no system can remove the need for such decision-making. The acceptability judgement, therefore, is not equivalent to a division into good and bad arguments: there is wide scope for the recognition of varying argument strengths. Finally, and most importantly for our purposes here, although the definition of argumentation sets out the three main dimensions of evaluation, within those strands the list of PQs is open and can be added to. The current divisions are the result of extensive

work considering a broad range of scholarship, but we make no claim that every possible way for an argument to be unacceptable is covered. However, the addition of new questions, even of entire sub-categories, and hence new types of fallacy, would only strengthen the comprehensiveness of the procedure, not weaken its foundations.

6. Implications for fallacy theory

Taking a two-tier procedural approach to both the description and the evaluation of arguments has wide-ranging effects for our understanding of the notion of fallacy. The procedure clears away much of the vagueness and subjectivity surrounding the concept of fallaciousness and provides more solid grounds for the defence of particular judgements on the acceptability of arguments than more *ad hoc* approaches can boast. While this paper focuses on theoretical aspects, examples of how the approach works in practice can be seen in earlier work (Hinton & Wagemans 2022, 2023). The advantages of the procedural approach in general, and those associated with the procedures presented in this paper in particular, are set out in full below.

There are, we suggest, five important advantages of the proposed system. Most importantly, the clear separation of the identification and evaluation of arguments resolves several of the key problems with earlier conceptualizations and categorizations of fallacies. In addition, it provides a clear account of the nature of fallacy through an argumentification of the notion; it yields a systemization of the identification of informal fallacies thus allowing for the making of justifiable acceptability judgements on formally invalid arguments; and it reveals the deeper character of fallaciousness which is hidden behind the common fallacy names. Finally, due to its procedural nature, this system is potentially adaptable for use in the technological applications which are of ever-growing importance in argumentation study and practice.

The separation of descriptive and evaluative analysis of arguments

The fundamental issue which we have set out to highlight and to resolve is the conflation of descriptive and evaluative analyses in much of fallacy theory. By doing so we hope to avoid the problems highlighted above, in particular the ambiguity which comes when the two steps merge into one. We believe that while we are not the first by any means

to point out the need to avoid this confusion, the proposals we make provide practical tools for its elimination in argument analysis, which have hitherto been absent. We suggest that the principle of separation is more important than the specifics of the procedures we describe; even where one or both of those procedures is rejected, in part or in whole, the rationale for their creation remains. It is to be hoped that debate will shift towards a discussion on what the exact nature of the processes of description and evaluation should be, and away from approaches which avoid or ignore one of those two tiers, or, what is worse, conflate them into one step of analysis.

Argumentification of fallacy

By applying the term ‘fallacy’ only in cases of the negative evaluation of an argument, we make it clear that fallaciousness is a quality of arguments alone, and thus resolve the diverse reference and idealization problems. This goes, in particular, for teleological approaches to fallacies taking the behaviour of the arguer as the object of their fallacy judgements. Behavioural fallacies, we believe, are both too broad in their possible range, and too narrow in their applicability, and therefore should be excluded from discussions of fallacy in argumentation. Any behaviour which does not further the purpose of an argumentative discourse may be labelled a behavioural fallacy, bringing a heterogeneous collection of phenomena under the fallacy banner and making further discussion of fallacies fraught with the possibility of category errors. At the same time, the specification of a purpose rules out much activity in which arguments are applied and within which their evaluation may be required. Argumentification, therefore, preserves both simplicity and scope for fallacy theory.

Systematic identification of informal fallacies

The use of a systematic procedure for both the description and the evaluation of arguments allows for the retention of the importance of contextual aspects in the study of argument acceptability, without making claims of fallaciousness purely subjective and arguer-dependent. Scholars working from an epistemological perspective have long criticized those working within a pragma-dialectical framework for the latter’s reliance on intersubjective agreement - see Garssen (2024) for the latest

iteration of this debate - but have failed to provide sufficiently developed tools to enable fully-explainable, extra-subjective acceptability judgements. We believe that the proposals in this paper do much to fill that gap and can alleviate the problems of assignation and underdetermination. We provide a system which is formalized and repeatable but goes beyond the constraints of formal logic and the concept of validity. It allows for fallacy judgements to be made by both those involved in a dispute and external analysts, following the same procedure. The determination of acceptability and unacceptability for formally invalid arguments is made explicit in a way which has not been previously achieved.

The character of fallaciousness

The clear separation of the two tiers of assessment allows for a better view into the nature of fallaciousness. It is not a similarity to some pre-described form which can be identified through a descriptive analysis, as the same form may be instantiated in either a fallacious or an acceptable argument under different circumstances. Rather, it is a failure to comply with a certain characteristic which acceptable arguments must have. The word fallacy is used in a judgemental way; it must therefore be part of the evaluation of the argument. It makes little sense to discuss fallacies without clarifying what exactly it is that is unacceptable about them. The traditional, named fallacies are highly problematic on this score: in most cases, it is allowed that particular argument structures are sometimes acceptable, sometimes not. When they are not, they are labelled as fallacies, but the type of fallacy is dependent on the form. For instance, every argument of a certain form which is rejected will be a “slippery slope” fallacy, but that label tells us nothing about what is actually wrong with the particular argument. The CAPNA system of procedural questions means that every fallacy judgement is accompanied with an explicit statement about the problem which has been uncovered: if the expression is too vague, we have a Clarity fallacy, if the argument is circular, we have a Productivity fallacy, if the starting assumptions are false, we have a Premise fallacy, to give some examples. This leads us to the theoretical insight that fallacies are defects, not forms. They are, therefore, best categorized and studied from that perspective.

Automated fallacy detection

We are very aware that in spite of the advantages we have noted above, the procedures for characterizing and evaluating arguments set out in this paper are relatively complicated and that they run counter to the very human desire to rely on reasoning heuristics and ad hoc judgements based on experience and intuition. A certain amount of training and practice is necessary to employ both the ATIP and the CAPNA with facility, and a degree of reluctance in committing to that is understandable, especially amongst those who find they can usually detect common traditional fallacy types at a glance. We believe that the theoretical insights we have put forward retain their value even for those who are unwilling to adopt the tools we suggest for practical implementation, and also that those tools are ripe for development as the basis for automated fallacy detection systems to be used in artificially intelligent software applications. The procedures we describe are not yet algorithms, but can be viewed as a step towards pseudo-algorithms with a potential for integration into AI text generators (see Hinton & Wagemans 2023). In this context, worries over the laborious and thus time-consuming nature of the systematic analysis we suggest disappear, and those very elements which are off-putting for a human analyst - the careful following of a step-by-step procedure - become the greatest advantages for the programmer. Such a technological implementation would be of use both in the generation of fallacy-free text and in the automated evaluation and annotation of inputted texts. The availability of annotation of this sort would have an enormous impact on the study of argumentation, in particular on the increasingly popular use of large-scale corpora (e.g. Visser et al. 2020, Greco 2023, Younis et al. 2023) and argument mining (Lawrence & Reed 2020, Arora et al. 2023).

7. Conclusion

In this paper, we have provided an overview of the established approaches to fallacy theory, suggesting that they can be divided into four main groups which we name Formal, Informal, Teleological, and Cognitive. We have described their main characteristics in Section 2, focusing on answering questions as to what object is evaluated under which conditions and by which (type of) criteria. Whilst recognising the enor-

rious contributions each of these traditions has made in our understanding of arguments, in Section 3 we have also discussed particular weaknesses of their conceptualization of fallacies. To various extents, the four approaches suffer from one or more of the following problems: the names of fallacies (1) are ambiguous in various ways, (2) are assigned without indicating a reason or for unclear reasons, (3) underdetermine what is wrong with the argument, and (4) refer to diverse phenomena. Further, (5) the identification of the argument type is based on its evaluation, and (6) evaluations are based on idealizations. In our view, all these difficulties relate to a conflation of the descriptive and evaluative dimensions of fallacy identification and the lack of distinct procedures for each of them. In Sections 4 and 5, we describe the elements of our own two-tier system of analysis which we suggest avoids the problems of other approaches, not by rejecting their fundamental assumptions about argumentation, but by recasting the notion of fallacy. Our approach to fallacies has the advantages of unifying the object of study by argumentifying the concept, systematizing the process of identification, clarifying the essence of fallaciousness, and, most importantly, explicitly and comprehensively separating the descriptive and evaluative aspects of that identification.

Whilst we acknowledge that the details of the systems described may be criticised and improved upon in further work, we are confident that the principles which have led to their creation are clear and well-justified. Among the topics for further research is a comprehensive comparison between the traditional approaches to fallacies and the evaluations by means of our two-tier approach. While a list of fallacies has been mapped onto negative evaluations of procedural questions (Rentier 2023), a more complete inventory of differences and commonalities is still to be produced. In particular, it would be interesting to examine to what extent existing descriptions of fallacies - ‘the old gang’ - corresponds to the argument flaws that can be detected with our method. What could be learned from such comparative research is what exactly is wrong with an argument that is labelled in the old fashioned way and on which level (process, reasoning, language) the flaw can be situated.

Acknowledgements: This paper has been supported by a grant from the Scientific Development Fund of the Faculty of Philology, University of Lodz, and by a grant from the Polish National Science Centre (NCN): 2022/47/B/HS1/01725.

References

- Arora, Sakshi, Ajay Rana, and Archana Singh. 2023. Argument Mining: A Categorical Review. In *Modern Electronics Devices and Communication Systems: Lecture Notes in Electrical Engineering*, eds. Agrawal, R., Kishore Singh, C., Goyal, A., Singh, D.K., vol 948, 353-367. Singapore: Springer. https://doi.org/10.1007/978-981-19-6383-4_28
- Boudry, Maarten, Fabio Paglieri, and Massimo Pigliucci. 2015. The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation* 29: 431-456.
- Copi, Irving. and Carl Cohen. 1990. *Introduction to logic*. 10th ed. Prentice Hall.
- Davies, Richard. 2020. In defence of a fallacy. *Studia Semiotyczne* 34(2): 25-42.
- Davies, Richard. 2023. Locke and “ad”. *Argumentation* 37: 473–492.
- van Eemeren, Frans, H. and Rob Grootendorst, Rob. 2004. *A Systematic Theory of Argumentation*. Cambridge University Press.
- van Eemeren, Frans H., Bart J. Garssen, Anne Snoeck Henkemans, Bart Verheij, and Jean H.M. Wagemans. 2014. *Handbook of Argumentation Theory*. Springer.
- Garssen, Bart. J. 2024. A Reaction to Critique from the Epistemological Sidelines. *Informal Logic* 43(4): 527-542. <https://doi.org/10.22329/il.v43i4.8415>
- Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.
- Govier, Trudy. 2010. *A Practical Study of Argument*. 7th ed. Belmont, CA: Wadsworth.
- Greco, Sara. 2023. Twitter Activists’ Argumentation Through Subdiscussions: Theory, Method and Illustration of the Controversy Surrounding Sustainable Fashion. *Argumentation* 37: 1–23. <https://doi.org/10.1007/s10503-022-09579-1>
- Hamblin, Charles L. 1970. *Fallacies*. Methuen.
- Hansen, Hans V. 2002. The straw thing of fallacy theory: The standard definition of ‘fallacy’. *Argumentation*, 16, 133-155.
- Hansen, Hans V. 2023. Fallacies. In *Stanford Encyclopedia of Philosophy: Spring 2023 edition*, eds. E. Zalta, U. Nodelman. URL = <https://plato.stanford.edu/archives/spr2023/entries/fallacies/>
- Herman, Thierry and Steve Oswald, eds. 2014. *Rhetoric & Cognition: Theoretical Perspectives and Persuasive Strategies*. Peter Lang.
- Hinton, Martin. (2018). Slippery Slopes and Other Consequences. *Logic and Logical Philosophy*, 27 (4), 453-470.

- Hinton, Martin. (2021). *Evaluating the Language of Argument*. Springer.
- Hinton, Martin. (2024). Argumentation and Identity: A Normative Evaluation of the Arguments of Delegates to the COP26 UN Climate Change Conference. *Argumentation* 38, 85-108. <https://doi.org/10.1007/s10503-022-09589-z>
- Hinton, Martin, & Jean H.M. Wagemans. (2022). Evaluating reasoning in natural arguments: A procedural approach. *Argumentation*, 36, 61-84.
- Hinton, Martin & Jean H.M. Wagemans, (2023). How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument & Computation*, 14(1), 59-74.
- Johnson, Christopher M. 2009. Reconsidering the Ad Hominem. *Philosophy* 84: 251-266.
- Johnson, Ralph & John A. Blair. 1993. *Logical Self-Defence*. 3rd edition. Toronto: McGraw-Hill Ryerson.
- Lawrence, John and Christopher Reed. 2020. Argument mining: A survey. *Computational Linguistics* 45(4): 765-818.
- Lumer, Christoph. 2005. Introduction. The epistemological approach to argumentation. A map. *Informal Logic* 25(3), 189-212.
- Mercier, Hugo and Daniel Sperber. 2011. Why do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences* 34(2): 57-74.
- Mill, John S. 1882. *A System of Logic, Ratiocinative and Inductive*. 8th ed. Harper & Brothers.
- Paglieri, Fabio, Laura Bonelli, and Silva Felletti, eds. 2016. *The Psychology of Argument: Cognitive Approaches to Argumentation and Persuasion*. College Publications.
- Rentier, Roosmarjin. 2023. *Connecting Procedural Questions and Fallacy Labels: A Heuristic Approach Using ChatGPT*. MA Thesis Linguistics and Communication, University of Amsterdam.
- Rizzo, Mario and Douglas Whitman. 2003. The Camel's Nose in the Tent: Rules, Theories and Slippery Slopes. *UCLA Law Review* 51: 539-592.
- Siegel, Havey and John Biro. 2021. Walton on Argument, Arguments, and Argumentation. *FLAP* 8(1): 183-194.
- de Swart, Harrie. 2018. *Philosophical and Mathematical Logic*. Cham: Springer.
- Visser, Jacky, John Lawrence, Christopher Reed, Jean H.M. Wagemans, and Douglas Walton. 2020. Annotating argument schemes. *Argumentation* 35: 101-139. DOI: <https://doi.org/10.1007/s10503-020-09519-x>
- Wagemans, Jean H.M. (2016). Constructing a Periodic Table of Arguments. In P. Bondy & L. Benacquista (Eds.), *Argumentation, Objectivity, and*

- Bias: Proceedings of the 11th International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, 18-21 May 2016 (pp. 1-12). Windsor, ON: OSSA.
- Wagemans, Jean H.M. (2018). Assertoric syllogistic and the Periodic Table of Arguments. In S. Oswald & D. Maillat (Eds.), *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation, Fribourg 2017*. (Volume I, pp. 573-588). College Publications.
- Wagemans, Jean H.M. (2020). Why missing premises can be missed: Evaluating arguments by determining their lever. *OSSA Conference Archive*, 1, 1-16. <https://scholar.uwindsor.ca/ossaarchive/OSSA12/Saturday/1>
- Wagemans, Jean H.M. (2022). The Philosophy of Argument. In P. Stalmaszczyk (Ed.), *The Cambridge Handbook of the Philosophy of Language* (pp. 571-589). Cambridge: Cambridge University Press.
- Wagemans, Jean H.M. (2023). How to identify an argument type? On the hermeneutics of persuasive discourse. *Journal of Pragmatics*, 203, 117-129. <https://doi.org/10.1016/j.pragma.2022.11.015>
- Wagemans, Jean H.M. (2025). Argument Type Identification Procedure (ATIP) – Version 5. Published online March 25, 2025. URL = www.periodic-table-of-arguments.org/argument-type-identification-procedure
- Walton, Douglas. 1992a. *Plausible argument in everyday conversation*. State University of New York Press.
- Walton, Douglas. 1992b. Nonfallacious arguments from ignorance. *American philosophical quarterly* 29(4): 381-387.
- Walton, Douglas. 1998. *The New Dialectic: Conversational Contexts of Argument*. University of Toronto Press.
- Walton, Douglas. 2006. Examination dialogue: An argumentation framework for critically questioning an expert opinion. *Journal of Pragmatics* 38(5): 745-777.
- Walton, Douglas. 2011. Defeasible reasoning and informal fallacies. *Synthese* 179: 377-407.
- Walton, Douglas. 2015. The Basic Slippery Slope Argument. *Informal Logic* 35(3): 273-311.
- Walton, Douglas, Christopher Reed and Fabrizio Macagno. 2008. *Argumentation schemes*. Cambridge University Press.
- Whatley, Richard. 1857. *Elements of Logic*. 9th ed. John Parker and Son.
- Woods, John. 2007. The concept of fallacy is empty: a resource-bound approach to error. In *Model-based reasoning in science, technology, and medicine*, L. Magnani, P. Li, eds., 69-90. Springer.
- Wreen, Michael. 2003. Arguments from Ignorance and the Presumption of Ignorance. *Logique et Analyse* 46(183-184): 365-382.

Younis, Ramy, Daniel de Oliveira Fernandes, Pascal Gygax, Marcin Koszowy, Steve Oswald. 2023. Rephrasing is not arguing, but it is still persuasive: An experimental approach to perlocutionary effects of rephrase. *Journal of Pragmatics* 210: 12-23.

Zenker, Frank, Jan Albert van Laar, Bianca Cepollaro, Anca Gâță, Martin Hinton, Colin Guthrie King, Brian Larson, Marcin Lewiński, Christoph Lumer, Steve Oswald, Maciej Pichlak, Blake Scott, Mariusz Urbański, and Jean. H.M Wagemans. 2024. Norms of public argumentation and the ideals of correctness and participation. *Argumentation* 38: 7–40. <https://doi.org/10.1007/s10503-023-09598-6>