



Evaluating Reasoning in Natural Arguments: A Procedural Approach

Martin Hinton¹ · Jean H. M. Wagemans²

Accepted: 14 September 2021 / Published online: 30 September 2021
© The Author(s) 2021

Abstract

In this paper, we formulate a procedure for assessing reasoning as it is expressed in natural arguments. The procedure is a specification of one of the three aspects of argumentation assessment distinguished in the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) (Hinton, 2021) that makes use of the argument categorisation framework of the Periodic Table of Arguments (PTA) (Wagemans, 2016, 2019, 2020c). The theoretical framework and practical application of both the CAPNA and the PTA are described, as well as the evaluation procedure that combines the two. The procedure is illustrated through an evaluation of the reasoning of two example arguments from a recently published text.

Keywords Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) · Argument evaluation · Natural argumentation · Periodic Table of Arguments (PTA) · Reasoning

1 Introduction

Reasoning is an essential part of many forms of discourse, and is of particular importance in the spheres of science and policy debate, where it is expressed in argumentation aimed at convincing others of the acceptability of a particular point of view. As poor or deceptive argumentation can have serious consequences for the lives of us all, developing methods for evaluating reasoning in argumentation is of great importance. Methods centring around the notion of ‘fallacy’ can be called ‘informal’ since they are based on criteria that do not require the evaluator to first

✉ Martin Hinton
Martin.hinton@uni.lodz.pl

Jean H. M. Wagemans
j.h.m.wagemans@uva.nl

¹ University of Lodz, Lodz, Poland

² Amsterdam Center for Language and Communication (ACLC), University of Amsterdam, Amsterdam, The Netherlands

provide a completely formal description of the characteristics of the argument or behaviour under scrutiny. They can also be considered ‘vague’, since the term ‘fallacy’ comprises negative judgments about a great many different aspects of argumentative discourse, varying from the acceptability of an individual argument to the reasonableness of the behaviour of participants in an argumentative exchange. And from a methodological point of view, they can be characterised as ‘comparative’ since they are based on the evaluator’s finding some degree of similarity in the comparison of an ideal argument with a real one or idealized behaviour with real behaviour (Wagemans 2020c: pp. 1–3).

This paper aims to contribute to the development of a less informal and comparative approach to argumentation assessment by providing a procedure for evaluating the underlying reasoning of individual arguments. It does so by combining insights from the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) set out in Hinton (2021) and the argument categorisation framework of the Periodic Table of Arguments (PTA) as presented in Wagemans (2016, 2019, 2020c). The combination of these two theoretical frameworks feeds a specification of the various steps of a procedure for the assessment of the reasoning aspect of individual arguments.

The following two sections describe the relevant insights from the two frameworks, CAPNA (Sect. 2) and PTA (Sect. 3), which we mean to utilise in designing our procedure for assessing the reasoning underlying an argument. Given that such reasoning varies with the type of argument under scrutiny, we first detail in Sect. 4 the steps taken in order to identify the argument type. Section 5 then provides a full description of how the reasoning assessment itself is to be conducted, with an illustration of the various steps of that assessment procedure and a list of procedural questions by which it is to be navigated. The subsequent Sect. 6 gives an example of the evaluation of arguments in operation, which draws out different features of the procedure, showing the power and effectiveness of its systematicity. In Sect. 7, which concludes the paper, we reflect upon the methodological status of the developed method. While stressing that the assessment still requires the use of subjective judgements, we argue that our procedural method supplies a clear framework within which those judgements should be made.

2 Comprehensive Assessment Procedure for Natural Argumentation

The procedure for the assessment of reasoning in argumentation which we propose in this paper is based on the three-part definition of argumentation put forward and extensively discussed in Hinton (2021: pp. 45–58), where it is defined as:

The expression of reasoning within a process.

This definition highlights the relationship between reasoning and argumentation as it clarifies that arguments are a linguistic realization of reasoning, and that that realization when placed in the context of a communicative process constitutes the act of argumentation. It differs from those put forward by other theorists in that no

aim, motivation, or method for arguing is proposed: it is designed to be as general as possible, encompassing any form of discursive reason giving.¹ The given definition is not suggested as a replacement for those proposed previously: every author has an individual perspective on argumentation and focus in his research of it, but it is one which we find lends itself well to the construction of a tool for the analysis and evaluation of all aspects of naturally occurring argumentation. That is partly due to its generality, which allows the tool to be applied in any situation where an argument is found, and also because the definition clearly identifies three components of argumentation which can then be considered separately in the evaluation procedure: the Process, which refers to the discursive context; the Reasoning, which refers to the inference of conclusion from premises, and is the focus of this paper; and the Expression, which will generally mean the linguistic realization of that reasoning, but might also include visual representation, for example. It is important to stress here that while the definition of argumentation offered and the division into three aspects for evaluation lie behind the structure of the CAPNA, the employment of the procedure itself is not reliant on the complete acceptance of these theoretical assumptions: the definition is deliberately broad, but the tool does not preclude a far narrower conception of what argumentation is on the part of whoever is applying it.

The CAPNA is a procedure designed to assess all aspects of argumentation. Here, after giving a general overview of the procedure as a whole, we present in detail those parts relevant for assessing the reasoning. We believe that the CAPNA differs from other attempts at putting forward criteria for the assessment of arguments in that it offers a clearly defined, reproducible procedure to be followed in that assessment.

The assessment is conducted as illustrated in Fig. 1 below, and progresses via the posing of a series of procedural questions which establish whether an outright error or potential weakness is contained within the argument under evaluation.

The procedure applies only to individual arguments, understood as premise-conclusion relation sets, therefore, in the first stage, a text which contains an “Apparent argument” (indicated with a squared box) is subjected to an “Initial analysis” (indicated with a round-cornered box). This determines whether there is, in fact, an argument present and reconstructs it as a token of an “Argument type” (indicated with a squared box) as described in Sect. 4 below, on the basis of the form of reasoning employed. It is likely that a degree of reconstruction and reformulation will be necessary at this stage. Although there is a level labelled as Language analysis, it is important to note that this refers to a deeper consideration of the text, largely, though not exclusively, of its semantics: at all points, a degree of linguistic assessment is necessary in order to make any evaluation. This ordinary language competence in the evaluator is taken for granted.

¹ Examples of such definitions are ‘Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint’ (van Eemeren & Grootendorst 2004: p. 1) and ‘[A]ll argumentation aims at gaining the adherence of minds, and, by this very fact, assumes the existence of an intellectual contact’ (Perelman & Olbrechts-Tyteca 1969: p. 14). For a general comparison of classical and present-day definitions of argumentation, see Wagemans (2019: pp. 58–59).

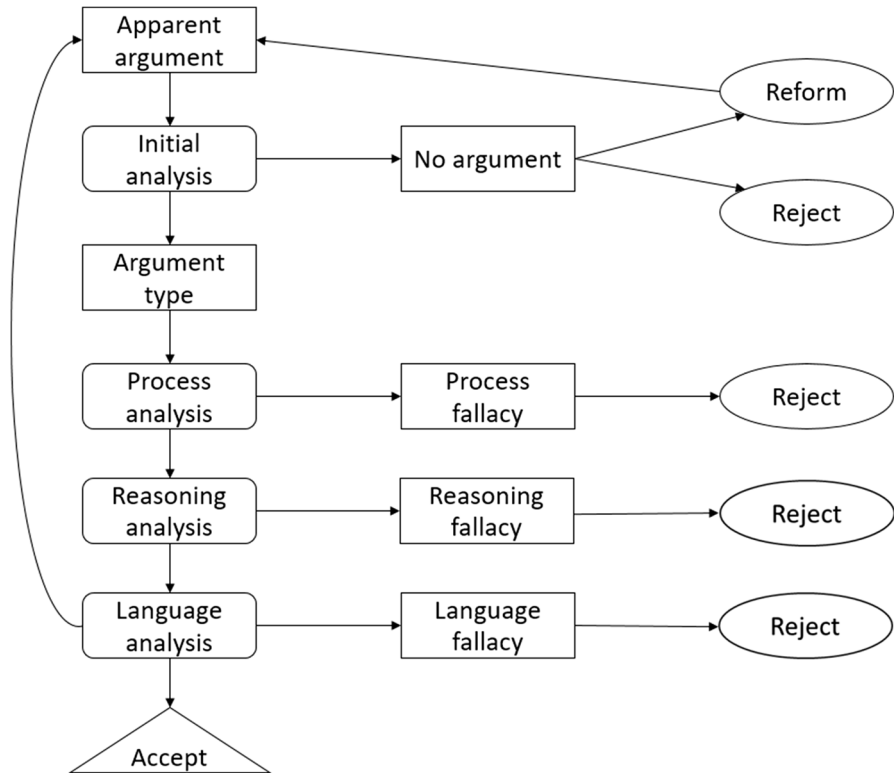


Fig. 1 The Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) (Hinton 2021: p. 169)

An experienced assessor may immediately recognise where an argument is likely to be found wanting and move quickly from the “Initial analysis” to the relevant place in the procedure: there is no hard requirement for the steps to be followed in the order described, although, of course, if the argument were not found to fail at that point, the assessment would pick up again from where it left off. The “Initial analysis” includes a determination of the “Argument type” within the framework of the Periodic Table of Arguments (Wagemans 2016, 2019, 2020c), which can then go on to the level of “Process analysis”. The procedural questions at this stage are to some degree relative to the context of the argumentation: courts of law have different criteria of admissibility from friendly deliberations.

Once it has been established that an argument is acceptable within the ongoing process of argumentation, its underlying reasoning can be assessed. This procedure is carried out on the basis of the determined argument type and is described in detail in Sects. 4 and 5 below.

If the arguer is found to have committed neither a Process fallacy nor a Reasoning fallacy, then the final stage is the Language analysis which is conducted using the Informal Argument Semantics (IAS) developed and described at length

in Hinton (2021). It should be reiterated at this point that IAS is a form of deep linguistic assessment designed to analyse and evaluate meaning at a level beyond what would be expected as a matter of course from a normally competent language user. The aim of the Semantics is to draw out subtle, hidden characteristics of the language of the text, such as emotional or evaluative connotations, implicatures and textual argumentativity, as well as a range of ‘philosophical’ misuses of language, such as persuasive definition and the fetishisation of language.² These linguistic characteristics are not discussed in this paper, but it should be noted that they may often cause a reassessment of the reasoning as the full meaning of the terms used becomes apparent.

There is a sense in which all fallacies of expression are fallacies precisely because they obscure or invalidate the reasoning of the argument (see Hinton 2021: p. 129): the separation of the three elements of argumentation for the purposes of the procedure should not be taken to suggest that there are strong barriers between those elements preventing mutual influence.

To forestall possible criticisms, it should be noted that when a weakness is found in an argument it does not mean automatic rejection: it may lead instead to a qualified acceptance and further investigation. Indeed, even when serious flaws, catastrophic to the success of the argument, are discovered, the evaluator may choose to continue the assessment, in order to have a fuller understanding of the various aspects of the argument. The presence of one fallacy is sufficient for acceptance to be withheld, but uncovering the presence of several would put an opponent of the argument in a far stronger dialectical position.

It is also worth pointing out the asymmetry between the acts of acceptance and rejection. Rejection is taken to mean that the argument is flawed, while acceptance means merely that it has not yet been found to be flawed: that is to say, there is always presumptive, not final, acceptance. Any argument which has passed through the entire procedure without having any serious fault exposed cannot be dismissed, but it is not necessarily decisive—similarly acceptable counter-arguments may exist. The assessment procedure is designed to evaluate natural arguments: it cannot guarantee that issues are actually settled or disagreements resolved.

Finally, the term ‘fallacy’ is used throughout the procedure, but this should not be confused with the general, somewhat vaguely defined, meaning often given to the word. We do not assess argumentation by looking for fallacies which have previously been identified, named and described: instead, arguments are found to be ‘fallacious’ because they fail at certain points in the assessment procedure. They may be found to be fallacious at more points than one, which raises no difficulty as we make no attempt to label individual arguments as examples of one particular fallacy in one particular category. The naming of fallacies, to the extent to which we indulge in that

² Figure 1 is a slight simplification and there are, in fact, three possible outcomes of the language analysis: if there is a clear failure in the argument, it is rejected as containing a language fallacy; if there is a failure which can be rectified by a reformulation of the argument, or there are significant implicatures which change the nature of the argument, then it can be returned to the beginning of the assessment procedure to be evaluated afresh.

activity, is a naming of what has been discovered to be wrong with a certain aspect of the argumentation at a certain level in the evaluation. Thus, the three basic groups of fallacies are those of Process, of Reasoning, and of Expression, not because we believe that all the known fallacies can be placed under one of those headings, but because those are the three aspects of argumentation which are assessed in the general procedure, and, thus, every fallacy is discovered at one of those stages of analysis. To a degree, we follow here the example of pragma-dialectics, in which fallacies are defined as violations of rules of the ‘code of conduct for reasonable discussants’ and are not necessarily given names taken from the traditional lists (van Eemeren and Grootendorst 2004: pp. 158–186). There are, however, important methodological differences between our proposal and that of the pragma-dialecticians. Firstly, we provide a step by step procedure for a third party to follow in evaluating an argument, not a set of rules governing the behaviour of disputants. Secondly, that procedure is based on norms for arguments, rather than rules for arguers, thus returning the notion of fallacy to its roots as a quality belonging to an argument rather than a discourse move. Thirdly, regarding these arguments, we provide a more fine-grained method of analysis that is based on a formal(izable) classification of argument types rather than an arbitrary choice of three argumentation schemes. Thus, we follow the pragma-dialectical idea that a fallacy is the breaking of a norm, rather than a named argument type, but the way in which we approach that normative evaluation is quite different.

3 The Periodic Table of Arguments

Having explained the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA), we turn to describe the second framework we use for designing a procedure for assessing the reasoning aspect of natural argument: the argument categorisation framework of the Periodic Table of Arguments (PTA).³ Using this framework enables a specification of the Reasoning level analysis of CAPNA. At the same time, while the PTA so far has been mainly used as an analytical tool for argument type identification and annotation (Visser et al. 2018, 2021; Gobbo et al. 2019) its use in specifying parts of the comprehensive assessment procedure articulates how it can also be used as an evaluative tool.

The categorisation framework of the PTA has in common with logical taxonomies of argument that it takes logical form as a defining characteristic of an argument type. However, it also differs from these taxonomies in that logical form is not the only characteristic taken into account. Inspired by the classical dialectical and rhetorical taxonomies of arguments, which define argument types on the basis of their content rather than form, two other characteristics are added to the theoretical framework of the PTA: argument substance and argument lever. Since the

³ The explanations of the aspects of the PTA in this section are based on Wagemans (2016, 2019, 2020c). For updates, example analyses, and associated research projects see www.periodic-table-of-arguments.org.

PTA defines an argument type as the specific combination of these three parameters (form, substance, and lever), it can be characterized as a ‘combinatorial’ or ‘factorial’ taxonomy of arguments.

Within the theoretical framework of the PTA, an argument is conceptualized as a combination of two statements—a conclusion and a premise. It is further assumed that an arguer puts forward the premise in order to support the conclusion, i.e., to make the conclusion (more) acceptable in the eyes of the addressee. In other words, when viewed from a pragmatic perspective, the arguer aims at changing the epistemic status of the conclusion from ‘doubted’ to ‘accepted’. We refer to this projected change in epistemic status of the conclusion with the term ‘acceptability leverage’.

To explain how the acceptability leverage from the premise to the conclusion works, the PTA assumes the ‘law of the common term’. This law states that the premise, in order to fulfil its pragmatic aim of rendering the conclusion (more) acceptable, should share exactly one common term with the conclusion. While this common term functions as the ‘fulcrum’ of the leverage of acceptability taking place within the argument, the relationship between the non-common terms, which expresses the underlying mechanism of the argument, functions as its ‘lever’.

The law of the common term yields two basic possibilities of argument forms. If the statements share the same subject, the argument has the form ‘*a is X, because a is Y*’ and is characterised as a ‘predicate argument’ (pre). In this case, the subject (*a*) functions as the fulcrum and the relationship between the predicates (*Y* and *X*) as the lever of the argument. A concrete example is *Unauthorized downloading (a) is not theft (X), because unauthorized downloading (a) does not deprive the original owner of the use of an object (Y)*, which has *unauthorized downloading (a)* as its fulcrum and the relationship between *does not deprive the original owner of the use of an object (Y)* and *is not theft (X)* as its lever.

The other basic possibility is when the common term is the predicate, which means the argument has the form ‘*a is X, because b is X*’. In this case, the predicate (*X*) is the fulcrum and the leverage of acceptability can be explained by assuming that there is some kind of relationship between the non-common terms of the premise and the conclusion, namely their subjects (*a* and *b*). Within the framework of the PTA, such arguments are called ‘subject arguments’ (sub). An example is *Cycling on the grass (a) is prohibited (X), because walking on the grass (b) is prohibited (X)*, which has *is prohibited (X)* as its fulcrum and the relationship between *cycling on the grass (a)* and *walking on the grass (b)* as its lever.

In natural argumentative discourse, any statement can be expressed as a proposition or as an assertion. The difference between the two modes of expression is that in the latter, the arguer’s doxastic attitude regarding the statement is explicitly present in the discourse. The statement *The president is doing a great job*, for example, is expressed as a proposition, while the statement *I believe that the president is doing a great job* is expressed as an assertion. While both statements contain the proposition *the president is doing a great job*, the assertion additionally contains the doxastic attitude marker *I believe that* (see Fig. 2).

Within the theoretical framework of the PTA, the distinction between propositions and assertions is used to characterise arguments as ‘first-order arguments’ (1)

proposition	<i>The president is doing a great job</i>
assertion	<u><i>I believe that the president is doing a great job</i></u>

Fig. 2 The difference between a proposition and an assertion

or ‘second-order arguments’ (2). If the propositions of the statements share a common subject or predicate, as in the examples above, the argument is characterised as a first-order predicate argument (1 pre) or first-order subject argument (1 sub) respectively. If the statements have the proposition of the conclusion as their common term, the argument has the form ‘ q is T, because q is Z’ – ‘T’ standing for ‘true’, a standard formulation of the doxastic attitude marker that may or may not have been expressed in the actual discourse and can be added or substituted by the analyst. Such a ‘second-order predicate argument’ has the shared proposition (q) as its fulcrum, while the leverage of acceptability can be explained by assuming that there is some kind of relationship between the predicate of the premise (Z) and that of the conclusion (T). An example is *We only use 10% of our brain (q) is true (T), because we only use 10% of our brain (q) is said by Einstein (Z)*, which has *we only use 10% of our brain (q)* as its fulcrum and the relationship between *is said by Einstein (Z)* and *is true (T)* as its lever.

If the statements contain different propositions, they have the doxastic attitude marker as their common element and the acceptability leverage is based on a relationship between the propositions. Such arguments are called ‘second-order subject arguments’ and have the form ‘ q is T, because r is T’. An example is *He must have gone to the pub (q) is true (T), because the interview is cancelled (r) is true (T)*.

As said above, the theoretical framework of the PTA takes the conclusion and the premise of an argument to be expressed by statements. The third characteristic of arguments that constitutes this framework is the so-called ‘argument substance’, i.e., the specific combination of types of statements. This characteristic is determined on the basis of a widely used typology of statements that is developed in debate theory and distinguishes between statements of fact (F), statements of value (V), and statements of policy (P) (see, e.g., Broda-Bahm et al. (2004), Skorupski (2010), Freeley and Steinberg (2014)). An argument can thus be said to substantiate one of nine possible different combinations of types of statements, conventionally starting with the type of statement expressed in the conclusion followed by that in the premise: PP, PV, PF, VP, VV, VF, FP, FV, FF. *The government should invest in jobs, because this will lead to economic growth*, for instance, can be characterized as a PF argument, since it combines a statement of policy (P) in its conclusion with a statement of fact (F) in its premise.

In sum, in identifying the type of argument, the analyst should classify it as (1) a first-order or second-order argument; (2) a predicate or subject argument; and (3) as one out of nine possible combinations of types of statements. The superposition of these three partial characterizations yields the systematic name of the argument. In order to illustrate this notion, we provide in Table 1 the systematic name of the examples given for each of the four basic argument forms.

Table 1 Systematic names of examples following the four basic argument forms

Example	Argument form	Argument substance	Systematic name
<i>Unauthorized downloading (a) is not theft (X) (V), because unauthorized downloading (a) does not deprive the original owner of the use of an object (Y) (F)</i>	$a \text{ is } X, \text{ because } a \text{ is } Y$	VF	1 pre VF
<i>Cycling on the grass (a) is prohibited (X), because walking on the grass (b) is prohibited (X)</i>	$a \text{ is } X, \text{ because } b \text{ is } X$	VV	1 sub VV
<i>He must have gone to the pub (q) is true (T), because the interview is cancelled (r) is true (T)</i>	$q \text{ is } T, \text{ because } r \text{ is } T$	VV	2 sub VV
<i>We only use 10% of our brain (q) is true (T), because we only use 10% of our brain (q) is said by Einstein (Z)</i>	$q \text{ is } T, \text{ because } q \text{ is } Z$	VF	2 pre VF

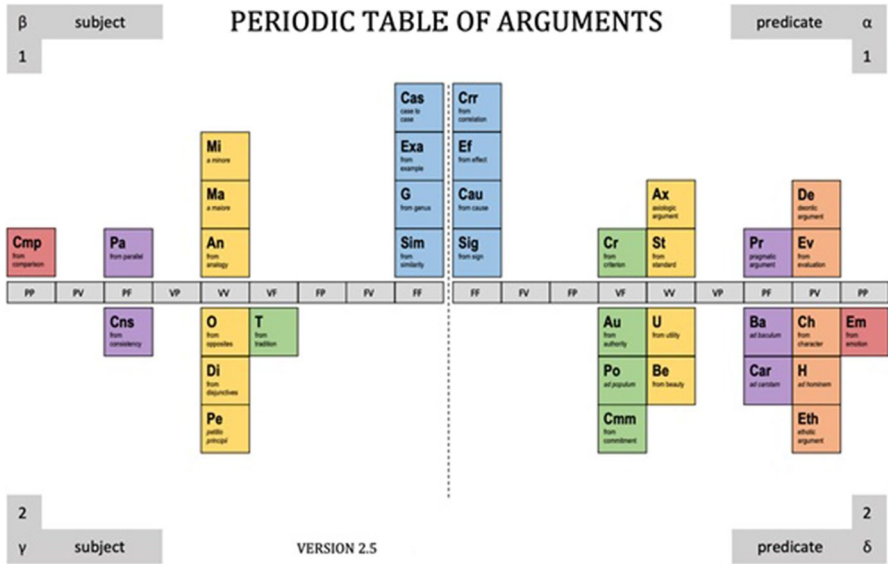


Fig. 3 Visual representation of the PTA – Version 2.5 (Wagemans 2020b)

In the visual representation of the PTA pictured in Fig. 3, the argument types that substantiate the four basic argument forms are situated in four different quadrants, which are indicated with the Greek letters alpha, beta, gamma, and delta respectively. Within each quadrant, arguments are further differentiated depending on the specific combination of types of statements. As a result, arguments sharing the same form are to be found in the same quadrant, while arguments sharing the same argument substance are to be found in the same column.

Given the respective possibilities of the three partial characterizations of argument, the theoretical framework of the PTA allows for $2 \times 2 \times 9 = 36$ systematic types of arguments. However, not all possible combinations in theory are found in practice. For instance, in the Alpha Quadrant there is no PP element, while in the Beta Quadrant there is no VF element. On the other hand, there can be more than one element corresponding to an argument type, depending on the linguistic formulation of the lever, i.e., the relationship between the non-common terms of the premise and the conclusion. Each element representing the above mentioned systematic types of argument may host a number of ‘isotopes’, which are named in accordance with the existing dialectical and rhetorical traditions of argument classification. We list in Table 2 the formulation of the lever and name of the isotope of the examples mentioned above.

Table 2 Levers and traditional names of examples following the four basic argument forms

Example	Lever	Isotope
Unauthorized downloading (a) is not theft (X) (V), because unauthorized downloading (a) does not deprive the original owner of the use of an object (Y) (F)	Not depriving the original owner of the use of an object (Y) is a criterion for not being theft (X)	Argument from criterion (Cr)
Cycling on the grass (a) is prohibited (X), because walking on the grass (b) is prohibited (X)	Cycling on the grass (a) is analogous to walking on the grass (b)	Argument from analogy (An)
He must have gone to the pub (q) is true (T), because the interview is cancelled (r) is true (T)	Either he must have gone to the pub (q) or the interview is not cancelled (not-r)	Argument from disjunction (Di)
We only use 10% of our brain (q) is true (T), because we only use 10% of our brain (q) is said by Einstein (Z)	Being said by Einstein (Z) is authoritative for being true (T)	Argument from authority (Au)

4 Specifying the Initial Analysis: The Argument Type Identification Procedure (ATIP)

We set out in this paper to explain in detail how to use the argument categorisation framework of the Periodic Table of Arguments (PTA) to specify the Reasoning aspect of the Comprehensive Assessment Procedure for Natural Argument (CAPNA). Since within the PTA framework, every type of argument comes with its own underlying reasoning, such assessment is preceded by an identification of the type of argument. For this reason, we present in this section a slightly shortened version of the Argument Type Identification Procedure (ATIP) (Wagemans 2020a), which provides the analyst with an identification of the type of argument in terms of the PTA as outlined above. In the following section, we then explain how to assess the underlying reasoning of the argument under scrutiny based on its identification.

The ATIP consists of the following six steps:

Step 1–Label the Textual Elements

To identify the type of argument, the analyst should first label its textual elements based on their pragmatic function. The following labels are in use:

- the text may contain a ‘connector’ such as *because* or *therefore* indicating the function of the statements as ‘conclusion’ and ‘premise’ (for lists of such indicators see, e.g., van Eemeren et al., 2007; Stab & Gurevych 2017)
- the statements usually contain a ‘subject’, i.e., an entity about which something is said, and a ‘predicate’, i.e., what is said about that entity
- the subject and predicate together form the ‘propositional content’ of the statement
- apart from this propositional content, the statement may contain a ‘doxastic commissive’ such as *we believe that, it is true that, or in my humble opinion*, which are linguistic expressions of the arguer’s commitment regarding the acceptability of the propositional content
- the statement may also contain a ‘doxastic directive’ such as *you should accept*, which is a linguistic expression of the arguer’s goal of convincing the addressee of the acceptability of the propositional content of the conclusion.

Step 2–Reformulate the Argument

The labelling of the elements of the argument enables the analyst to reformulate it in the standard form “[subject (conclusion)] [predicate (conclusion)], because [subject (premise)] [predicate (premise)]”. Such reformulation may involve several transformations of the original text:

- regarding the statements

- reordering of the statements to reflect the standard form “conclusion, because premise”
- regarding the connector
 - addition of the standard connector *because* between the conclusion and the premise
 - substitution of the original connector by the standard connector *because*
- regarding the non-propositional elements of the statements
 - hiding of the doxastic commissives and directives
- regarding the propositional content of the statements
 - anaphora resolution, i.e., the substitution of specific elements so that identical entities are referred to by identical words (preferably the most informative ones)
 - passivization or activation, i.e., changing the statement from active to passive voice or the other way around

Step 3–Determine the Argument Form

For completing this step in the procedure, the analyst can use the decision tree pictured in Fig. 4, which contains three heuristic questions as well as the corresponding instructions and outcomes depending on the answers to these questions.

Step 4–Determine the Argument Substance

The labelling of the type of statement is done in accordance with a widely used tripartite typology of statements developed within debate theory that consists of statements of fact (F), statements of value (V), and statements of policy (P).

- a statement of fact (F) is defined as a description of a particular state of affairs that is or can be empirically observed in reality or that is or can be imagined to exist. In order for the analyst to distinguish them from statements of value, it may be helpful to consider the following subtypes and examples:
 - empirical statements, such as ‘The suspect left a long trace of rubber on the road’.
 - existential statements, such as ‘God exists’
- a statement of value (V) is defined as an evaluative judgment about a particular entity based on a subjective selection and weighing of assessment criteria. In order for the analyst to distinguish them from statements of fact, it may be helpful to consider the following subtypes and examples:

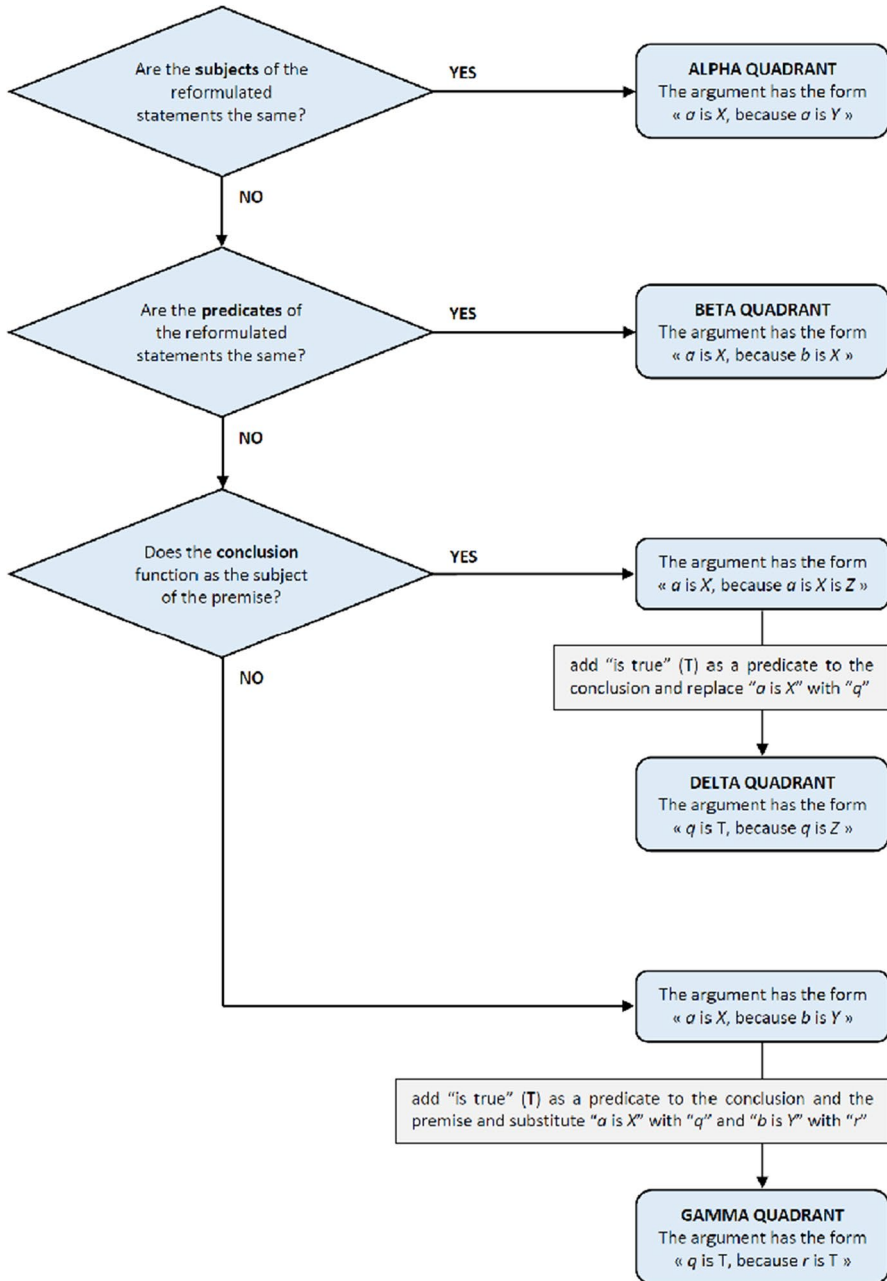


Fig. 4 Decision tree for determining the argument form

Table 3 Fulcrum and abstract lever of the four basic argument forms

Argument form	Fulcrum	Abstract lever
a is X , because a is Y	a	$Y R X$
a is X , because b is X	X	$a R b$
q is T , because r is T	T	$q R r$
q is T , because q is Z	q	$Z R T$

- aesthetic judgments, such as ‘*The Corrections* is a great novel’
- moral or ethical judgments, such as ‘Circumcision is reprehensible’
- legal judgments, such as ‘Unauthorized copying is not theft’
- pragmatic judgments, such as ‘Our plan for reducing CO₂-emission is feasible’
- logical judgments, such as ‘This proposition is true’
- hedonistic judgments, such as ‘Paragliding is fun’
- a statement of policy (P), which is defined as a directive statement that expresses advice, an incitement, or an imperative. The analyst may recognize statements of policy because of the presence of the term ‘should’ in combination with a verb expressing a particular action. Examples are:
 - advice, such as ‘Children should not sleep with artificial lightning’
 - incitements, such as ‘You should go to the gym’
 - imperatives, such as ‘Go to your room’

By labelling both the conclusion and the premise of the argument in this way, the argument substance can be determined as one of the nine possible combinations of types of statements (FF, VF, PF, FV, VV, PV, FP, VP, PP).

Step 5—Provide the Systematic Name of the Argument

The systematic name of an argument is a symbolic representation of the results of Step 3 and 4 of this procedure, and thus contains information regarding the argument form and the argument substance. It consists of:

- the prefix “1” or “2”, indicating a first-order or a second-order argument
- the infix “pre” or “sub”, indicating a predicate or subject argument
- the suffix “FF”, “VF”, etc., indicating the types of statements instantiated by the argument

Step 6—Provide the Traditional Name of the Argument

The traditional name of an argument is the name as it would occur in the dialectical lists of argument schemes and fallacies and the rhetorical lists of means of persuasion. Within the theoretical framework of the PTA, the traditional name of an

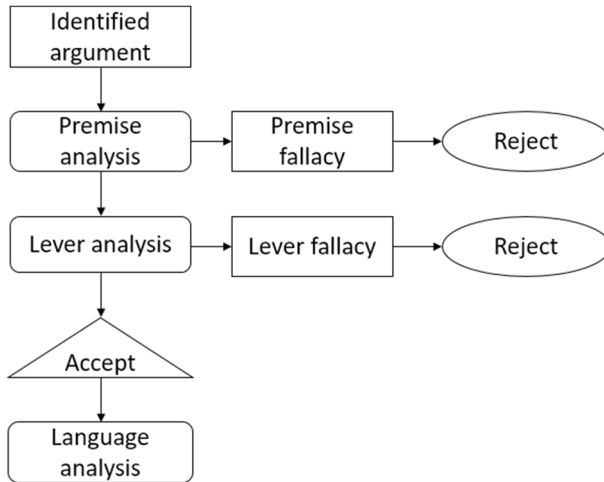


Fig. 5 The Reasoning Assessment Scheme for Natural Argumentation (Hinton 2021: p. 177)

argument is related to its ‘lever’, the relationship between the non-common terms of its conclusion and premise. Each of the four basic argument forms has a different common element (fulcrum) as well as a different set of non-common terms, which determines the abstract lever (see Table 3).

The concrete lever can be formulated by (1) finding the abstract lever related to the argument form; (2) substituting the actual predicates or subjects in the abstract lever, and (3) finding a fitting keyword for expressing the relationship. For the latter step, the current version of the PTA can be used as a heuristic instrument.⁴

5 Specifying the Reasoning Analysis: Procedural Questions

Once the ATIP has been carried out, the argument can be tested with questions of two types: those which address the acceptability of the propositional content of the premise and those which consider the force of its underlying reasoning, which is linguistically expressed by the lever (warrant, bridging premise, etc.). This method reflects two long-standing traditions in informal logic and is set out graphically in Fig. 5.

Firstly, the questions we propose are related to Johnson and Blair’s (2006) ‘RAS’ criteria for argument evaluation: relevance, acceptability, and sufficiency of an argument and Govier’s (2010) ‘ARG’ conditions: acceptability, relevance, good grounds. These criteria, however, are developed and operationalised into firm procedural steps. Our first question is about the ‘acceptability’ of the argument (i.e., in our terminology, the acceptability or truth of the premise). Our second question captures

⁴ For updates of the PTA and ATIP, see www.periodic-table-of-arguments.org.

Table 4 Example arguments and procedural questions

Example	Procedural questions
<i>Unauthorized downloading (a) is not theft (X) (V), because unauthorized downloading (a) does not deprive the original owner of the use of an object (Y) (F)</i>	(1) Is it true that <i>unauthorized downloading does not deprive the original owner of the use of an object</i> ? (2) Is <i>not depriving the original owner of the use of an object (Y)</i> a relevant and sufficient criterion for <i>not being theft (X)</i> ?
<i>Cycling on the grass (a) is prohibited (X), because walking on the grass (b) is prohibited (X)</i>	(1) Is it true that <i>walking on the grass (b) is prohibited (X)</i> ? (2) Is <i>cycling on the grass (a)</i> analogous to <i>walking on the grass (b)</i> ?
<i>He must have gone to the pub (q) is true (T), because the interview is cancelled (r) is true (T)</i>	(1) Is it true that <i>the interview is cancelled</i> ? (2) Is <i>he must have gone to the pub (q)</i> disjunctive to the interview is cancelled (<i>not-r</i>)?
<i>We only use 10% of our brain (q) is true (T), because we only use 10% of our brain (q) is said by Einstein (Z)</i>	(1) Is it true that <i>we only use 10% of our brain (q) is said by Einstein (Z)</i> ? (2) Is <i>being said by Einstein (Z)</i> authoritative for <i>being true (T)</i> ?

the relevance and sufficiency of the warrant (i.e., in our terminology, the solidity of the lever). Rather than these criteria, the procedural questions are named after their object of assessment. The first question is called the “Premise analysis” and the second one the “Lever analysis”. Our approach also reflects the dialectical tradition of ‘critical questions’ employed alongside argument schemes, which usually pertain to three aspects of the argumentation: the premise content, the warrant, and the context in which the argument has been put forward (see de Jong 2019: pp. 7–13). Of these three categories, the last one is captured in the Process aspect of the CAPNA (see Sect. 2). There are, however, significant differences between the widely used critical questions and our procedural questions. The most important of these is standardisation: there is no regular pattern to either the form or the content of the questions accompanying the many schemes listed in the work *Argumentation Schemes* (Walton et al. 2008). Also, the procedural questions are designed to be part of a wider process of evaluation: while those relating to the reasoning may resemble critical questions attached to an argument scheme, other questions at other parts of the process do not.

The premise analysis consists of the question of the truth or acceptability of the content of the premise – that it cannot be shown to be false; and its precise nature is dependent on the nature of the statement. The lever analysis pertains to the relevance and sufficiency of the argument—the strength of its leverage, which is a subjective judgement of the evaluator.

The ordering of the stages of analysis could be reversed, especially in cases where the lever strikes the evaluator as being obviously weak, however, the factual level is placed first by default as it can deliver an objective rejection of the argument if the premise is found to be false, admitting of no further discussion, while the rejection of a lever is less final and more open to debate.

Table 4 provides some examples of procedural questions and illustrates how their basic form is adapted to meet the requirements of specific argument types. Naturally, these questions may need to be followed by a series of sub-questions before an answer is reached: the procedure does not necessarily make the evaluation of arguments quick and easy, it seeks to make it systematic and transparent. These sub-questions may be inspired by the traditional critical questions of which various specifications have been provided (see, e.g., Walton et al. 2008).

6 An Example

In this section, we provide an example analysis of a passage of natural language argumentation with a step-by-step walk through of the identification of the argument type and evaluation of its underlying reasoning. The following extract is from a recent opinion piece in The Guardian online newspaper.

As long as the UK lacks a statutory law with a clear and binding code of practice, it simply isn't ready for the mass deployment of this technology. At the very least, we need to have a genuine public debate. As hard as it may be, democratic governments need to resist the temptation to undermine civil liberties in the name of safety and security. The stakes are far too high. (Kaltheuner 2020)

This passage, which is from an article criticising the decision to employ live facial recognition (LFR) software on the streets of London, appears to contain two separate arguments: one in the first sentence, and one in the third. The second sentence is an assertion for which no argument is given.

One way in which the initial analysis detecting the presence of arguments can be performed is through applying the ATIP and seeing what it can identify. Let us do that to the first sentence.

Step one—label the conclusion and the premise: the construction 'as long as' something, something else, is a clear sign that the first clause is the premise and the second the conclusion.

Step two—reformulate the argument in its standard form: changing to a 'because' structure, and replacing the pronoun with the implied noun phrase, we get:

The UK isn't ready for the mass deployment of this (LFR) technology, because the UK lacks a statutory law with a clear and binding code of practice (for the use of LFR technology).

Step three—identify the argument form: as the subjects are the same (the UK) and the predicates different (isn't ready for... and lacks a statutory law...), we have the form a is X , because a is Y – this is a first-order predicate argument from the Alpha Quadrant.

Step four—determine the argument substance: the premise, that the UK has no law of a certain type, is clearly a factual claim; while the conclusion, that the UK

is not ready to employ the new technology, is an evaluative judgement. The argument substance is, therefore, VF.

Step five—provide the systematic name of the argument: from the foregoing we have a first-order predicate argument with substance VF, so the systematic name is: 1 pre VF.

Step six—provide the traditional name of the argument: the table shows only one isotope for 1 pre VF, the ‘argument from criterion’ (Cr).

So, we are satisfied that our text contains an argument and we have identified it. Assuming that the argument passes through the Process stage of analysis, we now come to evaluate its reasoning, through the means of relevant procedural questions. Firstly, premise analysis:

Is it true that the UK lacks a statutory law with a clear and binding code of practice for the use of LFR?

The article links to the website of the Information Commissioner’s Office (ICO), which is described as ‘The UK’s independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals’ (ICO 2020). This site contains the sentence: ‘We reiterate our call for Government to introduce a statutory and binding code of practice for LFR as a matter of priority’, in a statement posted three days before the argument text, from which we can reasonably conclude that none currently existed, and the premise of the argument is true.

Secondly, Lever Analysis:

Is not having a statutory law with a clear and binding code of practice for the use of LFR a relevant and sufficient criterion for not being ready to employ LFR?

This is a far more complex question. The procedure itself cannot answer this question, but that is not its role. The procedure should ensure that whoever comes to evaluate the reasoning of the argument will arrive at the same question, which then becomes the key ground for the discussion to move into. The evaluator may choose to pause at this point in order to resolve the question, or may decide that as the answer is not obviously ‘no’ and no clear lever fallacy has been committed, the reasoning may be accepted presumptively, and further assessment of the Expression component of the argumentation carried out.

The second argument in the passage can be quickly assessed following the same steps. Step one—the premise is ‘The stakes are far too high’, and the conclusion ‘democratic governments need to resist the temptation to undermine civil liberties in the name of safety and security’.

Step two—a greater amount of reformulation is required this time in order to bring out the common term, and to make an idiomatic phrase clearer:

Undermining civil liberties in the name of safety and security should not be done by democratic governments, because undermining civil liberties in the name of safety and security bears too much risk.

Step three—in the premise, the subject is ‘undermining civil liberties in the name of safety and security’, and the predicate ‘bears too much risk’. In the conclusion, the subject is also ‘undermining civil liberties in the name of safety and security’, and the predicate ‘should not be done by democratic governments’. The subjects

are, therefore, the same, and the argument is a first-order predicate argument, a is X because a is Y .

Step four—the premise has the form of a statement of fact, and the conclusion is one of policy. The argument substance is PF.

Step five—the systematic name is 1 pre PF.

Step six—the isotope name is ‘pragmatic argument’ (Pra).

With this argument, the reasoning analysis highlights a different type of concern. We begin with premise analysis:

Is it true that undermining civil liberties in the name of safety and security bears too much risk?

This is an extremely difficult question to answer, since ‘how much is too much?’ is a subjective matter. Let us move on for a moment to the lever analysis:

Is bearing too much risk a relevant and sufficient pragmatic reason for a democratic government to not perform an action?

Here, there seems little doubt that the answer is yes; indeed, the phrase ‘too much’ makes that clear. The real question with this argument, then, is do the risks to civil liberties outweigh the benefits to safety and security? This will seem obvious, but, again, the role of the procedure is to provide a clear path to the crux of the argument, a path anyone can follow, and to highlight the next step in evaluating the reasoning, which will be to investigate the risks and the benefits of the policy carefully—not simply to accept the arguer’s word for their existence.

7 Conclusion

The ultimate aims of this work are ambitious: a procedure that allows an assessor to evaluate the underlying reasoning of arguments expressed in natural language would be a powerful tool and would need to be a tool of great flexibility while maintaining sufficient simplicity to make it practicable. In this paper, we have set out how such a tool can be created, and we have placed it within a broader framework of argument assessment.

Based upon a definition of argumentation as consisting of Process, Reasoning, and Expression aspects, the procedure we describe isolates Reasoning, as far as that is possible, from the other elements of argumentation. The procedure is employed as part of the CAPNA, which is a still-developing comprehensive and systematic tool for argument assessment that aims at capturing all three aspects. As with other elements of the CAPNA, our procedure progresses through the use of procedural questions, which are standardised and founded in widely-accepted principles of argumentation.

There are, we believe, a number of advantages to this system, which, although still relying on the subjective judgements of analysts in some places, is repeatable and transparent. Firstly, thanks to the ATIP, the identification of argument types is made systematic and, therefore, more justifiable than in ad hoc approaches. Secondly, the identification does not merely name the argument scheme, but picks out the key elements within it, giving the analyst a clear picture of how it works. Thirdly, this clearer picture greatly assists anyone assessing the argument in knowing exactly

what they should be evaluating, and which questions they should be asking to do so. Finally, while the tool as it currently stands is intended for use by human evaluators, the systematic procedure has the nature of a pseudo-algorithm which could be exploited in the development of computational applications. For comparison, it is worth noting that, in the case of reasoning, the criteria suggested by other argumentation theorists, such as Govier's 'ARG conditions', where premises must be acceptable and relevant, and provide good grounds for the conclusion (2010), and Johnson & Blair's 'RSA criteria'—relevance, sufficiency, and acceptability (2006), are certainly fair and well-supported, and they inform the structure of the CAPNA, but they are not in themselves procedures, and it is for the analyst to decide how those criteria might best be applied. Establishing such conditions is a first step, but we hope to have gone further by introducing systematicity and clarity into the way in which they are utilized. Our approach is different because it introduces procedural thinking into the analysis and evaluation of argumentation.

We acknowledge that the current model of the CAPNA has a number of limitations. As stated above, it can be used only in the evaluation of what might be termed 'linear' argumentation, and does not take into account counter arguments, or more complex multi-premise argument forms. It should be noted that the procedure does not seek to 'map' all the arguments and refutations involved in a debate, simply to assist in evaluating each individual premise/conclusion structure. Future developments may include such functions. We accept also that the reformulation of the actual argument into a suitable 'conclusion because premise' structure may cause difficulty, but would point out that the problem of paraphrasing original texts to extract coherent arguments is a wider one within argumentation theory.

We are aware too that the procedure as presented in this paper may be considered difficult to employ, especially for students, which might lead to accusations that it is impractical, and irrelevant in teaching. We believe, on the contrary, that there are a number of insights from this method of analysing arguments which can be highlighted as part of courses on argumentation, without necessarily replacing more traditional approaches. Indeed, both authors (Hinton forthcoming, King and Wagemans forthcoming) are currently preparing textbooks containing more didactically focussed versions of the CAPNA and ATIP respectively, which employ less technical formulations and can therefore more easily be applied in educational settings. It is worth noting, however, that there is no 'easy' way to evaluate reasoning systematically, and it is hoped that elements of the process will soon be automatised, reducing the workload of the evaluator.

One other element which has come out of this research is a reconceptualization of fallacy in argument. Arguments are not rejected because they appear to be similar to a known fallacious form, they are rejected because they fail to successfully negotiate one of the procedural questions. If that failure occurs during the Reasoning evaluation, then a fallacy of reasoning is present; if it occurs earlier, a fallacy of process has been identified. In this way, we avoid the necessity of giving a definition of fallacy which covers all the varied types of problems which are normally given that name, and we avoid the need to name fallacies and then show how individual instances are connected to paradigmatic examples. A complete explication of this theoretical shift will require a paper of its own, but we believe that there are many advantages

to moving away from the traditional model of fallacy lists, divided into categories based on unclear principles, and with no stated process for their identification.

Of course, the traditional fallacies did not appear from nowhere and are likely to correspond with positions in the procedure where negative evaluations are frequent. Indeed, the accumulated wisdom on argument error which they represent plays no small part in the construction of the procedure itself. However, within this system, no argument should, then, be cited as an example of a fallacy: if an argument turns out to be fallacious in some aspect, it should be cited as an example of an argument which fails at a particular moment in the evaluation. That it may also fit a traditional fallacy name is to be expected in many cases, but is incidental to the procedure. Outside the evaluation, ‘fallacy’ is a vague and uncertain term of limited theoretical basis or application. To conceive the central notion of ‘fallacy’ in this way implies a fundamental change in our thinking about fallacies and one to which it may not be easy to get accustomed.

Regarding the two theoretical frameworks used in this paper, we believe that our effort to combine them makes it clear how the PTA enables greater specification of the previously undeveloped Reasoning level analysis of CAPNA, and how synchronisation into that procedure allows the descriptive PTA tool to be operationalised in an evaluative, normative manner. Further work is required to elaborate the procedural questions of our tool, providing clear steps for the evaluation of all types of argument lever. This work goes hand-in-hand with the continuing evolution of the PTA, and the discovery and examination of more isotopes and the ways in which these are expressed linguistically. Finally, the role of the underlying Reasoning analysis within the broader CAPNA framework, and its precise use in practice, can only be fully described once the other constituents of that assessment framework, i.e., Process and Expression have been properly constructed. While much work has been done on the system of informal argument semantics for the evaluation of linguistic expression of arguments, the creation of procedural questions for the evaluation of argument process has only just begun.

Funding This paper has been supported by an STSM grant from COST Action CA 17132 – European Network for Argumentation and Public Policy Analysis.

Availability of data and materials Not applicable.

Code availability of data and materials Not applicable.

Declarations

Conflicts of interest The authors declare they have no such interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Broda-Bahm, K., D. Kempf, and W. Driscoll. 2004. *Argument and audience: Presenting debates in public settings*. Amsterdam: International Debate Education Association.
- Freeley, A.J., and D.L. Steinberg. 2014. *Argumentation and debate: Critical thinking for reasoned decision making*. Thirteenth. Boston, MA: Wadsworth.
- Gobbo, F., M. Benini, and J.H.M. Wagemans. 2019. Annotation with Adpositional Argumentation: Guidelines for building a Gold Standard Corpus of argumentative discourse. *Intelligenza Artificiale* 13 (2): 155–172.
- Govier, T. 2010. *A practical study of argument*, 7th ed. Belmont, CA: Wadsworth.
- Hinton, M. 2021. *Evaluating the Language of Argument*. Cham: Springer.
- Hinton, M. forthcoming. *Ways to Argumentation*. Lodz: University of Lodz Press.
- Information Commissioner's Office. 2020. *ICO statement in response to an announcement made by the Metropolitan Police Service on the use of live facial recognition*. [updated 2020 Jan 24; cited 2020 Oct 20]. Available from: <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/01/ico-statement-in-response-to-an-announcement-made-by-the-met-police/>.
- Johnson, R.H., and J.A. Blair. 2006. *Logical self-defense*. New York: Idebate Press.
- Jong, A. de. 2019. *Analyzing and systematizing Walton's critical questions*. MA Thesis University of Amsterdam.
- Kaltheuner F. 2020. Facial recognition cameras will put us all in an identity parade. *The Guardian*. [updated 2020 Jan 27; cited 2020 Oct 20]. Available from: <https://www.theguardian.com/commentsfree/2020/jan/27/facial-recognition-cameras-technology-police>.
- King, C.G., and Wagemans, J.H.M. forthcoming. *Argumentation in the Wild*. Cambridge, MA: MIT Press.
- Perelman, C., and L. Olbrechts-Tyteca. 1969. *The New Rhetoric*. Notre Dame: University of Notre Dame Press.
- Skorupski, J. 2010. *The domain of reasons*. Oxford: Oxford University Press.
- Stab, C., and I. Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics* 43 (3): 619–659. https://doi.org/10.1162/COL1_a_00295.
- van Eemeren, F.H., and R. Grootendorst. 2004. *A systematic theory of argumentation*. Cambridge: Cambridge University Press.
- van Eemeren, F.H., P. Houtlosser, and A.F. Snoeck Henkemans. 2007. *Argumentative indicators in discourse*. Dordrecht: Springer.
- Visser, J., Lawrence, J., Wagemans, J.H.M., and Reed, C.A. 2018. Revisiting computational models of argument schemes: Classification, annotation, comparison. In S. Modgil, K. Budzynska & J. Lawrence (Eds.), *Computational models of argument: Proceedings of COMMA 2018. Frontiers in Artificial Intelligence and Applications. Volume 305* (pp. 313–324). Amsterdam: IOS Press. <https://doi.org/10.3233/978-1-61499-906-5-313>.
- Visser, J., Lawrence, J., Reed, C.A., Wagemans, J.H.M., and Walton, D.N. 2021. Annotating argument schemes. Published online May 7, 2020. *Argumentation*, 35, 101–139.
- Wagemans, J.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, J.H.M. 2019. Four basic argument forms. *Research in Language* 17 (1): 57–69. <https://doi.org/10.2478/rela-2019-0005>.
- Wagemans, J.H.M. 2020a. *Argument Type Identification Procedure (ATIP) - Version 3*. [updated 2020 Feb 21; cited 2020 Oct 20]. Available from: www.periodic-table-of-arguments.org/argument-type-identification-procedure.
- Wagemans, J.H.M. 2020b. *PDF Periodic Table of Arguments 2.5*. [cited 2020 Oct 20]. Available from: https://www.academia.edu/36621185/PDF_Periodic_Table_of_Arguments_2_5.

- Wagemans, J.H.M. 2020c. Why missing premises can be missed: Evaluating arguments by determining their lever. In J. Cook (Ed.), *Proceedings of OSSA 12: Evidence, Persuasion & Diversity*. Windsor, ON: OSSA Conference Archive. URL = <https://scholar.uwindsor.ca/ossaarchive/OSSA12/Saturday/1>
- Walton, D.N., C. Reed, and F. Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.