Five approaches to argument strength: probabilistic, dialectical, structural, empirical, and computational

FRANK ZENKER[1,2,3,*]
[1] *Department of Philosophy, Lund University, Sweden*
[2] *Int. Center for Formal Ontology, Warsaw U. of Technology, Poland*
[3] *Dpt. of Philosophy, Boğaziçi University, 34342 Istanbul, Turkey*
[*] *corresponding author: frank.zenker@fil.lu.se;*
frank.zenker@pw.edu.pl; ORCID: 0000-0001-7173-7964

KAMILA DĘBOWSKA-KOZŁOWSKA
Faculty of English, Department of Pragmatics of English
Adam Mickiewicz University in Poznań, Poland
kamila@wa.amu.edu.pl; ORCID: 0000-0002-5039-6133

DAVID GODDEN
Philosophy Department,
Michigan State University, USA
dgodden@msu.edu; ORCID: 0000-0002-4870-6331

MARCIN SELINGER
Department of Logic and Methodology of Sciences,
University of Wrocław, Poland
marcisel@uni.wroc.pl, ORCID: 0000-0002-9685-6684

SIMON WELLS
School of Computing
Edinburgh Napier University, Scotland
s.wells@napier.ac.uk; ORCID: 0000-0003-4512-7868

Originating from the 2019 Symposium on Argument held at the 3rd European Conference on Argumentation (ECA 2019) in Groningen, The Netherlands, this paper outlines and briefly compares central tenets in five approaches to argument strength: the probabilistic (Bayesian), the dialectical, the structural, the empirical, and the computational approach.

KEYWORDS: acceptability, argument, attack, Bayesian, computation, defense, empirical, experiment, defeat, diagram, dialectical, empirical, force, map, model, measurement, probability, relevance, semantic, strength, structure, sufficiency, obligation, permission, probability, retract

1. INTRODUCTION

The symposium's purpose was to review approaches to argument strength as a central notion for a theory of natural language argumentation. This paper presents five such approaches, namely: the probabilistic (Bayesian) approach, the dialectical, the structural, the empirical, and the computational approach.

Already in the planning phase we bracketed the rhetorical approach. It would presumably have suggested discourse items that the empirical approach might test statistically, or have provided post-hoc explanations why persuasion did (not) occur. A more complete overview of theoretical approaches to argument strength would no doubt include this approach.

Prior to the symposium, authors independently developed their own analyses of the three main arguments in *Epicureans on Squandering Life* (Aikin & Talisse, 2019). The motivating idea was to reveal differences between the five approaches (if any) as applied to the *same* text. The symposium not only included initial empirical findings regarding the perceived strength of these three arguments, but also addressed their structure. Visualizations thereof are part of the slide-set available at http://tiny.cc/ArgStrth. Here we present the approaches' theoretical gist (Sects. 2 to 6), the section-order mirroring the speaker-order at the symposium. Our discussion suggests that all five approaches can relate fruitfully (Sect. 7). Instead of conclusions, we give a brief summary (Sect. 8).

2. THE BAYESIAN APPROACH[1]

If one interprets probability values subjectively, they represent not objective chances of singular or repeatable events, but *degrees of belief* (credences) or *graded commitments* in reasons and claims (Korb, 2004; Oaksford & Hahn, 2004). To a given natural language argument featuring at least one reason and a claim, the Bayesian machinery applies readily given these abbreviations:

C = claim, conclusion, or standpoint
R = reason, or the set of conjoined premises {R_1 & R_2 & ... & R_n}
P = probability (a measure of credence, subjective belief, or committment)
P(C/R) = probability of a claim given a reason
P(R/C) = probability of a reason given a claim
t = an arbitrary threshold value
~ = negation

[1] Section adapted from Godden & Zenker (2016), which explicates formal steps here omitted and provides additional references (see Zenker, 2013).

Bayes' Theorem (BT) (Bayes, 1763) defines the posterior probability of a claim *C* given a reason *R* as the probability of the claim *and* the reason, divided by the probability of the reason irrespective of the claim:

(BT) $P(\mathrm{C} \mid \mathrm{R})=[P(\mathrm{R}|\mathrm{C})\times P(\mathrm{C})]/(P(\mathrm{R})$

The factor by which the prior probability of claim C, $P(\mathrm{C})$, is multiplied to yield the posterior probability of C conditional on the reason R, $P(\mathrm{R}|\mathrm{C})$, we may call the *impact* of the reason, $i=P(\mathrm{R}|\mathrm{C})/P(\mathrm{R})$. In full form:

(Impact term) $i=P(\mathrm{R}|\mathrm{C})/[(P(\mathrm{R} \mid \mathrm{C})\times P(\mathrm{C})+P(\mathrm{R}|{\sim}\mathrm{C})\times P({\sim}\mathrm{C})]$

The impact term thus *is* the ratio of how probable the reason is given the claim to how probable the reason is regardless (irrespective of the claim). Under the constraint that $P({\sim}\mathrm{C})=1-P(\mathrm{C})$, the impact term thus expresses a conditional expectation *of the reason* if the claim holds, as against a prior expectation on the reason regardless. The priors, $P(\mathrm{C})$ and $P({\sim}\mathrm{C})$, express commitments to (or degrees of belief in) the truth or falsity of C given background information.

The other two terms, $P(\mathrm{R}|\mathrm{C})$ and $P(\mathrm{R}|{\sim}\mathrm{C})$, express *likelihoods,* i.e., prior judgements about (or attitudes towards) the probative value of reasons under the constraints $P(\mathrm{R}|\mathrm{C})=1-P({\sim}\mathrm{R}|\mathrm{C})$ and $P(\mathrm{R}|{\sim}\mathrm{C})=1-P({\sim}\mathrm{R}|{\sim}\mathrm{C})$. The first term, $P(\mathrm{R}|\mathrm{C})$, expresses the *sensitivity* of the reason to the claim. When evaluating the reliability of an empirical test, for instance–and given that 'hypothesis (H)' replaces 'claim', and 'evidence (E)' replaces 'reason'–$P(\mathrm{R}|\mathrm{C})$ reports the *true positive rate* (i.e., the ratio of *correct* positive test-results to all test-results). The second term, $P(\mathrm{R}|{\sim}\mathrm{C})$, expresses the complement of the *specificity* of the reason to the claim: the *false positive rate* (the ratio of *incorrect* positive test-results to all test-results). Both terms together express how well *R* "correlates" with C, except that this correlation *need not* ground in the frequency of event tokens. Thus, if C logically entails R, then $P(\mathrm{R}|\mathrm{C})=1$; and if no "correlation" obtains (statistical independence), then $P(\mathrm{R}|\mathrm{C})=P(\mathrm{R}|{\sim}\mathrm{C})=P(\mathrm{R})$.

Sensitivity and specificity are readily meaningful for long-run frequencies of event tokens. In contexts of natural language argumentation, however, reasons can provide support for claims *irrespective* of frequency considerations. A suitable non-frequentist interpretation that leaves both likelihood terms meaningful is: reason *R* is *sensitive* to claim *C* to the extent that R supports C more than R supports *any* other claim C* (itself entailing ~C), i.e., $P(\mathrm{C}|\mathrm{R})>0.5>P({\sim}\mathrm{C}|\mathrm{R})$. And: R is *specific* to C to the extent that R rather than *any* other reason R* (itself entailing ~R) supports C, i.e., $P(\mathrm{C}|{\sim}\mathrm{R})<0.5<P({\sim}\mathrm{C}|{\sim}\mathrm{R})$. Drawing sensitivity and specificity together, the extent to which R supports C thus

depends on the extent to which the *C*-supporting-reason R *fails* to support ~C, and that to which argumentative support for *C cannot* be generated by reasons besides R.

In the two extremal cases, P(C|R)=1 and P(C|R)=0, support thus is strongest if R is an *exclusive and decisive* supporting reason-for-C; and support is weakest if R is a *common and indecisive* supporting reason-for-C. An example of an exclusive and decisive supporting reason (outside the argumentative domain) is a modern litmus test in the form of a universal pH-indicator (hydrogen ion), where the indicator paper's coloring red is a causal effect of a solution's hydrogen ion concentration. Assume, unrealistically, that the test is *perfectly* sensitive, i.e., P(R|C)=1. This means any pH-level<3 always turns the indicator paper red. Equally unrealistically, assume the test is *perfectly* specific, i.e., P(C|~R)=0, so that any pH-level≥3 always fails to color the paper red. And again unrealistically, assume that for this purpose *no* other test is available. The paper's *not* turning red now *decisively* indicates that the solution is basic (not strongly acidic), so that P(~R|~C)=1-P(R|~C)=1—and it does so *exclusively* too, for by assumption no other test can. In this case, i=1/P(C); thus BT (equation 12) reports P(C/R)=1. (For application to arguments denying the antecedent or affirming the consequent, see Godden & Zenker, 2015).

Given non-extremal prior probability values for R and C, selecting suitable values for the sensitivity and selectivity terms provides one way of modeling the comparative support that R provides to C, e.g., as P(C|R)−P(C). (Other measures are possible (see, e.g., Pfeifer, 2013). What these suitable values *are*, this depends on what arguers believe or commit to, or on what they recognize as evidence. In fact, advice on which values to select is not an integral part of the Bayesian approach, but rather pertains to considerations originating as much in philosophy of science as in practical philosophy. Presupposed in an analysis, moreover, is access to the structure of reasons and claims themselves, via an argument map, for instance, or a Bayesian network.

What particularly recommends the Bayesian approach to modeling argument strength is that the threshold t can specify not only arguers own commitments to the 'is a (necessary, insufficient, sufficient, supererogatory) reason for''-relation with sufficient numeric precision (Spohn, 2012). The threshold can also serve to spell our contextual constraints on this relation. Indeed, the approach specifies the informal logic criteria of relevance, sufficiency, and acceptability ("RSA conditions"), and thus connects them with quantitative notions of *evidential strength* in applied statistics, for instance, or *evidential value* in (criminal) law (Godden & Zenker, 2016).

The Bayesian approach is a *Pascalian* approach to probability. Among others, it requires that changing one's credence in a proposition P (in response to reasons) entails a corresponding and well-defined change of credence in its negation, ~P. This constraint runs counter to an important intuitions that the so-called *Baconian* approach to probability seeks to secure (Cohen, 1980; Spohn, 2012; Zenker, 2015, Sect. 5.3).

3. THE DIALECTICAL APPROACH

3.1 Overview

There are two classes of dialectical theories of argument: *informal* (van Eemeren & Grootendorst 2004; Walton, Reed & Macagno 2008) and *formal* (Barth and Krabbe 1982; Hamblin 1970, 1971; Kieff 2011; Krabbe 2013, 2017; Krabbe & Walton 2011; Rescher 1977; Walton & Krabbe 1995). Dialectical approaches to modelling argumentation represent the turn-by-turn sequence of conversational moves, or speech acts, in an argumentative exchange, thereby facilitating their analysis and appraisal (e.g., van Eemeren & Grootendorst 2004, Walton & Krabbe 1995). The analytical task involves reconstructing argumentative text or discourse into a sequence of such moves. Once a set of protocols, or rules for the dialogue game, are provided, moves and move sequences can be appraised.

Dialectically, argumentative norms are modelled as procedural rules (protocols) that permit, oblige, or prohibit particular moves or move sequences. Thus, the force of argumentative norms is operationalized in terms of creating obligations, or granting permissions, for discussants to make moves of specified kinds at future stages in the dialogue. For example, discussants have the right to assert any standpoint. Having asserted a standpoint, a proponent undertakes the obligation to defend that standpoint if challenged (thus acquiring a "burden of proof"), while respondents gain the right to challenge asserted standpoints.

Argument strength, then, is modelled as the set of commitments, entitlements, and obligations pertaining to discussants at any stage in an argumentative dialogue. Collectively, these comprise the "move space" available to that discussant. In this context, argument strength can be operationalized as the (un)availability of participant moves that constrain further interlocutor moves. Minimally, argument strength thus is a function of the (un)availability of non-losing future participant moves. In this sense, the strongest proponent-argument leaves no further opponent-move except concession (i.e., retraction of either a standpoint or of critical doubt), and the weakest proponent argument constrains *no* opponent-move, given the "move-space."

3.2 Informal Dialectical Theories

Informal dialectical theories draw from a conceptual, analytical, and evaluative toolkit that includes argumentation schemes, critical questions, and fallacies. In presenting reasons, arguers draw upon a repertoire of schematic argument forms (i.e., argument schemes) to construct and compose their arguments. These schemes are rather like recipes, prompting arguers to provide all the right ingredients, properly assembled, for the argument they seek to produce.

One way to distinguish different discursive domains (or fields of argument) is according to the currency that a schematic argument has in each domain. Thus, the moves available to an arguer at any point in an argumentative dialogue are a function of both the claims they may draw upon as ingredients (i.e., premises) and the repertoire of schematic arguments (i.e., moves, or rules) that may be activated if those ingredients are provided as inputs.

By contrast, the fallacies represent a negative move space in the discursive domain. While fallacies schematically represent arguments, they are arguments of the unacceptable variety, which may be criticized merely on the grounds that they instantiate the fallacy. Typically, however, the fallaciousness of a given episode of argumentation depends not merely on its schematic form but also on background information and various contextual features specific to that episode.

To help determine whether a given instance of an argument scheme is deployed felicitously or fallaciously, one applies a prescribed set of critical questions accompanying each scheme. These questions are designed as tools to test the argument for stereotypical ways that arguments of that schematic type can default.

Dialogue moves are made by applying these tools on a turn-by-turn basis in ways such that the burden-of-proof shifts (back and forth) between discussants over the course of the dialogue according to *which* discussant has incurred *what* obligation to provide reasons, grant concessions, or retract standpoints. If at the end of the dialogue a proponent has failed to discharge their initial burden of proof, then they must retract their standpoint. For, their argument was not strong enough to support the standpoint. Alternatively, if the proponent meets their burden of proof, then the respondent must withdraw their critical doubts, and so the argument was strong enough, dialectically speaking.

In determining argumentative norms, these informal tools are of *heuristic* value. But whether singularly or in combination, no such tool yet delivers a *comprehensive* catalogue of (im-)permissible arguments, let alone a *complete* list of the cogency-criteria that would apply. None, for instance, provides an exhaustive list of the (im-)permissible discussant moves. This makes a more systematic approach desirable, as found in formal dialectical theories of argument.

3.3 Formal Dialectical Theories

Formal dialectical theories of argument depict arguments as *profiles of dialogue* that are distinguished by their different protocols (Krabbe 1999, 2002; Walton 1999, 54f.; 2015, 96f.; van Eemeren et al., 2014, 366-367). Structurally, dialogue profiles are directed graphs with a tree-structure. Nodes represent possible moves in the argumentative dialogue. Edges (joining the nodes) represent paths to permissible discussant-moves (that are *available* according to the dialogue game's rules) given a particular dialogue-state. Actual dialogue can thus instantiate some "branch" (i.e., a specific path from root to tip). An obligatory move is a single path emanating from the previous move; while several paths lead to permissible moves. Thus, as all edges are weighted equally, *path connection strength* is modeled as a constraint on the available response moves.

Each participant's goal is to strategically execute a move sequence that compels their interlocutor to make a game-ending move: either proponent standpoint retraction or respondent standpoint concession. A discussant has a *winning strategy* just in case they have available to them a sequence of moves such that, whatever response their interlocutor makes, each branch ends in a losing interlocutor move.

3.4 Argument Strength, Dialectically Conceived

Following such an analysis, what can be said about argument strength? Viewed dialectically, argument strength is a function of the (un)availability of permissible move sequences originating at the present dialogue stage, and ending in a discussant's role-specified goal being achieved.

Since the dialectical role of a respondent is to raise critical doubts, rather than to defend a standpoint from critical doubt, respondents technically don't offer arguments understood as presentations of reasons in support of standpoints, unless they advance counter-arguments (e.g., claims motivating their critical doubt, or alternative standpoints). As such, "argument strength" as it applies to each discussant might be better labelled "position strength," understood as the opportunity of the discussant to make a non-losing move. Positively, this amounts to the (un)availability of participant moves that lead to losing interlocutor moves. Minimally, it is a function of the (un)availability of non-losing future participant moves.

3.5 Further Considerations

Nevertheless, other evaluative considerations also might be brought to bear, as argument strength, generally understood, seems also to depend on them. Consider, for example:

(a) *Background Commitments*: Determinations of the availability of a discussant's non-conceding moves requires knowledge of their other commitments. It cannot be assumed that material that goes unchallenged is accepted. Rather, proponents strategically select as premises claims that they think are most invulnerable to challenge or likely to go unchallenged, while respondents strategically direct their critical attentions to those moves they deem most vulnerable to challenge or most likely to be indefensible.

(b) *Commitment Set Dynamics*: Dialectically, one "wins" an argument by obliging the opponent to retract either their standpoint or their critical doubt. Seemingly, arguments that are better positioned for such "wins" are stronger. But respondents can "win," as skeptics do, by persisting with their critical doubts simply by refusing to grant proponent claims that are otherwise unobjectionable. Similarly, proponents might "win" only by taking on so many otherwise implausible commitments that, were they not committed to their standpoint, they would rather give them up. Sometimes, to constrain an interlocutor's dialectical room to maneuver, discussants end up giving up, or denying, so much of the rest of their ordinary commitments that they come away from the argumentation bearing little cognitive resemblance to the discussant who entered into it. Especially if retraction is permitted, a more complete evaluation of argument strength would incorporate a measure of minimum mutilation (Quine, 1961, 44; 1992, 14-16) and considerations of (in)coherence when comparing the opening and closing commitment sets of each discussant.

(c) *Meta-argumentation*: The ordinary transaction of reasons arguably involves the meta-argumentative task of evaluating them, as well as the meta-dialogical critique of the applicable rules and standards (Finocchiaro, 2007; 2013; Krabbe, 2003). Because fallacy accusation, for instance, constitutes a meta-argumentative move, the "move space" available to discussants should include *meta-argumentative* moves. In fact, the critical point of the Squandering Argument (SA) (Aikin & Talisse, 2019) is meta-argumentative, pertaining not to what is said but *to being entitled to say it*. For the SA attributes a pragmatic inconsistency (performative absurdity) not to the speech act's content, but to the discussant's *performing* the act of asserting their view. A dialectical analysis and appraisal therefore requires that meta-argumentative moves are part of a dialectical system.

In sum, dialectical theories offer a straightforward account of argument strength, operationalized in terms of the availability of non-conceding moves to a discussant and the lack of availability of non-conceding moves to their interlocutor. Yet the application of this apparently clear and simple standard can be complicated by many factors, both situational and structural.

4. THE STRUCTURAL APPROACH

4.1 Overview

We present the structural approach as a framework for evaluating the strength of structured arguments and counterarguments based on the kinds of diagrams used in informal logic. Evaluation here is bottom up, general and abstract, making it easier to compare specific models of evaluation and to formulate a consistent methodology (Tokarz, 2006; Gordon & Walton, 2006; Prakken, 2010; Selinger, 2019). The generalized model, after all, abstracts not only from the particular set of values that represent argument strength, but also from particular algorithms that transform the acceptability of premises into the acceptability of conclusions.

4.2 The underlying idea of argument structure

According to the underlying model of argument structure we apply here, the focal objects consist of separate inferences constituting atomic (simple or linked) arguments. Complex arguments are formed from atomic ones via syntactic operations corresponding to convergent, divergent and serial arguments (standard part), as enriched with counter-considerations and undercutters (dialectical extensions). Counter-considerations (or 'con-arguments' *vs*. 'pro-arguments') attack sentences—underminers attack premises, rebuttals attack conclusions—while undercutters attack inferences, i.e. inferential links between sentences. Such structures consist of the sentences of some predefined language L. They can be represented symbolically as finite, non-empty sets of *sequents*, i.e. quadruples of the form $<P, c, d, R>$, where $P \subseteq L$ is a finite, non-empty set of premises; $c \in L$ is a conclusion; $d \in \{0, 1\}$ is a Boolean value (1 for pro- and 0 for con-sequents); $R \subseteq 2^L$ is a finite set whose elements are non-empty, finite sets of (linked) undercutters (Selinger, 2019).

4.3 Abstract evaluation

For the purpose of evaluation, two types of values are introduced. Those that are assigned to sentences we simply call *values*. By contrast, those assigned to inferential links between sentences we call *weights*. The set of values V can be any set containing at least two elements, which are assigned to the sentences of L by a partial function v. The elements of a distinguished (non-empty) proper subset $V^* \subseteq V$ are assigned to audience-accepted sentences. The set of weights W is any set containing at least two elements assigned to the strengths of direct inferences, regardless of the premises' actual values. Also in this set, we distinguish a proper (non-empty) subset $W^* \subseteq W$ whose elements correspond to valid inferences.

Both sets, V and W, are ordered by the 'being stronger'-relation. Each specific way of ordering thus determines the corresponding concept of argument strength (if unordered, V and W determine no such concept). A linear (total) order is prima facie a natural choice. However, the minimum assumption imposed on such an ordering is that any distinguished value/weight be stronger than any undistinguished one. If an order is linear, then each two computable arguments are comparable. The greatest element of V, if any, can be interpreted as corresponding to full acceptability of sentences, and the least element, if any, to their total rejection. The greatest element of W, if any, can be interpreted as corresponding to deductive (strict) inference, and the least, if any, to deductive rejection.

In the evaluation process, the (bottom) values of the first premises combine with the weights of the component inferences in an appropriate order, corresponding to the structure of the examined whole. Thus, by using suitable *operations* on both values and weights, the domain of the evaluation function is extended step by step to eventually obtain the (upper) value of the final conclusion. Per definition, this value is the *strength* of the focal argument.

This procedure can be viewed as an implementation of the RSA-triangle requirements (Johnson & Blair, 1977; see our Sect. 2). The initial evaluation function corresponds to premise acceptability, the weighing function corresponds to premise relevance, and the suitable combination of both values and weights regarding the structure of argument corresponds to premise sufficiency.

4.4 Evaluation of counter-arguments

Since counter-arguments are arguments, they too can be evaluated as separate arguments. But they may also be combined, or aggregated, with the arguments they themselves attack, so as to evaluate the aggregated whole (Selinger, 2019). The *relative strength* of a counter-argument can thus be defined as the "gap" between the strength of the attacked argument and the strength of the aggregated argument. Such an attack can be called *successful* if the value of the aggregate's conclusion does not belong to the set V^*.

4.5 Problematic issues

The components of examined arguments exhibit various kinds of logical interdependencies. These can impede the evaluation in particular *argumentation systems* that are defined by (i) specific sets of values and weights, (ii) their distribution, and (iii) operations on them. When gathering convergent arguments, this can result in overestimating argument strength (*double counting fallacy*), or in underestimating the overall acceptability of sets of sentences when computing the value of linked premises or undercutters (a dual form of the *double counting fallacy*). The recalculation of values that are already assigned to sentences (e.g., to some undermined first premises) is also vulnerable to this fallacy. On the other hand, if some interrelations among components are overlooked in the course of analysis, then their impact upon argument strength may remain uncounted at all. In specific cases, for example, one may well ask whether a rebuttal only rebuts some conclusion, or whether it also undercuts each (or some) of the convergent pro-arguments supporting the conclusion. After all, pro-arguments may constitute a kind of rebuttal to con-arguments, and perhaps pros also undercut cons.

5. THE EMPIRICAL APPROACH

5.1 Introduction

Rather than *define* argument strength, the empirical branch of communication studies *operationalizes* this concept. Experimental persuasion research regularly manipulates argument strength as a variable relevant to message content. In constructing this variable properly, special attention is paid to questions such as: How to properly pre-test strong/weak argument-stimuli, and how to operationalize the focal concept reliably?

5.2 The underlying idea of argument strength

Although lots of research within the empirical approach to argument strength has been carried out, there remains significant disagreement among researchers concerning the evaluation of this concept (Zhao & Cappella, 2016). One typically distinguishes a direct from an indirect way. The direct way relies on the pre-given features of argument strength, for instance the presence of evidence (Kononova et al., 2017). By contrast, the indirect way determines argument strength based on cognitive responses of a (sampled) population.

The indirect way is of particular interest because it presents a data-driven account that can be tailored to meet specific contextual or situational factors. An indirect pre-test procedure typically involves three steps: (1) develop a pool of (non-)cogent arguments regarding some attitude-object (e.g., an issue, person, place, etc.); (2) gather cognitive responses by sampling from a relevant population; (3) select those arguments with the highest and lowest acceptability ratings for further investigation.

A central problem is the reliability of the (open-ended or closed) techniques that this procedure incorporates. For instance, the main limitation of an open-ended technique such as thought listing—itself developed within the Elaboration Likelihood Model (ELM)—is the focus on a single dimension of argument strength, i.e. the valence of the population's thoughts. As Darke & Chaiken's (2005) study shows, some factors that interact with the valence of the population's thoughts are not included into the ELM method. To refine stimuli, the ELM generally elicits participants' thought-profiles in a high elaboration condition, where participants are asked to think carefully about a message (Petty & Cacioppo, 1986; Borgstede et al., 2017). If subjects list favourable thoughts with reference to it, then the message is perceived as strong, whereas a weak message is expected to generate unfavourable thoughts.

Several researchers have expressed doubts about the ELM's technique, because the valence of thoughts is not an exclusive indicator of argument strength (Zhao & Cappella, 2016). A number of attempts have also been made to either elaborate on the valence factor or to find alternative argument strength indicators (see Carpenter, 2015).

5.3 Extending the ELM evaluation

An operational definition of argument strength in the ELM omits the question whether a pre-test must standardize the relations between valence and personal gains and losses. In attempting to overcome the ELM's limitations, Darke & Chaiken (2005) do standardize those relations through a manipulation check. They suggest that evaluating a message (in terms of eliciting favourable, unfavourable or neutral thoughts) should give room for degrees of self-interest within the thought profile. Specifically, the self-interest component should be considered in relation to both 'immediate personal benefit onset' vs. 'delayed benefit onset' and 'immediate personal costs' vs. 'delayed costs'. In their manipulation check, messages producing immediate personal benefits were seen as stronger than those producing delayed benefits for the population. More so than the presence of immediate personal benefits, however, what proved to be a decisive factor for argument strength indication was the expectation of immediate personal costs, which decreased the strength of arguments irrespective of the kinds of benefits being presented.

Since experimental settings see the argument strength-variable interact with a host of other variables (Stiff & Mongeau, 2016), Darke and

Chaiken's manipulation provides a valuable insight into the relation between a thought-favourability aspect of argument strength and such additional message features as benefits and costs for the self. The main weakness of the manipulation, however, is that the presumed direction of the interaction effects does not reflect the comprehensive nature of argument strength itself. In other words, it is difficult to foresee whether similar persuasive effects would obtain if aspects other than valence and self-interest were manipulated.

5.4 A multi-item scale evaluation

In search of alternative methods, scholars have proposed multi-item scales that go beyond a single indicator of thought-favourability. For instance, Munch and Swasy (1988) pre-tested argument strength using a multi-item scale including strong vs. weak contraries, as well as five additional binary opposites for factors such as relevance, convincingness, importance, logicality and agreement. Unfortunately, they provide no data on internal consistency for the set of factors. A more comprehensive method (Lavine & Snyder, 1996) not only pre-tested ten factors for evaluating argument strength, it also showed a satisfactory level of internal consistency for the chosen set of factors.

For a given data set, both an internal consistency check and a confirmatory factor analysis (CFA) is required to tell how well the scale's model is suited for the specific data. Neither of the two pre-tests, however, was followed by checking the inter-factor reliability of the perceived argument strength scale or by the investigation of their factorial validity for the specific data.

Based on previous work, Zhao et al. (2011) offer a model of a multi-item scale and subsequently adjust it to evaluate the strength of arguments extracted from anti-drug and anti-smoking public service announcements (PSAs). To overcome one of the thought listing-procedure's potential limitations, viz. no capacity or need to provide precise thoughts among younger populations, Zhao et al. (2011) performed two studies in different age groups. In study 1, a model of a perceived argument strength scale was given to adolescents who assessed arguments from anti-drug PSAs. In study 2, the model was presented to adults who used it to evaluate arguments from anti-smoking PSAs. In both studies, the scale's model included nine indicators of argument strength (believability, novelty, convincingness, importance, confidence, friend, thoughts, agreement and reason).

These indicators were used to gather argument strength-ratings from a relevant population, and a CFA served to evaluate how well the multi-item model fitted the data. Inter-correlations of the perceived argument strength scale, Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) were also provided. As a result, only if CFI and RMSEA values were at an acceptable level did the model adequately fit the data. To improve the fit, variables with the lowest loadings on the latent factor were removed from the model. Although removing convincingness and novelty from the model in study 1, and novelty from study 2, improved the model's fit, the overall fit still remained unacceptable. Proper fit was obtained only after allowing for (easily explainable) error correlations in the CFA.

5.5 Problematic issues

In Zhao et al.'s (2011) multi-factor scale, itself a rating procedure that introduces a standardized model of argument strength indicators, the rationale for choosing a factor is generally unclear. Factors such as valence of thoughts, agreement, importance, believability and novelty belong to prior close-ended measures of argument strength. Although Zhao et al.'s scale is well-grounded in the experimental literature, its application is often based on an unsystematic choice of factors used in pre-tests of argument strength. It stands to reason that the choice of factors should instead be supported by using critical questions relevant to an argumentation type or scheme (Walton et al., 2008). Therefore, future research should investigate the factorial validity of indicators having been extracted from critical questions to better understand the nature of argument strength.

6. THE COMPUTATIONAL APPROACH

6.1 Overview

Whilst there are many computational models of argument, for example structured approaches reminiscent of those in Sect. 4, the computational perspective has predominantly, but not exclusively, focussed upon Dung's (1995) abstract approach to evaluating arguments. Rather than studying the exact statements used to express an argument and evaluating how those statements work together to strengthen or weaken a position, the computational approach instead abstracts away the linguistic detail, subsuming premise-conclusion structures into distinct, individual, atomic, abstract arguments. Each argument still has internal structure, the premises and conclusions still exist. But the focus in many abstract computational approaches has shifted from studying the nature of support *within* an argument, to studying the nature of conflict *between* arguments.

This approach is, admittedly, a gross simplification of the rich domain of argumentative discourse. But it enables powerful computational machinery to be deployed whilst recognising the limits of what computers can currently do. More plainly, human analysts can make leaps of intuition and tease out meaning from unexpressed, under-formed, or badly written texts, something that machines cannot yet achieve. By momentarily setting aside many of the rich historical, socio-psychological, and linguistic aspects of argumentation, we establish a basic context in which arguments can be handled computationally.

6.2 Conflicts & Semantics

Abstract arguments are studied in terms of their conflicts with other arguments, for example when one argument undercuts or rebuts another, leading to the notion that one argument can *attack* another argument (Walton, 2009). A network of such abstract arguments and attacks forms a directed graph and is referred to as an *argumentation framework*. The status of each argument can then be computed in relation to the pattern of attacks that exist within any given framework. For example, if an argument is attacked, and the attacker is in turn attacked, then the first argument is said to be *defended* by the second attacker. This idea of attack and defence can also be applied to sets of arguments so that an argument can be deemed *acceptable* to a set of arguments if the argument is defended by a member of that set. Any set of arguments can also be *conflict-free* if there are no attacks between members of the set.

Taken together, any set of arguments that is both conflict free and whose members are reflexively acceptable is an *admissible set*. These properties (conflict-freeness and acceptability) and the derived property (admissibility) subsequently play a central role in the identification of subsets of the overall framework, a process referred to as the application of a semantics. Semantics are, in essence, an evaluation process, determining which consistent groups of arguments, referred to as extensions, can be accepted.

Given that the application of a semantics can lead to multiple, consistent sets of arguments, which set of the multiple extensions should a person adopt as their position? This would appear to be a problem where the strength of arguments could play a role. The obvious answer might be: the strongest extension. However, there is very little explicit reference to the strength of arguments within the computational literature.

6.3 Notions of Strength

Nevertheless, some notions can be wrestled from the computational approach and shown to align with common concepts of strength. These relate to, firstly, the relationship between different semantics, secondly, the relationship between sets of arguments calculated by a given semantics, and finally, to the use of preferences to order the members of an argumentation framework.

If we pose the question, 'Given an argument framework, which argument is the strongest?', Dung and subsequent abstract argumentation authors actually have little to say on the subject. A *first* concept of argument strength could be based upon the output from a given semantics by applying the "gunfighter" analogy. Those arguments (gunfighters) that are acceptable (survive) at the end of the evaluation process (gunfight) can be deemed to be strong on the basis that they were not defeated. Put more plainly, for a given framework and a specified semantics, an argument that is acceptable is objectively strong.

A *second concept* of argument strength might be to compare how the set of arguments identified by one semantics can be considered to be stronger or weaker than those identified by another semantics. Note that there are many semantics, and an equivalent, slightly more expressive approach called labelling (Verheij, 1996). But these approaches all fit within a hierarchical organisation (Baroni et al., 2018) such that the stable semantics is a semi-stable semantics, which is in turn an instance of the preferred semantics, and so on until the base condition is reached with the conflict-free set. A difference in strength could be identified based upon the difference in requirements for set-membership under a semantics such that, for example, the members of an extension under the complete semantics are stronger arguments than those that are merely members of the admissible set.

In both these cases, however, the size of the extension could be very large, yet neither notion can tell us which individual argument within the extension is the strongest. This leads us to a *third concept* of argument strength. People are convinced for many reasons, whether by good or poor arguments, and often by things whose status as arguments we might debate. An argument that persuades one person might not persuade another, and might even be counterproductive. An argument defined as objectively strong under the gunfighter criteria, then, need not entail that that same argument would also be persuasive to an individual.

Recognizing that human psychology is problematic in this respect yields a more subjective sense of strength. It manifests in the computational approach through the use of preferences (Amgoud & Cayrol, 2002). Based upon a set of criteria, an ordering can be made over the members of any given set of arguments within a framework. Thus, when a set of acceptable arguments is identified, these can be arranged in a preferred order such that the most preferred argument can be deemed the strongest argument. One approach to defining an ordering over a set of arguments could stem from the kind of empirical approach discussed in Sect. 5. In this approach, however, neither is there an agreed and generally applicable method for constructing a preference ordering, nor a guarantee that the specific ordering is accurate for the audience in whose context it is being applied.

To summarise, *three concepts of argumentative strength* have been identified from the abstract argumentation literature: as a function of survival of the application of a semantics, as a result of comparison between semantics, and due to preference ordering. The first two of these concepts are objective, scalable, algorithmic, and generally applicable, but also highly granular—leading to possibly many, equivalent sets of "strong" arguments. In contrast, the third is a subjective, effort-intensive, more specific sense that can be finely tuned to an individual in terms of their preferences. In all cases, however, the computational literature rarely refers explicitly to the strength of an argument.

7. DISCUSSION

All five approaches are in some sense computational. In the Bayesian, the structural, and the strict computational approach, one first reconstructs structures by diagramming them, then computes some final value. (While the Bayesian approach rests on a probabilistic semantics, the other two approaches can deal with various, different semantics). On the dialectical approach, by contrast, analysts must first compute participants' possible moves, obligations, and scenarios. The subsequent reconstruction of argumentation as a diagram does not so much prepare the grounds for a computation; it rather is the computation itself. On the empirical approach, finally, computations are used to process, elaborate, and present empirical data.

The dialectical method prescribes, indeed only provides for, a procedural account of argumentative norms, including norms that operationalize argument strength. This gives rise to a responsibilist picture of argumentative norms expressed in terms of entitlements and obligations (i.e., permissible and required moves). A consequence is that the norms determining argument strength are accessible to the arguers, as norms they can apply themselves.

The empirical approach should, in order to avoid an unlimited, unsystematic, or unpredictable addition of factors to multi-item scales, rely more strongly on informal dialectical tools. As an example, the critical questions for argument schemes provide direction whenever empirical researchers must select plausible indicators for multi-item scales. For instance, one of the critical questions for the symptomatic argumentation scheme—"Is (characteristic) Y indeed typical of (property) Z ?"—motivates adding a believability-factor to empirical models that evaluate the strength of symptomatic argumentation. Extracting such factors from the critical questions thus operationalizes argument strength as the comparative degree of argument scheme-complicity.

Insights transfer not only from dialectical to empirical research. Empirical research also informs the third concept of argument strength on the computational approach, insofar as preference orderings would increase in validity were the preferences themselves based on dialectically-informed multi-item scales that fit well to data. Similarly, although the structural and the computational approach both abstract away from the order of dialogue moves, both nevertheless incorporate dialectical elements. One difference is that in Dung's frameworks the attack relation is primitive, while in the structural approach it must be defined. The Bayesian approach, by contrast, simply models attacks as evidence-nodes that, if relevant, bear on a claim's posterior probability.

8. SUMMARY

Each of the five approaches deals with the intricacies of modelling and evaluating an *order* among arguments given as their comparative strength.

The *Bayesian approach* models any argument's strength, provided that non-zero prior probabilities are assigned. Posterior probabilities then are derived given likelihoods (or vice versa). This approach "scales up" to networks of arguments. 'Argument strength' narrowly refers to the numerical difference, as measured, that credendes in reasons make to credences in claims. The all-things-considered *best* formal measure of argument strength (rightly) remains contended. The model's subjective "bent" shows when a *specific* reason-claim complex (RCC) rests on, or conveys, a single-event probability (as some RCC's do). But in argument evaluation, one cannot readily ground this probability in objective frequencies *other than* the number of those adhering to RCC's content. As a decisive evaluation criterion, however, this would only invite circularity.

On the *dialectical approach*, argument strength is modelled as the set of commitments, entitlements, and obligations pertaining to discussants at any stage in an argumentative dialogue. These collectively comprise the "move space" that discussants have available to them. Argument strength is operationalized as the (un)availability of discussant-moves that constrain further discussant-moves. Minimally, argument strength is a function of non-losing future participant moves: the strongest proponent-argument leaves no available opponent-move, while the weakest proponent-argument constrains *no* opponent-move, given the "move-space."

The structural approach's focal objects are separate inferences constituting atomic (simple or linked) arguments. Complex arguments are formed from atomic ones via syntactic operations (that correspond to convergent, divergent and serial arguments) (standard part), as enriched with counter-considerations and undercutters (dialectical extensions). To evaluate arguments, analysts combine the values of first premises with the weights of component inferences in an order corresponding to the structure of the examined whole. By using suitable operations on both values and weights, the domain of the evaluation function can be extended, to thus obtain the value of the final conclusion. This value represents the focal argument's *strength*.

On the empirical approach, analysts typically study argument strength either in a direct or an indirect way. The *direct way* relies on a prior notion of argument strength, determined, for instance, by the presence and quality of evidence). Rather than offer a measure of argument strength, the direct way presupposes one. The *indirect way* determines argument strength based on the cognitive responses of a (sampled) population. A data-driven account, it can be tailored to specific contextual factors. The challenge is to identify the set of measurement dimensions that provide an all-things-considered best model of perceived argument strength, i.e., an empirically adequate descriptive model that fits well to the data.

The computational approach rarely refers explicitly to the strength of an argument. One can nevertheless distinguish three concepts of argumentative strength: (i) as a function of survival of the application of a semantics, (ii) as a result of comparison between semantics, and (iii) due to preference orderings. The first two concepts are objective, scalable, algorithmic, and generally applicable. But their high granularity can possibly lead to many, equivalent sets of "strong" arguments. By contrast, the third concept offers a subjective, effort-intensive, and more specific sense of argument strength. It can be fine-tuned to individual arguers in terms of their preferences.

REFERENCES

Aikin, S. F., & Talisse, R. B. (2019). Epicureans on squandering life. *Three Quarks Daily*. URL=https://www.3quarksdaily.com/3quarksdaily/2019/02/epicureans-on-squandering-life.html (accessed 5 OCT 2019).

Amgoud, L. & Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments, *Annals of Mathematics and Artificial Intelligence*, 34 (1-3), 197-215.

Barth E., & Krabbe, E. C. W. (1982). *From axiom to dialogue*. Berlin: de Gruyter.

Baroni, P., Caminada, M., & Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. *Handbook of Formal Argumentation* (chapter 4, pp. 159-236.) London: College Publications.

Bayes, T. (1763/1958). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418. (Reprinted in *Biometrika*, 45, 296-315.)

Darke, P. R., & Chaiken, S. (2005). The pursuit of self-interest: Self-interest bias in attitude judgment and persuasion. *Journal of Personality and Social Psychology*, 89, 864–883.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in non-monotonic reasoning and logic programming and n-person games. *Artificial Intelligence*, 77, 321-357.

Finocchiaro, M. (2007). Arguments, meta-arguments, and metadialogues: A reconstruction of Krabbe, Govier, and Woods. *Argumentation*, 21, 253-268.

Finocchiaro, M. (2013). *Meta-argumentation: An approach to logic and argumentation theory*. London: College Publications.

Godden, D., and Zenker F. (2015). Denying antecedents and affirming consequents: The state of the art. *Informal Logic, 35*, 88-134.

Gordon, T. F., & Walton, D. (2006). The Carneades argumentation framework. In P. E. Dunne & T. J. M. Bench-Capon (Eds.), *Computational Models of Argument: Proceedings of COMMA 2006* (pp. 195-207). Amsterdam: IOS Press.

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

Hamblin, C. (1971). Mathematical models of dialogue. *Theoria*, 37, 130-155.

Johnson, R. H. & Blair, J. A., (1977). *Logical self-defense*. Toronto: McGraw-Hill Ryerson.

Keiff, L. (2011). Dialogue logic. In E. Zalta et al. (Eds.), *Stanford encyclopedia of philosophy*. Stanford, CA: Center for the Study of Language and Information. http://plato.stanford.edu/archives/sum2011/entries/logic-dialogical/ (accessed 5 OCT 2019)

Korb, K. (2004). Bayesian informal logic and fallacy. *Informal Logic*, *24*, 41-70.

Krabbe, E. C. W. (1992). So what? Profiles for relevance criticism in persuasion dialogue. *Argumentation*, 6, 271-283.

Krabbe, E. C. W. (1999). Profiles of dialogue. In J. Gerbrandy, M., Marx, M. de Rijke, & Y. Venema (Eds.), *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th birthday* (pp. 25–36). Amsterdam: Amsterdam University Press.

Krabbe, E. C. W. (2002). Profiles of dialogue as a dialectical tool. In F.H. van Eemeren (Ed.), *Advances in pragma-dialectics* (pp. 153–167). Amsterdam-Newport News, VA: Sic Sat/Vale Press.

Krabbe E. C. W. (2003). Metadialogues In: van Eemeren F.H., J.A. Blair, C. Willard C., and A.F. Snoeck Henkemans (Eds.), *Anyone who has a view* (pp. 83-90). Dordrecht: Springer.

Krabbe, E. C. W. (2013). The topical roots of formal dialectic. *Argumentation*, 27, 71–87.

Krabbe, E.C.W. (2017). The formalization of a critical discussion. *Argumentation*, 31, 101–119.

Krabbe, E. C. W., & Walton, D. (2011). Formal dialectical systems and their uses in the study of argumentation. In E. Feteris, B. Garssen, & A.F. Snoeck Henkemans (Eds.), *Keeping in touch with pragma-dialectics* (pp. 245-263). Amsterdam: John Benjamins.

Kononova, A., Yuan, S., & Joo, E. (2017). Reading about the flu online: How health-protective behavioral intentions are influenced by media multitasking, polychronicity, and strength of health-related arguments. *Health Communication*, 32(6), 759-767.

Lavine, H. & Snyder, M. (1996). Cognitive Processing and the Functional Effect in Persuasion: The Mediating Role of Subjective Perceptions of Message Quality. *Journal of Experimental Social Psychology* 32(6), 580–604.
Munch, J. M., & Swasy, J. L. (1988). Rhetorical question, summarization frequency, and argument strength effects on recall. *Journal of Consumer Research*, 15, 69-76.
Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal for Experimental Psychology*, 58, 75-85.
Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1, 93–124.
Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to persuasion.* New York, NY: Springer-Verlag.
Pfeifer, N. (2013). On argument strength. In F. Zenker (Ed.), *Bayesian argumentation: The practical side of probability* (pp. 185-193). Dordrecht: Springer.
Quine, W. V. O. (1961). Two dogmas of empiricism. In *From a logical point of view* (pp. 20-46). Cambridge, MA: Harvard UP.
Quine, W. V. O. (1992). *Pursuit of truth*. Cambridge, MA: Harvard UP.
Rescher, N. (1977). *Dialectics: A controversy-oriented approach to the theory of knowledge*. Albany, NY: SUNY Press.
Selinger, M. (2019). Towards defining the relation of attack. In B. Garssen, D. Godden, G. R. Mitchell, & J. H. M. Wagemans (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation* (pp. 1064-1073). Amsterdam: Sic Sat.
Spohn, W. (2012). *The Laws of Belief*. Oxford: Oxford University Press.
Stiff, J. B. & Mongeau, P. A. (2016). *Persuasive communication*. New York: The Guilford Press.
Tokarz, M. (2006). *Argumentacja, perswazja, manipulacja* [*Argumentation, persuasion, manipulation*]. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation*. Cambridge: Cambridge UP.
van Eemeren, F.H., Garssen, B., Krabbe, E.C.W., Snoeck Henkemans, A.F., Verheij, B., & Wagemans, J. (2014). Formal dialectical approaches. In van Eemeren et al. (Eds.) *Handbook of argumentation theory* (ch.6, pp. 301-372). Dordrecht: Springer.
von Borgstede, C., Andersson, M., & Hansla, A. (2014). Value-congruent information processing: The role of issue involvement and argument strength. *Basic and Applied Social Psychology,* 36(6), 461-477.
Verheij, B. (1996). Two approaches to dialectical argumentation. Admissible sets and argumentation stages. *Proceedings of the eighth Dutch conference on artificial intelligence* (NAIC'96) (pp. 357-368)..Utrecht: Utrecht University.
Walton, D. (1984). *Logical dialogue-games and fallacies*. London: Lanham.
Walton, D. (2009). Objections, rebuttals and refutations. *Argument Cultures: Proceedings of OSSA 09*, CD-ROM (pp. 1-10), Windsor, ON: OSSA.
Walton, D. (2011). Modeling burden of proof in a modified Hamblin dialogue system. *Informal Logic*, 31, 279-304.

Walton, D. (2015). Profiles of dialogue: A method of argument fault diagnosis and repair. *Argumentation & Advocacy*, 52, 91-108.

Walton, D. & Krabbe, E.C.W. (1995). *Commitment in dialogue*. Albany: SUNY Press.

Walton, D., Reed, C. & Macagno, F. (2008). *Argumentation schemes*. New York: Cambridge University Press.

Zenker, F. (Ed.) (2013.) *Bayesian argumentation: The practical side of probability*. Dordrecht: Springer.

Zenker, F. (2015). Denying the antecedent probabilized: a dialectical view. In: Eemeren, F.H. van, Garssen, B., Godden, D, and Mitchell, G. (eds). *Proceedings of the 8th International Conference of the International Society for the Study of Argumentation* (ISSA), July 2014 (pp. 1–14). Amsterdam: Rozenberg.

Zhao. X., Strasser, A., Cappella, J. N., Lerman, C., & Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. *Communication Methods & Measures*, 5, 48-75.

Zhao, X., & Cappella, J. N. (2016). Perceived argument strength. In D. K. Kim & J. Dearing (eds.), *Health Communication Research Measures* (pp. 119-126). New York: Peter Lang.