



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Rational interpretation of numerical quantity in argumentative contexts

Citation for published version:

Cummins, C & Franke, M 2021, 'Rational interpretation of numerical quantity in argumentative contexts', *Frontiers in Communication*, vol. 6, 662027. <https://doi.org/10.3389/fcomm.2021.662027>

Digital Object Identifier (DOI):

[10.3389/fcomm.2021.662027](https://doi.org/10.3389/fcomm.2021.662027)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Frontiers in Communication

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Rational Interpretation of Numerical Quantity in Argumentative Contexts

Chris Cummins^{1*} and Michael Franke²

¹Linguistics and English Language, University of Edinburgh, Edinburgh, United Kingdom, ²Institute for Cognitive Science, University of Osnabrück, Osnabrück, Germany

Numerical descriptions furnish us with an apparently precise and objective way of summarising complex datasets. In practice, the issue is less clear-cut, partly because the use of numerical expressions in natural language invites inferences that go beyond their mathematical meaning, and consequently quantitative descriptions can be true but misleading. This raises important practical questions for the hearer: how should they interpret a quantitative description that is being used to further a particular argumentative agenda, and to what extent should they treat it as a good argument for a particular conclusion? In this paper, we discuss this issue with reference to notions of argumentative strength, and consider the strategy that a rational hearer should adopt in interpreting quantitative information that is being used argumentatively by the speaker. We exemplify this with reference to United Kingdom universities' reporting of their REF 2014 evaluations. We argue that this reporting is typical of argumentative discourse involving quantitative information in two important respects. Firstly, a hearer must take into account the speaker's agenda in order not to be misled by the information provided; but secondly, the speaker's choice of utterance is typically suboptimal in its argumentative strength, and this creates a considerable challenge for accurate interpretation.

Keywords: pragmatic inference, argumentative language use, non-cooperative dialogue, argument strength, information selection, quantity expressions

INTRODUCTION

How should a rational hearer interpret a statement of numerical quantity, such as 1)?

- 1) More than 30 states voted Democrat in the 1996 United States Presidential election.

Assuming that the speaker is accurate, the hearer can begin by deriving the semantic meaning of the quantity expression, and arrive at the interpretation that the cardinality of the set of Democrat-voting states in the 1996 election is greater than 30. If the hearer is willing to make additional assumptions about the speaker's cooperativity and knowledgeability, they can derive additional pragmatic inferences. Specifically, they can potentially infer that the speaker is unable to assert informationally stronger alternatives to 1), and hence either that these alternatives are false or that the speaker is ignorant as to their truth-value. In this case, informationally stronger alternatives potentially include those which give larger or more precise numbers (*more than 40, 35*) or which describe wider date ranges (*in every Presidential election*).

But what if the speaker is strategic, in the sense that they wish to present information that will optimally support a particular argumentative agenda? For the rational hearer, this creates both a problem and an opportunity. On the one hand, the standard pragmatic inferences mentioned above

OPEN ACCESS

Edited by:

Matthew W. Crocker,
Saarland University, Germany

Reviewed by:

Christoph Hesse,
Leibniz Center for General Linguistics
(ZAS), Germany
Anouschka Foltz,
University of Graz, Austria

*Correspondence:

Chris Cummins
c.r.cummins@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 31 January 2021

Accepted: 27 April 2021

Published: 19 May 2021

Citation:

Cummins C and Franke M (2021)
Rational Interpretation of Numerical
Quantity in Argumentative Contexts.
Front. Commun. 6:662027.
doi: 10.3389/fcomm.2021.662027

may be unavailable, on the basis that the speaker may simply be declining to utter stronger alternatives that are known to be true, for purely strategic reasons. Thus, in 1), perhaps the speaker wishes to discuss the results of the 1996 election in isolation, in order to make a point about the relative strength of the candidates that particular year. On the other hand, if the speaker is known to be pursuing a particular argumentative agenda, this opens up the possibility of the hearer drawing inferences about the falsity of alternatives that would have been argumentatively stronger, whether or not these are informationally stronger in the usual pragmatic sense. For instance, a speaker who wished to argue that the Democrats can win a comfortable majority of states might choose to discuss the most recent example of them doing so, in which case in 1) they would have said *2008* rather than *1996* if the resulting sentence had still been true.

In this paper, we outline issues of rational use of language in argumentative discourse. Rational communication in non-cooperative contexts has been studied before, e.g., from the perspective of game theory (Franke et al., 2012; de Jaegher and van Rooij, 2014) and also via experimental methods (Franke et al., 2020). The argumentative dimension has been stressed as an important perspective on language use (Anscombe and Ducrot, 1983) that offers an alternative to purely information-based accounts of interaction. It has been used to explain a variety of natural language phenomena, such as the meaning and distribution of particles like *also* and *even* (Merin, 1999) or that of adversarial connectives such as *but* (Winterstein, 2012). Here, we focus specifically on argumentative language use in the domain of numerical quantity expressions. We first survey some of the relevant issues in current research on the semantics and pragmatics of numerical quantity, under standard assumptions about cooperativity in *Standard Semantic and Pragmatic Meanings of Numerical Expressions*. We then discuss, in *Argumentative Framing for a Single Numerical Quantity*, how argumentative motives affect a speaker's choice of utterance when describing a single numerical quantity. *Argumentative Framing for Complex Information States With Complex Utterances* extends these considerations to more complex cases where more than one numerical feature is potentially relevant for argumentative framing. *Quantifying Argumentative Strength, and Allowing for Uncooperativity* then introduces a notion of argumentative strength, following Merin (1999), which aims to subsume the considerations laid out in the foregoing discussion. *A Case Study: Reporting the Research Excellence Framework* subsequently derives some more concrete predictions of this approach and tests them with reference to a small corpus of argumentative usages of quantity expressions, drawn from the public statements made by United Kingdom universities concerning their rankings in the 2014 Research Excellence Framework (REF). We show that these usages can usefully be understood by appeal to the notion of argumentativity that we propose, but also that they present a particular interpretive challenge to the hearer as a consequence of their argumentative strength typically being suboptimal.

STANDARD SEMANTIC AND PRAGMATIC MEANINGS OF NUMERICAL EXPRESSIONS

It is tempting to assume that expressions of numerical quantity will be easy to formalise semantically. However, as the enduring debates in the semantics and pragmatics literature testify, many turn out on closer inspection to require sophisticated and subtle analyses. The question of how to formalise these meanings is important for semantic and pragmatic theory, but also for real-life communication, given the crucial role that number plays in conveying precise information that feeds into high-stakes decision-making.

As used in natural language, 'bare' (unmodified) numerals already admit multiple possible interpretations. Horn (1972) noted that bare numerals can express both exact readings, as in 2), and lower-bound ("at least") readings, as seems to be preferred for 3). In some cases, as pointed out by Carston (1998), bare numerals appear to contribute to upper-bound ("at most") readings, as in 4). And round numbers in particular can also convey approximate meanings, as discussed by Krifka (2009), as in 5), which is widely judged true, or at least true enough, if the number of people in the room is, for instance, 99 or 101 (Lasersohn, 1999).

- 2) I have three children.
- 3) People with three children are entitled to extra benefits.
- 4) You can have 2000 calories without putting on weight.
- 5) There are a hundred people in the room.

This ambiguity creates a potential challenge for the hearer: are they able to recover the speaker's intended meaning, given that this is not linguistically signalled? This is a widespread issue. Taking the case of approximate readings as in 5), speakers frequently round values before reporting them, and do not typically state that they have done so (for instance in telling the time, e.g. 7:30pm, cf. Van der Henst et al., 2002; and indeed in providing summary statistics for an experiment, e.g. "mean RT = 345 ms"). Hence, the way bare numerals are routinely interpreted in natural language gives rise to some pitfalls when we attempt to convey information with them at any given level of precision.

When speakers use modified numerals such as *more than/at most/up to 100*, a different set of issues arises. The ambiguity discussed above does not occur, as pointed out by Solt (2014): in this case, the semantic meaning contributed by the numeral is clearly exact. This imposes an additional constraint on the speaker. For instance, if there are 98 people in the room, a speaker can utter 5) and be judged to have told the truth, but if one further person then entered the room, a speaker who uttered 6) would still be judged to have spoken falsely, because *100* is interpreted as obligatorily exact in 6). That is to say, *more than 100* means *more than precisely 100* rather than merely *more than are present in a situation of which '100' could be truthfully asserted*.

- 6) There are more than 100 people in the room.

However, a different kind of ambiguity, at a pragmatic level, arises from utterances such as 6). In addition to conveying a (semantic) lower bound on the possible value under discussion, an expression such as *more than 100* appears to convey a (pragmatic) upper bound (Cummins et al., 2012). For instance, the utterance of 6) typically appears to convey the falsity of 7).

7) There are more than 200 people in the room.

Cummins et al. (2012) propose that these enriched meanings can be treated as quantity implicatures, and more specifically scalar implicatures: the use of *more than 100* implicates the falsity of the corresponding sentence with the stronger scalar alternative *more than 200*. But this analysis predicts further scope for misunderstanding between speaker and hearer, as it is not clear which stronger alternatives should be considered to have been negated. Should *more than 100* be taken to convey the falsity of *more than 110*, *more than 125*, *more than 150*, or none of these?

A partial solution to this problem, in the spirit of traditional approaches to scalar implicature, is to argue that the relevant stronger alternatives—which give rise to implicatures—involve numerals which are at least as salient as the original numeral. The notion of granularity, as discussed by Krifka (2009), offers one way of fleshing out this idea. The idea is that round numbers are scale points of scales with differing granularities—60 is at once a scale point in scales graduated by units, tens, perhaps twenties, and so on—and only numbers which are scale points on equally coarse-grained scales constitute scalar alternatives.

However, the limits of this approach are clear. As applied to round numbers in neutral contexts, the hearer still needs to understand which scale a speaker means to evoke—when they say *more than 100*, are they thinking of 100 as a scale point on a scale of tens, or 25s, or 100s? This will determine whether the scalar alternative is *more than 110*, *more than 125*, or *more than 200*. Various considerations might influence how hearers attempt to resolve this problem (see Hesse and Benz, 2020). And specific contexts may be associated with particular scales which supervene. For instance, salient milestones in the United Kingdom Singles Chart traditionally include Top 75 and Top 40, but not Top 50: a song that peaked at #48 could reasonably just be called a Top 75 hit, contrary to the predictions of a general granularity-driven account.

Both at a semantic and pragmatic level, then, the interpretation of numerical expressions creates challenges for the hearer, as the speaker is not obliged to signal the precise sense in which they intend a numeral to be interpreted. And so far we have assumed throughout that we are dealing with a cooperative discourse environment, in which the speaker intends their message to be perfectly reconstructed by the hearer.

What about discourses that are not fully cooperative in the sense of aiming for accurate, precise information transmission? Suppose, in particular, that the speaker wishes the hearer to get a false impression about a particular quantity. We have already seen how this situation might arise by accident—the hearer might take a precise numeral to be an approximation, a lower-bound numeral to be precise, or a modified numeral to give rise to an

implicature that was not intended. Can an argumentative speaker exploit these natural possibilities for misunderstanding in order to mislead the hearer in a particular direction? And if so, how should a rational hearer respond in order not to be misled?

The following sections look in more detail at the interplay between, on the one hand, the pragmatic interpretation of quantity words as studied in the context of standard information-seeking cooperative discourse, and, on the other hand, a speaker's interest in presenting a known state of affairs in a particularly favourable light. *Argumentative Framing for a Single Numerical Quantity* looks at the arguably more basic case in which the relevant information is just a single numerical quality, and the speaker knows this precisely, but wishes the hearer to perceive it to be as high as possible. *Argumentative Framing for Complex Information States With Complex Utterances* extends this analysis to more complex situations where more than one feature matters for the speaker's argumentative framing.

ARGUMENTATIVE FRAMING FOR A SINGLE NUMERICAL QUANTITY

The goal of this section is to investigate how the pragmatic inferences discussed in the previous section, stemming from the usually assumed ideal of a cooperative information-conveying discourse, may be exploited by a speaker who knows the true value N of some numerical property but wishes to induce in the hearer an impression that this quantity is in fact higher than N . We refer to this situation as *high-framing* of a single quantity. We first look at possibilities of high-framing of a single quantity by using pragmatic slack, or pragmatic halos, associated with unmodified numerals in *Exploiting Pragmatic Slack in Round Bare Numerals for High-Framing*. *Exploiting the Imprecision of Round Modified Numerals for High-Framing* then looks at roundness effects associated with modified numerals. Finally, *The Potential Sub-Optimality of Non-Round Numbers* explores the potential sub-optimality of using precise non-round number terms for high-framing.

Exploiting Pragmatic Slack in Round Bare Numerals for High-Framing

Suppose that a speaker, fully knowledgeable about a precise numerical quantity N , wishes to give a hearer a maximal impression of this quantity without speaking falsely¹. What strategies might they adopt, given what we know about the interpretation of numerical quantity expressions?

One option is to make good use of imprecision and pragmatic slack. If N is just below a round number, the speaker might try

¹We assume throughout that we are dealing with speakers who are disposed to be honest, in the minimal sense of not making assertions that could be judged semantically false. However, this leaves open the possibility that such speakers may choose to mislead their hearers pragmatically, appealing to plausible deniability (Pinker et al., 2008).

using that round number M : for instance, uttering 5) when there are in fact 98 people in the room. The hearer might interpret this as exact, or better yet from the speaker's point of view, as a lower-bound, i.e. as a commitment on the speaker's part to the existence of a set of 100 people who are in the room.

However, if the true attendance were 102, uttering 5) would risk the hearer getting a needlessly low impression of it, contrary to the speaker's interests; and if the attendance were not within the 'pragmatic halo' (Lasersohn, 1999) of 100, the speaker could not truthfully utter 5) at all.

In sum, we expect high-framing speakers who know true N to be able to use pragmatic slack to their advantage in the following way: use round number $M > N$ to describe N if N is plausibly contained in a pragmatic halo around M .

Exploiting the Imprecision of Round Modified Numerals for High-Framing

A related but perhaps more powerful means of high-framing is to use round modified numerals which ensure a lower-bound interpretation in the semantics. For example, if N is the true known number, the high-framing speaker can use *more than M* , relating the quantity under discussion to some reference point M . Semantically, it would be natural to choose M to be as large as possible, thus ruling out as many (low) potential values as possible. However, pragmatically, as discussed above, the optimal choice of M is not straightforward, because *more than M* can implicate *not more than O* for various values $O > M$. Indeed, according to Cummins et al. (2012), the values that hearers associate with the description *more than 110* may be generally lower than those they associate with *more than 100* (although Hesse and Benz, 2020, have apparently conflicting data on this point). If this is so, a speaker wishing to emphasise the largeness of a crowd of 111 might be better off uttering 6), repeated below, rather than the semantically stronger 8).

8) There are more than 110 people in the room.

On a granularity-based account, this counterintuitive result arises because 8) effectively leaks information about the level of precision at which the speaker is operating—it seems highly likely that the speaker of 8) would have uttered 9) if they could do so. By contrast, it is not clear that the speaker of 6) is operating at such a fine-grained level, and they might not utter 9) even if they knew it to be true. Hence, the hearer may be more confident that 8) implicates the falsity of 9) than they could be that 6) implicates the falsity of 9).

9) There are more than 120 people in the room.

We conclude that speakers may choose to describe true known N for the purpose of high-framing by using a modified numeral like *more than M* , which semantically only contains a lower-bound. If they do so, they should select M in such a way that the expected pragmatic interpretation of *more than M* conveys higher values in information-seeking cooperative discourse than any

other reference point or round number $M' < N$ would in the phrase *more than M'* .

The Potential Sub-Optimality of Non-Round Numbers

So far, we have focused on round numbers and their potential usefulness for high-framing. Let us now consider whether high-framing might benefit from the use of non-round numbers.

We note first that, even with non-round numbers, the speaker can convey additional quantity information, such as in 10) where the non-round 19 is selected as the endpoint of a particular range.

10) If restored to operation, it would be one of the 19 largest telescopes existing today, all of which are in constant demand (<https://www.nytimes.com/1988/11/15/science/volunteers-look-revival-of-famed-telescope.html>, retrieved 24/03/20).

Describing the telescope as *one of the 19 largest* rather than *one of the 20 largest* clearly makes a semantically stronger claim, which supports the speaker's apparent point that it would be an exceptionally large telescope. However, using 19 rather than 20 invites the hearer to draw inferences about the motivation for this precise choice—an available inference in this case being that the telescope would rank precisely 19th in size (unless there is some reason why we should care about precisely the 19 largest telescopes in particular). If the hearer infers this, the speaker has perhaps been less argumentatively effective than if the hearer had merely concluded that the telescope would be somewhere among the largest 20.

Similarly, in 11), the use of *top 19* strongly invites the inference that the salient stronger (given the entailment direction of the utterance) alternative *top 20* doesn't hold—i.e. that the team currently 20th in the CFP rankings, like Clemson, has not faced a team currently in the committee's top 25, which in turn suggests that Clemson's status is less special than the speaker seems to want to suggest.

11) Clemson is the only team among the top 19 in the CFP rankings that hasn't faced a team currently in the committee's top 25 (https://www.espn.co.uk/college-football/story/_/id/28196686/dabo-swinney-says-clemson-held-different-standard-cfp-voters, retrieved 24/03/20)

A similarly complex example occurs in 12).

12) Disappointingly, 10 of the world's 19 most unequal countries are in sub-Saharan Africa (<https://www.un.org/africarenewal/magazine/december-2017-march-2018/closing-africa%E2%80%99s-wealth-gap>, retrieved 24/03/20).

Here, by similar reasoning, the hearer can infer that 19 could not be replaced by 18, as otherwise the speaker would have done so. It follows that the 19th most unequal country in the world is in sub-Saharan Africa, and thus only nine of the world's 18 most

unequal countries are in that region. This is presumably considered to be a less compelling argument for the speaker's overall thesis than *10 of the world's 19*, as otherwise they would have uttered it in the first place.

In each of these cases, then, choosing the semantically strongest description invites pragmatic inferences which appear to push back against the speaker's argumentative goals (namely, in 11), that Clemson is distinguished by its lack of strong opposition so far, and in 12) that inequality is widespread in sub-Saharan Africa). Of course, the extent to which hearers draw these inferences is an empirical question, so it is not self-evident that these utterances constitute less effective arguments than informationally weaker alternatives would (for instance, *one of only two teams in the top 20*, or *10 of the world's 20*, respectively). However, it is equally unclear that they constitute better arguments than informationally weaker alternatives would.

In summary, then, the use of non-salient numbers in utterances such as 10)–12) invites inferences about the falsity of corresponding stronger statements involving more salient numbers. For this reason, we might expect non-salient numbers to be generally poor choices for high-framing.

ARGUMENTATIVE FRAMING FOR COMPLEX INFORMATION STATES WITH COMPLEX UTTERANCES

Examples 11) and 12) begin to show some of the complexity that is typical of argumentative language use. In these, unlike the previous examples, the speaker is not merely expressing one quantity as to make it sound large or small: rather, they have chosen two numbers with which to make a particular argument. In 12), the speaker has not only chosen the frame *X of the world's Y* but has made a deliberate choice about how to populate it, out of all the possible number pairs (X, Y) that would make the sentence true, and has presumably chosen numbers which they feel are rhetorically effective.

The broader point that this illustrates is that a speaker citing complex data in support of their argument can do so in many ways. An effective choice may invite the hearer to draw additional inferences that support the speaker's argument². On the flip side, an ineffective choice may invite the hearer to draw inferences that undermine the speaker's argument. Bill Bryson (1998): 112f describes drawing just such inferences in response to a car advertisement:

"[The advert] says something like 'The new Dodge Backfire. Rated number one against the Chrysler Inert for handling. Rated number one against the Plymouth Repellent for mileage. Rated number one against the Ford Eczema for repair costs.' As you will notice . . . in each category the Dodge is rated against only one other competitor. . . . [I]f the Dodge were rated top against ten or

twelve or fifteen competitors in any of those categories, then presumably the ad would have said so. Because it doesn't say so, one must naturally conclude that the Dodge performed worse than all its competitors except the one cited."

In this scenario, the sceptical hearer's inferences derive ultimately from the perception that a knowledgeable speaker, with a particular argumentative agenda, has chosen to present a very limited amount of information. The hearer infers that this reflects a strategic decision, motivated by the fact that presenting additional information (how the Dodge compares to the Chrysler in mileage, etc.) would undermine the speaker's broader communicative goal (presenting the Dodge as the most attractive choice).

From the standpoint of pragmatic analysis, we could formalise this idea by noting that the advert, as described, would give rise to a series of ad hoc implicatures to the effect that the Dodge is inferior to the Chrysler and Plymouth (and perhaps other competitors) in repair costs, inferior to the Chrysler and the Ford (and perhaps other competitors) in mileage, and inferior to the Plymouth and the Ford (and perhaps other competitors) in handling. These ad hoc implicatures are proposed to arise on the basis that entailment relations exist between sentence pairs such as 13) and 14), with 14) entailing 13); and given a context in which the stronger sentence 14) would be relevant, the utterance of the weaker sentence 13) is taken to implicate the stronger sentence's falsity each time.

- 13) The Dodge is rated higher than the Chrysler for handling.
- 14) The Dodge is rated higher than the Chrysler and the Plymouth for handling.

Given a sufficiently complex set of quantitative data, the set of true statements that could be made about the data will be very large. Under these circumstances, the speaker's decision to say whatever they decide to say, rather than any of the alternatives, could give rise to a rich array of inferences. As an example, consider a scenario in which 15) and 16) would each be plausible descriptions of a situation.

- 15) All of the students got some of the questions right.
- 16) Some of the students got all of the questions right.

In purely semantic terms, neither of these sentences is strictly more informative than the other, in the sense that no entailment relation obtains between them. However, a hearer might feel that one of them is more valuable than the other, as a conversational contribution, in a world where both are true. Suppose that such a hearer thinks that 16) is clearly the more valuable option. They should then take the utterance of 15) by a knowledgeable speaker to convey the negation of 16). An argumentative speaker who is aware of the hearer's preference can then potentially exploit it: they can cause the hearer to believe that 16) is false (perhaps incorrectly) by asserting 15).

In its effect, this would be much like a speaker asserting *some* in order to convey *not all* when they know that *all* is the case. But a speaker who asserts *some* when they know that *all* is the case could be argued to be dishonest, because there is a widespread

²As discussed earlier, whether or not those inferences are true may not be important to the argumentative speaker, although the speaker may wish them to be covered by plausible deniability.

understanding that *some* typically conveys *not all* in declarative contexts—a point discussed in more detail by Meibauer (2014) and Franke et al. (2020). By contrast, a speaker who asserts 15) in order to (misleadingly) convey the falsity of 16) might have some measure of plausible deniability against the claim of dishonesty, because speakers and hearers do not share contextually stable intuitions about the relative usefulness of these two possible utterances.

In summary, the above examples suggest that the effectiveness of a particular utterance, construed as an argument towards a particular goal, depends both on the semantic content of the utterance and the pragmatic inferences drawn by the hearer as a result of the utterance. Moreover, the eventual interpretation of a hearer who takes into account that the speaker has an argumentative agenda may diverge considerably from the pragmatic interpretation that they would be predicted to arrive at in cooperative contexts. Consequently, the usual tools with which we analyse the semantics and pragmatics of cooperative discourse are of limited use in helping us to systematise these ideas. In the following section, we explore how we can address this challenge by appeal to the notion of argumentative strength.

QUANTIFYING ARGUMENTATIVE STRENGTH, AND ALLOWING FOR UNCOOPERATIVITY

In the context of cooperative communication, we can use ideas around informativity and relevance to quantify the extent to which a candidate utterance would be a useful contribution to the discourse, in the sense of bringing about positive cognitive effects in the hearer, in Sperber and Wilson's (1986) terms. Somewhat analogously, given a (not necessarily cooperative) situation in which a speaker wishes to make a particular point, we can explore their choice of utterance by considering the extent to which candidate utterances would represent good arguments in support of that point. In the following we therefore explore a quantitative measure of argumentative strength of an utterance and consider the predictions that it makes about usage under various different assumptions. In *Argumentative Strength for a Semantic Interpretation of an Utterance* we consider argument strength in the case where hearers adopt a purely semantic interpretation of the speaker's utterance, and in *Argumentative Strength for a Pragmatic Interpretation of an Utterance* we expand this to the case where hearers are presumed to take into account the usual pragmatic inferences that would be available in a cooperative context. In *Argumentative Strength for Complex Cases* we exemplify how complex contexts invite the speaker to be more selective in their choice of utterance than standard pragmatic theories usually accommodate. Finally, in *Rational Interpretation in an Argumentative Context*, we consider the perspective of a sceptical hearer confronted with a speaker who is selective in this way, and examine how argumentative strength can be evaluated in this kind of non-cooperative setting.

Argumentative Strength for a Semantic Interpretation of an Utterance

Working within the tradition of argumentative approaches to language (Anscombe and Ducrot, 1983), Merin (1999) proposes to model the argumentative strength (arg_str) of an utterance u with reference to the weight of evidence that it provides in support of the speaker's communicative goal hypothesis G . This notion of weight of evidence can be unpacked (following Good, 1950, and others) as a log-likelihood ratio as in Eq. (17)³.

$$\text{arg_str}(u, G) = \log \frac{P(u|G)}{P(u|\neg G)} \quad (17)$$

Here, $P(u|G)$ denotes the probability that utterance u is true if hypothesis G is true, and $P(u|\neg G)$ denotes the probability that utterance u is true if hypothesis G is false. The idea is that an utterance with a positive argumentative strength with respect to hypothesis G is, by definition, one that is more likely to be true if G is true than it is to be true if G is false⁴.

For simple examples, it is easy to evaluate argumentative strength according to this measure. For instance, 18) has positive (indeed, infinitely large) argumentative strength in support of the contention G that the Poincaré conjecture holds, because $P(u|\neg G) = 0$ and $P(u|G) > 0$.

- 18) Grigori Perelman proved the Poincaré conjecture in 2006.

However, in more complex cases, it can be difficult to precisely calculate argumentative strength, while it is still possible to evaluate at least qualitative predictions based on intuition. To illustrate this, we can revisit 11), repeated here (omitting *disappointingly*) as 19). We might take it that the speaker's communicative goal in this context is something like 20).

- 19) 10 of the world's 19 most unequal countries are in sub-Saharan Africa.
20) Inequality is widespread in sub-Saharan Africa.

The argumentative strength of the utterance, as defined above, is calculated from the probability that 19) is true given 20) and the probability that 19) is true given the negation of 20). But the latter probability, in particular, is not readily calculable for speaker or hearer, because even if we define *widespread* crisply, *not widespread* clearly covers a range of values. However, the speaker and hearer may still have intuitions about the probabilistic relations between 19) and 20). For instance, we might say that 19) has positive argumentative

³The use of log-likelihood in Good's formalism ensures that the weight of evidence has desirable additive properties.

⁴Note that this definition of argumentative strength only makes reference to the truth conditions associated with the quantity expression. For the purposes of this paper, we do not explore the idea developed by Mira Ariel (2004) and subsequently that the use of particular quantity expressions is conventionally associated with particular argumentative effects, although this work is compatible with that idea: we could, for instance, take the nature of the quantity expression to inform the hearer's understanding of the identity of G .

strength with respect to 20) if 21) is judged more probable than 22), and negative argumentative strength if the reverse is true.

- 21) Inequality is widespread in sub-Saharan Africa, and 10 of the world's 19 most unequal countries are located there.
- 22) Inequality is not widespread in sub-Saharan Africa, and 10 of the world's 19 most unequal countries are located there.

By definition, an utterance with positive argumentative strength should constitute positive evidence in favour of the speaker's communicative goal G over its negation, and hence a rational hearer should respond to such an utterance by increasing the strength of their belief in G . However, as discussed in *Argumentative Framing for a Single Numerical Quantity*, an utterance might also give rise to pragmatic enrichments that would tend to oppose the argument being made by the speaker. This possibility is not taken into account in Eq. (17), which is concerned purely with the semantic content of the utterance.

Argumentative Strength for a Pragmatic Interpretation of an Utterance

To see how pragmatic inferences which are ordinarily associated with utterances of single-number descriptions (see *Standard Semantic and Pragmatic Meanings of Numerical Expressions* and *Argumentative Framing for a Single Numerical Quantity*) might affect the notion of argumentative strength, let us return to a simpler example. Suppose that we, as hearers, believe that our conference will be a success if and only if more than 120 people attend. Let S be the event that more than 120 people attend the conference, and assume that it is common knowledge that people will attend if and only if they have registered. A speaker who (privately) has the argumentative goal of convincing us that the conference will be a success then utters either 23) or 24).

- 23) More than 100 people registered.
- 24) More than 110 people registered.

On semantic grounds, $P(S|24) \geq P(S|23)$: that is to say, the probability that S is true given that 24) is true is at least as great as the probability that S is true given that 23) is true. This holds because S is false in all worlds in which 23) is true and 24) is false. Therefore 24) should be a better argument for S than 23) is. However, as discussed earlier, 24) strongly invites the pragmatic inference that S is false, which is arguably not true of 23). If this pragmatic analysis is correct, taking that inference into account may change the picture and result in 23) being a better argument than 24) for the truth of S .

The general point here, once again, is that utterances which are effective arguments on their semantics may not be effective when pragmatic enrichments are included in the calculation. It would be helpful to have a notion of argumentative strength that takes this into account. More precisely, if we include pragmatic considerations, what is necessary for argument strength is not merely that the utterance u should be more likely true given G than given not- G , but rather that u should be more likely felicitously

assertable—in the sense of both being true and not giving rise to false implicatures—given G than given not- G . Let $A(u)$ stand for the fact that u is felicitously assertable. We could then propose a notion of pragmatic argument strength (prag_arg_str) as in Eq. (25).

$$\text{prag_arg_str}(u, G) = \log \frac{P(A(u)|G)}{P(A(u)|\neg G)} \quad (25)$$

To illustrate how this works, we can flesh out the example of 23) and 24) further with some additional assumptions: these are not intended to be realistic, but just serve to illustrate the calculation process. Suppose there is a 90% probability of an utterance being interpreted as conveying a pragmatic enrichment, and that for *more than 100* that enrichment is *not more than 150* while for *more than 110* it is *not more than 120*. For simplicity let us suppose that no other pragmatic interpretations are in play. Suppose further that the true value under discussion—the number of people who have registered for the conference—is uniformly distributed on the range $[0, 200]$. Recall that S is the event that more than 120 people will attend, and we are assuming that it is common knowledge that they will attend if and only if they have registered.

According to the measure in Eq. (17), the argumentative strength of utterance u toward the goal G is the log of the ratio of $P(u|G)$ and $P(u|\neg G)$. Here, $G = S$, and we consider first the utterance *more than 100*. The probability that *more than 100* is true given that *more than 120* is true equals 1; the probability that *more than 100* is true given that *more than 120* is false equals $1/6$ here. Recall that we assume that the true value is uniformly distributed on $[0, 200]$ —if *more than 120* is false, it must lie in the range $[0, 120]$, again uniformly distributed. Hence the probability that it exceeds 100 is $20/120 = 1/6$. So, according to Eq. (17), the argumentative strength of *more than 100* is equal to $\log(1/(1/6)) = \log 6 \approx 0.78$. Now we consider the utterance *more than 110*. Again, the probability that *more than 100* is true given that *more than 110* is true equals 1; the probability that *more than 100* is true given that *more than 110* is false equals $1/11$ here. If *more than 110* is false, the true value is uniformly distributed on $[0, 110]$ and has a $1/11$ chance of exceeding 100. So, per Eq. (17), the argumentative strength of *more than 110* is equal to $\log(1/(1/11)) = \log 11 \approx 1.04$, which exceeds the argumentative strength of *more than 100*.

Now let us consider instead the measure in Eq. (25), under which the argumentative strength of utterance u towards the goal G is the log of the ratio of $P(A(u)|G)$ and $P(A(u)|\neg G)$. Again, $G = S$, and here we have adopted the assumptions that there is a 90% probability of the utterance being pragmatically interpreted, and that if it is, *more than 100* will be interpreted as *not more than 150* and *more than 110* will be interpreted as *not more than 120*. Consider first *more than 100*. This is assertable in two disjoint eventualities: i) it attracts a pragmatic interpretation and the true value lies in the range $(100, 150]^5$, or ii) it does not attract a

⁵Here we use $(100, 150]$ to refer to the half-open interval comprising values that are greater than 100 and less than or equal to 150, which are those values for which *more than 100* can be felicitously asserted if it implicates *not more than 150*.

pragmatic interpretation and the true value lies in the range (100, 200]. If *S* is true, then the probability that the true value lies in the range (100, 150] is $3/8$ (because it is uniformly distributed on (120, 200]), and the probability that the true value lies in the range (100, 200] is 1. So the total probability that *more than 100* is assertable is $(90\% \times 3/8 + 10\% \times 1) = 35/80$. If *S* is false, then the probability that the true value lies in the range (100, 150] is $1/6$ (because it is uniformly distributed on [0, 120]), and the probability that the true value lies in the range (100, 200] is also $1/6$. So the total probability that *more than 100* is assertable is $(90\% \times 1/6 + 10\% \times 1/6) = 1/6$. Hence, under the measure in Eq. (25), the argumentative strength of *more than 100* is $\log((35/80)/(1/6)) = \log(21/8) = 0.419$.

Now consider *more than 110*. This is assertable in two disjoint eventualities: i) it attracts a pragmatic interpretation and the true value lies in the range (110, 120], or ii) it does not attract a pragmatic interpretation and the true value lies in the range (110, 200]. If *S* is true, then the probability that the true value lies in the range (110, 120] is zero, and the probability that the true value lies in the range (110, 200] is 1. So the total probability that *more than 110* is assertable is $(90\% \times 0 + 10\% \times 1) = 1/10$ (or, to put it another way, *more than 110* is only assertable if it attracts no pragmatic enrichment, and we are assuming this to happen with $1/10$ probability in this illustration). If *S* is false, then the probability that the true value lies in the range (110, 120] is $1/12$, and the probability that the true value lies in the range (110, 200] is also $1/12$. So the total probability that *more than 110* is assertable is $(90\% \times 1/12 + 10\% \times 1/12) = 1/12$. Hence, under the measure in Eq. (25), the argumentative strength of *more than 110* is $\log((1/10)/(1/12)) = \log(6/5) = 0.079$, which is lower than for *more than 100*.

Hence, under these illustrative assumptions, *more than 110* is argumentatively stronger than *more than 100* by the purely semantic measure in Eq. (17), but argumentatively weaker than *more than 100* by the pragmatic measure in Eq. (25). A rational hearer in a world where these assumptions held should take either utterance as positive evidence for the goal *S*, but if they are sensitive to pragmatic considerations they should interpret *more than 100* as appreciably stronger evidence than the (very weak) *more than 110*.

Argumentative Strength for Complex Cases

In practice, we can think of complex quantitative data as inviting the speaker who summarises it to choose among a wide range of semantically true options, and even if we restrict the speaker to utterances that do not invite false pragmatic inferences, there may still be many possibilities in play. A striking example is provided by 26), which appeared as a newspaper sub-headline in 2018 on the subject of Oxford University's undergraduate admissions.

- 26) Figures show one in four of [sic] colleges failed to admit a single black British student each year between 2015 and 2017 (<https://www.theguardian.com/education/2018/may/23/oxford-faces-anger-over-failure-to-improve-diversity-among-students>, retrieved 25/03/20)

From the context (provided by the main headline) it is clear that the speaker's communicative goal here is to make the point that Oxford is failing in racial equality, as regards British students, through its admissions policy. The factual claim offered in the headline in support of this point clearly satisfies the criterion of having positive argumentative strength, by the definition in Eq. (17). Moreover, although 26) does invite potential pragmatic inferences that weaken this effect (for instance, that three in four colleges succeeded in fulfilling this admissions criterion), it seems very likely that 26) also has positive argumentative strength by the pragmatic definition suggested in Eq. (25).

At the same time, the utterance makes a strikingly complex quantitative claim, and it does so in a way that gives rise to several ambiguities, raising a number of potential questions in the mind of the hearer. Should the statement be interpreted as referring to the same colleges each year? Why are the years 2015–2017 focused on? Does *one in four (of) colleges* mean “a quarter of the colleges of the university” or “one out of the four colleges studied”? And is the scope ambiguity of *(they) failed [to do this] each year* to be resolved as meaning “each year, they failed to do this” or “in at least one year, they failed to do this”?⁶

We stress that, in discussing this and other examples, we do not aim to take a position on whether the speaker's argumentative goal in each specific case is ultimately supported by the data that the speaker summarises. Rather, we wish to consider how a rational hearer should adjust their belief about the speaker's argumentative goal, given the statement that the speaker chose to make on this occasion.

In the case of 26), it appears clear from the context that the speaker has a specific communicative goal in mind, and it would be reasonable to expect the speaker to choose an utterance which constitutes a good argument for that goal, when summarising the large and complex dataset under discussion. We take this to be a fairly standard argumentative context, distinguished only by the complexity of the utterance in 26), a complexity which suggests that the speaker is willing to entertain a wide variety of possible utterances with which to summarise their data. In effect, a rational hearer is entitled to note that such circumstances naturally seem to call for post hoc descriptions that involve some cherry-picking of the data. However, if a hearer believes that this kind of cherry-picking is occurring, this should make a difference to the interpretation that they place on the data that is ultimately reported, much like it does to our interpretation of post hoc statistical tests. We discuss the implications of this in the following subsection.

Rational Interpretation in an Argumentative Context

So far, we have only considered the perspective of an argumentative speaker who assumes that the hearer either

⁶The text of the full article suggests that the answers to these questions are: 26) does refer to the same colleges each year; the scope of the study being reported on was just the years 2015–2017; *one in four* means “a quarter of the colleges”; and *failed . . . each year* in fact means “in at least one year, they failed to do this”. According to the article, “[t]he worst figures belonged to Corpus Christi College, which admitted a single black British student in those three years”.

interprets utterances semantically (*Argumentative Strength for a Semantic Interpretation of an Utterance*) or pragmatically (*Argumentative Strength for a Pragmatic Interpretation of an Utterance*) in the usual non-argumentative manner. This is a simplifying assumption but arguably legitimate if the speaker can expect the hearer to be unaware or unsuspecting of a possibly misleading framing intention. However, we should also consider the perspective of a suspecting rational interpreter who is quite aware of the speaker's framing intentions.

So how should a rational hearer interpret an utterance made by an argumentative speaker? If the speaker merely produced a semantically truthful utterance that was drawn at random from the whole set of semantically truthful possibilities, it would appear rational for the hearer to increase their belief in G if the utterance had positive argumentative strength according to the definition in Eq. (17). If the speaker produced a pragmatically felicitous truthful utterance that was drawn at random from the whole set of pragmatically felicitous truthful possibilities, it would appear rational for the hearer to increase their belief in G if the utterance had positive argumentative strength according to the definition in Eq. (25). However, it would not be reasonable to suppose that an argumentative speaker should act in this way: we expect them to produce a true and felicitous statement which is selected to serve their argumentative goals. Consequently, the behaviour of a rational hearer should also be more nuanced.

If we consider the set of pragmatically felicitous and truthful utterances by which a complex data set can be summarised, post hoc, these will vary considerably in their argumentative strength. Indeed, for complex data, we might reasonably expect these utterances to range from having negative to positive argumentative strength, by either of the measures proposed above. An optimally argumentative speaker, according to such a metric, would be one who selected the utterance with the greatest positive argumentative strength with respect to their communicative goal G.

One way of characterising a rational hearer's expectation in such a case would be to assume that the speaker is optimally argumentative, taking pragmatic inference into account, and hence selects the maximally argumentatively positive utterance (of those that are true and pragmatically felicitous) according to the definition in Eq. (25)⁷. But the rational hearer should then not take this at face value: they should be aware that an utterance selected at random from the set of possible utterances would likely have had much less positive argumentative strength than the one that was in fact uttered.

In fact, if the speaker is argumentatively effective, the rational hearer should be interested in how likely G is under the assumption that *u* is the best thing that could be said in support of G (rather than just 'a thing that could be felicitously said in support of G'). From this perspective, when the hearer determines whether to concur with the speaker's argumentative goal G on the basis of 26), the hearer should not merely be asking whether the data presented in 26) are more

compatible with a world in which Oxford's admissions policy is racist or one in which it is not. Rather, they should ask whether 26) exceeds in argumentative strength the most damning thing that could likely be asserted of Oxford's admissions policy in a world where it is not racist, and they should increase the strength of their belief in G only if that criterion is satisfied.⁸

To put it another way, if a rational hearer is aware that the speaker is trying to argue for G in an optimal way, and if *u* could likely be truthfully and felicitously asserted in a world where G was not the case (and the data under discussion reflected that G was not the case), the rational hearer should not take *u* as evidence in favour of G. Rather, as a criterion for increasing their belief in G, the rational hearer should adhere to a more stringent rule of interpretation, along the lines of 27).

- 27) Increase your belief in G on the basis of utterance *u* iff $\text{prag_arg_str}(u, G) > \text{prag_arg_str}(v, G)$ for all *v* that are likely to be true and assertable given $\neg G$.

The point we wish to emphasise here is that, given a large dataset from a world in which G does not hold, it may well still be possible to summarise that dataset in a way that has positive argumentative strength with respect to G, according to the measures proposed in Eqs (17 and 25)—searching through the set of pragmatically assertable propositions that are true in the not-G world, we can find some that are (perhaps highly) suggestive of the truth of G. Given a large dataset from a world in which G does hold, an argumentatively effective speaker should be able to do better than this—they should be able to find pragmatically assertable propositions that constitute stronger evidence for G than any of those which would be available in a non-G world.

In practice, we cannot guarantee that this will be the case, because data from a not-G world may by chance be suggestive of the truth of G, just as data from a G world may by chance be suggestive of its falsity—hence the use of *likely* in 27) and the above argument. If, by chance, although G is in fact true, the data do not indicate it, then 27) predicts that no statement can be made about those data which should induce a sceptical rational hearer to increase the degree of their belief in G: we take this to be a reasonable corollary⁹.

In practice, this approach appears to invite the hearer to be more sceptical than is warranted. For complex data, it is unlikely to be computationally tractable for the speaker to be able to find the argumentatively optimal utterance given their communicative

⁸Here we are assuming that the hearer is knowledgeable about which propositions are true in a world in which G is false. If the speaker takes the hearer to be less than perfectly knowledgeable, the picture becomes more complicated. We discuss this further in *General Discussion*.

⁹A sceptical hearer might, of course, take it that even data that is extremely favourable for G might have arisen in a non-G world, just as, in the context of experimental science, even data that admit a very small *p*-value might have arisen under the null hypothesis. Consequently, they might hold that the condition in 27) is never satisfied, because any *u* might be true and assertable in a non-G world. However, beyond a point, scepticism of this kind will not be rational, in terms of leading to a correct understanding of the likely world state. Here we do not attempt to characterise the optimal degree of scepticism for the rational hearer under this idealisation.

⁷Note that here we do not assume that the argumentative speaker is calibrating their choice of utterance to take into account the hearer's scepticism—although it is reasonable to think that an argumentative speaker may wish to do so. For ease of exposition we shall not attempt to address this case in this paper.

goals. Allowing for this, an appropriate rule of interpretation for a rational speaker might instead be along the lines of 28).

- 28) Increase your belief in G iff $\text{prag_arg_str}(u, G) > \text{prag_arg_str}(v, G)$ for all v that are likely assertable and accessible to the speaker given $\neg G$.

That is to say, the hearer should interpret an utterance as evidence for G if it has greater argumentative strength than any utterance that the speaker would, in practice, be able to produce in a world in which G was not true.

Interim Summary

The use of number in summarising data is associated with objectivity and precision, but these concepts are somewhat negotiable: number interpretation is pragmatically ambiguous in a number of ways, and the flexibility of numerical quantification makes it a particularly powerful domain in which a speaker can use language in the service of particular communicative goals that may not be shared by the hearer. If a speaker is argumentative in this sense, a rational hearer should strive to take this into account when determining whether to increase or decrease their belief in the proposition for which the speaker is ultimately arguing, based on the utterance(s) put forward in support of that proposition.

In the following section we exemplify some of these ideas with respect to a complex quantitative data set that is argumentatively described by a large number of distinct stakeholders with similar communicative goals, namely the results of REF 2014. Specifically, we will identify predictions that can be made about speaker behaviour in this context under the assumptions of the argumentative account, and examine the extent to which these are borne out.

A CASE STUDY: REPORTING THE RESEARCH EXCELLENCE FRAMEWORK

The approach outlined above allows us to make and test predictions about how speakers will use certain numerically quantified expressions in argumentative contexts. To do this, we wish to examine production data in a context in which speakers are summarising complex datasets with a clear argumentative goal in mind, and in order to evaluate the predictions we need to have access to the data as well as the speakers' productions. We would ideally be focusing on cases in which the speakers are expert users of argumentative language and are fully conversant with the details of the data they are summarising, as this is the scenario in which we expect speakers to produce argumentatively effective summaries of the data.

In all these respects, the public statements made by United Kingdom universities about their respective results in the REF 2014 assessment appear to constitute an appropriate object of study. In the following subsections, we briefly introduce the workings of the REF, consider the motivations and constraints that influence universities' public statements about the REF results, articulate a series of predictions about these

statements that follow from our theory, and evaluate these predictions against the data. We will show that there are clear indications that the argumentative considerations we discuss are indeed influencing speakers' production choices; however, these productions are nevertheless suboptimal, as anticipated in the foregoing discussion, and this poses interpretative challenges for the rational hearer.

The Nature of the Research Excellence Framework 2014

REF 2014 (Research Excellence Framework) was an exercise designed to assess the quality of research in United Kingdom Higher Education Institutions. Its stated aims were to inform the allocation of research grant funds; to provide accountability for public investment in research; and to "provide benchmarking information and establish reputational yardsticks, for use within the higher education (HE) sector and for public information" (<https://www.ref.ac.uk/2014/about/>, retrieved 04/04/20).

For REF 2014, institutions made submissions consisting of research outputs, case studies of impact derived from research, and information about the research environment. These submissions were evaluated by 36 appointed sub-panels and awarded one of five possible grades, ranging from 4* to U/C (unclassified). In the case of research outputs, these grades corresponded to quality that was "world-leading", "internationally excellent", "recognised internationally", "recognised nationally", and which "falls below the standard of nationally recognised work" respectively (<https://www.ref.ac.uk/2014/panels/assessmentcriteriaandleveldefinitions/>, retrieved 04/04/20).

Institutions typically submitted to multiple sub-panels, and these distinct submissions were evaluated separately. In all, REF 2014 evaluated 1911 submissions from 154 different institutions: these submissions comprised 191,150 research outputs and 6975 impact case studies (and represented work by 52,061 academic staff).

The overall quality profile for each submission comprised a weighted average of the grades for outputs (65%), impact (20%) and environment (15%). Across all submissions, 30% were graded 4*, 46% 3*, 20% 2*, 3% 1* and 1% unclassified. However, this varied appreciably across the three sub-profile measures: only 22% of outputs achieved the 4* rating, whereas 44% of impact submissions and 45% of environment submissions did so.

When the REF 2014 results were published (December 18, 2014), several media outlets compiled 'league tables', perhaps the most influential being Times Higher Education (THE), who provided three rankings:

- **Grade point average (GPA).** 4 points were awarded for 4* grades, 3 points for 3*, and so on. The overall GPA measure for an institution was the weighted mean of the GPA for its individual panel submissions (weighted by the number of full-time equivalent (FTE) staff whose work was submitted to each panel).
- **Research power.** This was computed by multiplying the GPA by the number of FTE staff submitted by the institution.

- **Research intensity.** This was computed by multiplying the GPA by the proportion of REF-eligible staff whose work was submitted by the institution. The ranking based on this was published subsequently to the other two rankings.

The THE main league tables included only multi-subject institutions (those which submitted to more than one panel), with single-subject institutions listed separately; we focus on multi-subject institutions in what follows.

To exemplify the methodology, consider the results from the Institute for Cancer Research (ranked first on GPA), which submitted to two sub-panels, namely Clinical Medicine and Biological Sciences. Its submission for Clinical Medicine comprised 69 FTE staff and achieved a GPA of 3.33 (which itself was comprised of scores of 3.09 for outputs, 3.90 for impact and 3.63 for environment), while that for Biological Sciences comprised 34 FTE staff and achieved a GPA of 3.55 (3.44 outputs, 3.80 impact, 3.75 environment). The overall weighted mean GPA was 3.40, which, multiplied by 103 FTE staff, yielded a power score of 351. The Institute for Cancer Research had 108 FTE REF-eligible staff, so its research intensity measure was calculated by multiplying its overall GPA by 103/108: the resulting intensity-weighted GPA was 3.25, on which measure it again ranked first.

Additional statistics were computed by Research Fortnight (RF) and published by the Guardian: these prioritised research power, but added one further measure:

- **Research quality.** This was calculated as the proportion of 4* research plus one-third of the proportion of 3* research, based on the overall quality profile¹⁰. As an example, the Institute for Cancer Research achieved 50% 4* and 41.7% 3* outputs, and hence a quality index score of 63.9 (= 50 + (41.7/3)).

The average GPA scores for the whole REF were 3.01 for outputs, 3.24 for impact, and 3.28 for environment. This represented an appreciable increase in scores from the previous assessment, the 2008 Research Assessment Exercise (RAE). Although the official REF results did not report GPA, they noted an increase in the percentage of outputs judged world-leading (22% against 14%) and internationally excellent (50% against 37%). The official summary further noted that “three-quarters of the universities had at least 10% of their submitted work graded as world-leading (4*). The top quarter had at least 30% graded as world-leading (4*)” (REF Brief Guide 2014, <https://www.ref.ac.uk/2014/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf>, retrieved 30/01/21).

Reporting the Research Excellence Framework

Many institutions issued press releases summarising their results, in keeping with the REF's stated goal to “establish reputational

yardsticks”. However, the REF team did not articulate an official line as to how the results should be interpreted as evidence of reputational strength. Consequently, institutions were largely free to interpret and present the results as they saw fit. This therefore represents a case in which expert communicators (the institutional press officers), with full access to a complex dataset, have the opportunity to select what information to present and how to present it, in the service of a clearly motivated argumentative agenda (advancing the perceived research reputation of their institution).

Against this, of course, it might be argued that—again in the absence of national policy as to what should be considered *prima facie* evidence of reputational strength—institutions were free to pursue different objectives, and their reportage of the results might merely reflect that. For example, if an institution had pursued a strategy of boosting research power at the expense of GPA, and this was successful, it would be reasonable for them to present research power data as evidence of their success. Thus, we cannot exclude an optimistic interpretation under which the selective reporting of results actually corresponds to the prior goals of the institutions. Even so, such reporting could mislead the (non-sceptical) hearer, who might interpret a press release focusing only on one metric as evidence that the institution in question could—if challenged—offer similarly strong evidence of its high reputation across a broader range of metrics, whereas this might in fact not be the case.

Hypotheses

Our overarching question is whether institutions use argumentatively effective strategies in the way our theoretical account predicts, when selectively reporting REF outcome data. From the rational hearer's point of view, the corresponding question is whether it is necessary to take the institutions' likely argumentative agenda into account when interpreting the data that they present. Here we aim to unpack this into specific testable predictions concerning how speakers will act under the assumption that they are argumentatively effective, judged by the standard that we proposed in *Quantifying Argumentative Strength, and Allowing for Uncooperativity*. That is to say, we aim to test whether the speakers in this study—the authors of the institutional reports about their REF results—are optimising the argumentative strength of their utterances.

Firstly, we expect argumentatively effective speakers to avoid presenting information that gives rise to inferences that run counter to their communicative goals. One potential source of such information is quantity implicature. We discussed how numerical expressions of the form *top M* might give rise to implicatures of this kind: not only do they convey that *top O* is not the case for salient $O < M$, but, particularly in the case of non-round *M*, they potentially convey that *top M-1* is not the case. Consequently, we expect argumentatively effective speakers to use *top M* formulations only when they can do so while avoiding argumentatively disadvantageous quantity implicatures.

Secondly, we expect argumentatively effective speakers to avoid presenting contextual information when doing so would promote inferences that run counter to their communicative

¹⁰A motivation for the use of this measure was the expectation that funding allocations would be based on the proportions of 4* and 3* research, with 4* weighted three times as heavily as 3* research.

goals. In the REF context, multiple rankings are available for discussion, most notably the GPA and power rankings, and this is evident to the speaker but not necessarily evident to the hearer. A rational hearer, aware of the existence of multiple rankings, might expect the speaker to quote the most favourable one and could infer that other unmentioned rankings were less favourable to that institution. We might therefore expect an argumentatively effective speaker to avoid indicating to the hearer that multiple rankings exist, in order to preserve the hearer's ignorance on this point and thus prevent the hearer from drawing an unfavourable inference.

Thirdly, we expect argumentatively effective speakers to avoid presenting information that fails to support their communicative goal more clearly than it supports the negation of that goal. In the context of the REF, we assume that the press releases issued are intended to bolster the reputation of the institution in question with reference to its competitors. Presenting statements in support of the institution's quality that would also be true of its competitors would therefore be an ineffective strategy in terms of argumentative force. Moreover, in a sceptical hearer (of the kind discussed in *Rational Interpretation in an Argumentative Context*), it would invite the inference that nothing more favourable could be said about the institution in question than that which could be said of its competitors. Thus, such statements would be ineffective (given a rational hearer unaware of the speaker's argumentative agenda) or actively counterproductive (given a sceptical hearer who takes the speaker's argumentative agenda into account), when considered as arguments for the institution's quality. Note that we assume, in making this prediction, that the speaker takes the hearer to be knowledgeable as regards what could be truthfully said of the institution's competitors: we return to the implications of this assumption in *General Discussion*.

Hence, in summary, we make the following predictions about the reporting of REF results:

H1: Speakers will use argumentatively appropriate reference points: an institution will be described as "top M " only if its ranking is near M , and speakers will avoid using non-round M .

H2: Speakers will prioritise favourable rankings and suppress unfavourable rankings: if the GPA and power rankings differ in how highly they place an institution, the more favourable ranking will be reported and the report will not convey the existence of an alternative ranking scheme.

H3: Speakers will avoid argumentatively unhelpful statements: they should not attempt to argue for the reputational strength of their institution on the basis of statements that would also be true of lower-ranked institutions.

Procedure

We collated data from the top 40 institutions, according to the GPA rankings, focusing in each case on descriptions of institution-wide accomplishments rather than those of individual faculties or departments. We first searched for press releases that had been issued at an institutional level on December 18, 2014 in connection with REF 2014 results, as archived on institutions' websites: these were available for 29 of the 40

institutions. Where these were not available we looked for summary pages detailing REF 2014 results as part of the institutions' general profiles: these were available for 10 of the remaining 11 institutions. In this way we obtained information from all institutions in the top 40 except the London School of Economics and Political Science (ranked third by GPA), which is hence excluded from the following analysis.

Results

H1: Use Best Available Reference Points

We predicted that expressions such as *top M* , used argumentatively, will be uttered only in connection with institutions that are ranked just above the relevant threshold, and only with round n , in order to avoid argumentatively unfavourable implicatures.

29)–38) represent all the uses of *top M* in the REF reports we examined that make reference to the overall institutional ranking. We indicate in square brackets the precise ranking that these quotes allude to.

- 29) Cardiff in top five for research excellence . . . The quality and impact of Cardiff's research has led to a meteoric rise in league tables, pushing it into the UK's top 5 universities [5th]
- 30) [King's College London is] Top 10 nationally for research 'power' and 'quality' [6th, 7th].
- 31) Warwick repeats top 10 success in UK research ranking exercise. [8th equal]
- 32) The [London School of Hygiene and Tropical Medicine] is ranked in the top 10 of all universities in the UK. [10th]
- 33) The results demonstrate excellence across research, putting Sheffield in the top 10 per cent of all UK universities. [16th equal = 10th percentile¹¹].
- 34) University of Leeds in top 10 for research and impact power [10th]
- 35) Royal Holloway is within the top 25 per cent of UK universities for research rated 'world-leading' or 'internationally excellent' [26th equal on unweighted measure = 21st percentile]
- 36) Swansea research breaks into UK top 30. [26th equal]
- 37) Essex has re-confirmed its position as one of the UK's top 20 research universities. [20th]
- 38) [Strathclyde is] Top 20 in the UK for Research Intensity. [18th]

As predicted, each of these descriptions uses round values of M in the *top M* formulation, and in each case no comparably salient $O < M$ exists for which the *top O* claim would be true. Hence we can see these examples as demonstrating a preference on the part

¹¹ Among multi-subject institutions, Sheffield ranks equal 14th out of 128 on the GPA measure; including single-subject institutions, it ranks equal 16th out of 154, hence on the cusp of the top decile. We assume this is the metric that the authors of the press release have in mind.

of the speaker to choose *top M* descriptions that are argumentatively effective, by the measures we discuss.

There are also indications in these data that the possibility of describing the institution as *top M* for some relatively small round value of *M* has motivated the choice of ranking criteria. Essex, in 37), and Strathclyde, in 38), both appeal to the research intensity measure, on which they are ranked considerably higher than on either of the measures initially published. Strikingly, Essex places 22nd on this measure, but 20th among universities—that is to say 37) is true if we do not consider the Institute for Cancer Research or the London School of Hygiene and Tropical Medicine to be universities (notwithstanding 32)). Similarly, Cardiff places 6th on research excellence as measured by GPA, but improves to 5th if we exclude the Institute for Cancer Research from consideration. Thus, their rhetorical move of focusing on universities may be motivated by the argumentative advantage of being able to make the *top five* claim rather than merely *top six*, which is semantically weaker but also gives rise to an argumentatively disadvantageous implicature *exactly 6th*.

Other uses of *top M* in these data involve generalisations over faculties or subject areas, and are sometimes combined with appeal to non-obvious ranking choices, as for example in 39) and (perhaps most extremely) 40). However, as we are restricting our attention here to descriptions of the institutions as a whole, we will not discuss these cases further, other than to note that they represent an alternative way to present the data for argumentative effect.

- 39) On actual research outputs 19 of Warwick's departments were ranked in the top 10 in the UK.
- 40) More than 25 per cent of the Durham University subjects entered for REF 2014 were in the top 5 subjects [sic] nationally for grade point average (overall score).

H2: Prioritise Favourable Rankings, Suppress Unfavourable Rankings

Taking the GPA and power rankings to be the most salient, we hypothesise that institutions will prefer to report the measure on which they rank more highly, as this constitutes better evidence of their high reputation. We also hypothesise that institutions will decline to mention the existence of the alternative measure, as this would invite inferences about their relative performance on that measure that would be detrimental to their reputational claim.

Of the 39 institutions for which we have data, 19 are ranked higher on GPA than power and 19 are ranked higher on power than GPA (the University of Durham places 20th on both rankings). Of the former group, nine mention GPA in their report and none mention power (a significant difference: $p < 0.01$, sign test), while ten do not make explicit reference to either measure. Of the latter group, 11 prioritise reference to power over GPA (eight of which do not mention the GPA measure at all) and two prioritise reference to GPA and do not mention power (again a significant difference: $p < 0.05$, sign test), while six do not make explicit reference to either measure. There is thus a significant

interaction ($p < 0.001$, Fisher's exact test), showing a clear preference for institutions to prefer the measure on which they rank higher and most commonly not to acknowledge the existence of the less favourable measure.

Outside of these two major statistics, the most popular measure for first mention was the combined proportion of research attaining a particular quality threshold, which was cited first by 14 institutions. 11 institutions focused on their proportion of 4* and 3* research: of these 11, 8 rank more highly on this measure than on either GPA or power. However, although this is compatible with a view in which the choice of this measure has been generally motivated by the wish to report a high ranking, in fact only two of these institutions comment on their rankings by this measure: Royal Holloway, in 40), which makes a claim that it could also make with reference to the GPA measure, and Queen Mary University of London, in 41), although the data from the summary table appears to place it 8th on this measure.

- 41) Royal Holloway is within the top 25 per cent of UK universities for research rated 'world-leading' or 'internationally excellent'.
- 42) Overall QMUL is ranked 5th in the UK [among multi-faculty institutions] for the percentage of its 3* and 4* research outputs.

Alongside the reference to the combined proportion of 4* and 3* research, two of these 11 institutions also make reference to GPA, one to power, and eight to neither. Thus, the general pattern is once again one in which institutions do not acknowledge the existence of alternative rankings which would describe them less favourably.

As we discussed earlier, the extent to which institutions acknowledge alternative measures could reasonably be expected to bear heavily on hearers' interpretations of the information provided. 43), from King's College London, represents a particularly transparent presentation of the alternative measures (the 'quality' measure here referring to GPA): the institution's preferred measure is complemented immediately by reference to the salient alternative. 44), from the University of East Anglia (UEA), is somewhat more opaque in this respect: the institution's preferred measure (focusing wholly on outputs, rather than the combined measure) is not one that is usually tabulated in its own right, and neither the overall GPA nor the power rating are alluded to in the following text. The hearer of 44) might reasonably be surprised to find UEA ranked 23rd by the THE for research quality.

- 43) King's has risen to 6th position nationally in the 'power' ranking—up from 11th in the Research Assessment Exercise (RAE) 2008. 'Power' takes into account both the quality and the quantity of research activity. King's has also risen to 7th position for quality—up from 22nd in 2008.
- 44) UEA is 10th in the UK for quality of research outputs. Over 82% of UEA research is rated as 'world-leading' or 'internationally excellent'.

H3: Avoid Argumentatively Unhelpful Statements

Our third prediction was that speakers would tend to avoid arguing for their institutions' reputational strength on the basis of statements that could also be truthfully asserted of lower-ranked institutions, on the basis that such statements would be argumentatively at best ineffective and at worst (given a sceptical hearer) counterproductive. However, there are a striking number of apparent counterexamples to this among the data, as exemplified by 45)–50), which include several article headlines.

- 45) The REF 2014 showed that the vast majority of Newcastle University's research was placed in the top two categories of 4*(world leading) or 3* (internationally excellent).
- 46) The University of Nottingham is a leading international institution carrying out world-class research, according to the Research Excellence Framework (REF) 2014.
- 47) [The University of Exeter has] world-leading research in all the units we submitted to REF. . .
- 48) Liverpool research ranked in UK top 10.
- 49) REF highlights world leading research at Aston.
- 50) Research at the University of Dundee has been ranked among the very best in the United Kingdom.
- 51) University of Sussex research is 'world-leading', major review finds.

Of these examples, 45) makes a quantitative claim, but the strength of this depends on the interpretation of *the vast majority*. The relevant figure in this case is 79%, which places Newcastle 34th on this metric. Were the claim merely *a majority*, Newcastle would share this distinction with all the top 88 institutions in the GPA ranking; if we interpret the threshold for *vast majority* at, for instance, 66%, then 63 institutions still meet this criterion. We note that the rest of the press release does not encourage the reader to contextualise the claim in this way, and does not present any information that would be helpful to them in doing so.

The headline 48) also makes a quantitative claim, but this turns out, on closer inspection, to be existential in character: the body text clarifies that seven subjects at Liverpool were ranked in the top 10 nationally (by the measure of "research excellence"). As there are 36 sub-panels in play, and given the possibility of appealing to multiple distinct measures, the claim of having "research ranked in UK top 10" is argumentatively a relatively weak one, although it is impossible to verify precisely how weak without detailed examination of the overall distribution of outcomes by sub-panel.

The subsequent examples here all focus on the existence of world-leading research at the respective institutions. In the context of the REF results, this is a surprisingly weak claim from an argumentative perspective. As noted earlier, three-quarters of the universities submitting to REF 2014 had at least 10% of their work graded as 4*. Indeed, only 72 of the 1911 submissions failed to have any work at all graded at 4*, so the claim made by Exeter as 47) is one that could be made by the majority of institutions submitting to REF, while the existential

claims of 49)–51) could be made by 151 of the 154 institutions. Thus, to the extent that these claims are to be understood as arguments to the effect that Aston, Dundee and Sussex are above-average institutions (which they are, according to the GPA measure), they appear to have very little argumentative strength, according to the measures proposed in this paper.

GENERAL DISCUSSION

Individual institutions' reporting of the results of REF 2014 represents a scenario in which speakers can be expected to summarise complex data in an argumentatively effective way, in the service of a generally clear communicative goal, namely to emphasise the high quality of the institution's research. Based on the approach to argumentation discussed in this paper, we were able to articulate three predictions as to how speakers would behave in this case. Two of these were borne out. Given a choice of rankings to report, institutions have broadly behaved in accordance with a strategy of selecting the ranking that is most favourable, and presenting little information to hint at the existence of other, less favourable, data. This would accord with a strategy of presenting argumentatively strong information while dissuading the hearer from drawing ad hoc inferences that undermine its argumentative point. The use of the formulation *top M* also adheres to the predicted principles: the formulation is used only when the precise ranking is close to *M* and *M* is a round number. Again, the effect is not to invite the hearer to draw inferences that would be deleterious to the argument being advanced.

Speaker behaviour in this case study, however, deviated strikingly from our third prediction: argumentatively weak information was frequently presented, as seen in 45)–51), where assertions are made that could equally truthfully be made of institutions which had performed much less well. This represents a challenge for the explanatory utility of the approach we suggest—how can we explain this choice of communicative strategy?

Recall that in *Quantifying Argumentative Strength, and Allowing for Uncooperativity* we raised the question of how sceptical a rational hearer should be about the use of simple descriptions of complex data, when evaluating the argumentative strength of these descriptions and using that to update beliefs. A minimally sceptical approach would be to increase one's belief in some proposition *G* given an utterance *u* if the probability that *u* is true given *G* exceeds the probability that *u* is true given the negation of *G* (and to decrease one's belief in *G* if the reverse is true). For example, if we consider 47) as *u* and assume *G* is the proposition that Exeter is an outstanding research university, this condition is clearly satisfied, and we should increase our belief in *G* on hearing 47). A maximally sceptical approach would be to increase one's belief in *G* given *u* only if *u* is argumentatively better than any of the things that could be said given that *G* were false (and to decrease one's belief in *G* otherwise). In this case, taking the same values of *u* and *G* as before, this condition is not satisfied: 47) would likely be true even if Exeter were not an outstanding research university (and

its REF results reflected that), so we should decrease our belief in G given 47).

In practice, 47) illustrates an intermediate case: it represents a relatively weak argumentative claim, but this could be for distinct reasons. One possibility is that the speaker of 47) thinks that the hearer will not be sceptical in the way suggested by the above account in how they update their beliefs, and therefore expects this argument to convey positive argumentative strength: we could think of this as the speaker being optimistic about the receptiveness of the audience to their argument. Another possibility is that the speaker has simply not considered that 47) is an objectively weak argument, given that it is something that tens of institutions ranked below Exeter on the standard metrics could also say¹². In this case, we could regard the speaker as being incompetent at maximising argumentative strength—and, to the extent that speakers behave this way, we could conclude that the model is inadequate for capturing speaker behaviour.

It is also worth considering a third possibility. Perhaps the speaker of 47) thinks that the hearer is not aware that this assertion would also be true for lower-ranked institutions, and consequently believes that the hearer will perceive the utterance to have positive argumentative strength, even if the speaker knows this not to be the case. This is somewhat analogous to the case of Hypothesis 2, in which the speaker exploits the hearer's ignorance about alternative measures: it is reasonable to expect the hearer to be less than fully informed about the REF results for competing institutions, and this would license the speaker to exploit the argumentative potential of utterances that would not be predicted to be argumentatively effective with fully knowledgeable hearers¹³. In general, we feel that this is a plausible explanation for argumentative speakers' divergence from the theoretically optimal strategy. However, in order to evaluate this explanation empirically, we would need to establish the hearers' knowledgeability (and specifically how this is perceived by the speakers), which cannot be read off the data we examine in this paper.

In summary, then, the picture presented by the REF reports is (perhaps characteristically) mixed. The authors of these reports are, collectively, not entirely consistent in maximising argument strength, by the measures proposed in *Quantifying Argumentative Strength, and Allowing for Uncooperativity*. However, at the same time, they are clearly not neutral in their treatment of the data. Consequently, these

texts place considerable demands on the rational hearer who wishes to interpret the claims being made. Given a hearer who accepts their reports at face value, perhaps 20 or more institutions might be able to convince that hearer that they belong in the top 10; however, given a maximally sceptical hearer, perhaps only about 10 institutions might be able to convince that hearer that they belong in the top 20.

Thus, as far as these press releases are concerned, the hearer cannot arrive at any close approximation of an objectively accurate interpretation of the results by adopting any of the first three strategies canvassed earlier in this paper. Adopting the straightforward semantic or pragmatic approaches to argumentative strength, the hearer will generally infer that the universities' research has been evaluated more favourably by REF than is in fact the case. Adopting the more demanding stance of expecting the best possible descriptions, the hearer will overcompensate and infer that the evaluations are in fact worse, in most instances, than was actually the case. To decipher the descriptions accurately, from the standpoint of argumentative strength, the hearer has to be aware that the speakers are systematically making efforts in the direction of maximising argumentative strength, but also that they are inconsistent in how effectively they achieve this.

These data exemplify a much more widespread problem, concerning both how complex information should be summarised in order not to mislead the hearer, and how the hearer should interpret summary information in order to reconstruct the best possible approximation to the underlying reality. The problem is clearly accentuated when a speaker has a particular argumentative agenda, even when they are determined to advance that agenda only through the presentation of true and accurate (albeit carefully selected) facts. It is perhaps rather unfortunate, although not entirely surprising, that this challenge is so strongly in evidence in the context of the reporting of REF 2014 results, in which some of the United Kingdom's most esteemed institutions participate in an exercise designed to determine their "reputational strength".

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CC collected data; CC and MF analysed the data and co-wrote the paper.

FUNDING

The authors gratefully acknowledge financial support via the project "Probabilistic models of content choice in manipulative language use" funded by the Lower Saxony Ministry of Science and Culture (15-50115-Schottland-13/19).

¹²This is rather analogous to the problem of chance capitalisation in statistical analysis: the speaker may not have considered that 47) is likely to be true irrespective of whether or not G holds.

¹³Note that a cooperative speaker might equally be aware that the hearer's understanding of the situation is limited and that the hearer may miscalculate the argumentative strength of utterances as a result. For example, addressed to a non-specialised audience, *Daniel Kahneman has won the Nobel Prize* would constitute an effective argument as to that person's academic credentials, but *Tim Gowers has won the Fields Medal* might not, and the speaker might need to add more context to explain how the argumentative strength of this should be evaluated.

REFERENCES

- Anscombe, J.-C., and Ducrot, O. (1983). *L'argumentation dans la langue*. Brussels: Mardaga.
- Ariel, M. (2004). Most. *Language*. 80 (4), 658–706. doi:10.1353/lan.2004.0162
- Bryson, B. (1998). *Notes from a Big Country*. New York: Doubleday.
- Carston, R. (1998). “Informativeness, Relevance and Scalar Implicature,” in *Relevance Theory: Applications and Implications*. Editors R. Carston and S. Uchida (Amsterdam: Benjamins), 179–236. doi:10.1075/pbns.37.11car
- Cummins, C., Sauerland, U., and Solt, S. (2012). Granularity and Scalar Implicature in Numerical Expressions. *Linguist Philos.* 35, 135–169. doi:10.1007/s10988-012-9114-0
- de Jaegher, K., and van Rooij, R. (2014). Game-theoretic Pragmatics under Conflicting and Common Interests. *Erkenntnis* 79, 769–820. doi:10.1007/s10670-013-9465-0
- Franke, M., de Jager, T., and van Rooij, R. (2012). Relevance in Cooperation and Conflict. *J. Logic Comput.* 22 (1), 23–54. doi:10.1093/logcom/exp070
- Franke, M., Dulcinati, G., and Pouscoulous, N. (2020). Strategies of Deception: Under-Informativity, Uninformativity, and Lies-Misleading with Different Kinds of Implicature. *Top. Cogn. Sci.* 12, 583–607. doi:10.1111/tops.12456
- Good, I. J. (1950). *Probability and the Weighing of Evidence*. London: Griffin.
- Hesse, C., and Benz, A. (2020). Scalar Bounds and Expected Values of Comparatively Modified Numerals. *J. Mem. Lang.* 111, 104068. doi:10.1016/j.jml.2019.104068
- Horn, L. R. (1972). On the Semantic Properties of Logical Operators in English. Ph.D. thesis. Los Angeles: University of California Distributed by Indiana University Linguistics Club.
- Krifka, M. (2009). “Approximate Interpretations of Number Words: A Case for Strategic Communication,” in *Theory and Evidence in Semantics*. Editors E. Hinrichs and J. Nerbonne (Stanford, CA: CSLI Publications), 109–132.
- Lasersohn, P. (1999). Pragmatic Halos. *Language*. 75, 522–551. doi:10.2307/417059
- Meibauer, J. (2014). *Lying at the Semantics-Pragmatics Interface*. Berlin: Mouton de Gruyter.
- Merin, A. (1999). “Information, Relevance, and Social Decision-Making: Some Principles and Results of Decision-Theoretic Semantics,” in *Logic, Language, and Computation*. Editors L. S. Moss, J. Ginzburg, and M. de Rijke (Stanford, CA: CSLI Publications), Vol. 2, 179–221.
- Pinker, S., Nowak, M. A., and Lee, J. J. (2008). The Logic of Indirect Speech. *Proc. Natl. Acad. Sci. U.S.A.* 105 (3), 883–888. doi:10.1073/pnas.0707192105
- Solt, S. (2014). “An Alternative Theory of Imprecision,” in *Proceedings of the 24th Semantics and Linguistic Theory Conference (SALT 24)*. Editors T. Snider, S. D’Antonio, and M. Weigand (Washington DC: Linguistics Society of America), 514–533.
- Sperber, D., and Wilson, D. (1986). *Relevance: Communication and Cognition* (2nd Edn 1995). , Harvard University Press, Cambridge, MA: Harvard University Press.
- Van der Henst, J.-B., Carles, L., and Sperber, D. (2002). Truthfulness and Relevance in Telling the Time. *Mind Lang.* 17, 457–466. doi:10.1111/1468-0017.00207
- Winterstein, G. (2012). What *But*-Sentences Argue for: An Argumentative Analysis of *but*. *Lingua* 122 (5), 1864–1885. doi:10.1016/j.lingua.2012.09.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cummins and Franke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.