Argument and Computation, 2014
Vol. 5, Nos. 2–3, 216–236, http://dx.doi.org/10.1080/19462166.2014.899270

Trust, relevance, and arguments

Fabio Paglieri* and Cristiano Castelfranchi

Goal-Oriented Agents Lab (GOAL), ISTC-CNR, Via S. Martino della Battaglia 44, Rome 00185, Italy

(*Received 30 August 2013; final version received 11 February 2014*)

This paper outlines an integrated approach to trust and relevance with respect to arguments: in particular, it is suggested that trust in relevance has a central role in argumentation. We first distinguish two types of argumentative relevance: internal relevance, i.e. the extent to which a premise has a bearing on its purported conclusion, and external relevance, i.e. a measure of how much a whole argument is pertinent to the matter under discussion, in the broader dialogical context where it is proposed. Then, we argue that judgements of internal relevance heavily rely on trust, and that such trust, although occasionally misplaced (e.g. in some so-called fallacies of relevance), is nonetheless based on several reasons, and thus often justified, by either epistemic or pragmatic considerations. We conclude by sketching potential methods to formally model trust in argumentative relevance, and briefly discussing the technological implications of this line of research.

Keywords: trust; relevance; formal model of argumentation; computational models of natural language argument

1. Introduction

In a recent paper (Paglieri & Castelfranchi 2012), we noted that trust and relevance, while being separately the focus of much research in philosophy (Baier, 1995; Cohen, 1994; Floridi, 2008; Gambetta, 1988; O'Neill, 2002), pragmatics (Borg, 2005; Grice, 1989; Sperber & Wilson, 1995; Wilson & Sperber, 2004), and computer science (Borlund, 2003; Bremer & Cohnitz, 2004; Castelfranchi & Falcone, 2010; Cooper, 1971; Crestani, Lalmas, Van Rijsbergen, & Campbell, 1998; Ramchurn, Huynh, & Jennings, 2004), have rarely been brought into contact with the study of information dynamics. Yet, we argued in that paper, their integration has much to offer. The same holds true for argumentation: relevance has often been studied in that context (for a comprehensive survey and treatment of the subject, see Walton, 2004), and trust is also often discussed, especially in relation to the evaluation of testimony (on that topic, see in particular Govier, 1997, 1998). However, the need for trusting (or not) the relevance of arguments has not been addressed, nor the rationality (or lack thereof) we manifest in doing so – with very few exceptions, such as Santibañez (2012). There seem to be a widespread tacit consensus that relevance is not a matter of trust, as far as arguments are concerned.

In this paper, we challenge that assumption, to show that, on the contrary, there is an inescapable element of trust built into our judgements of relevance with respect to arguments, and that such trust can have, and often has, rational foundations. This, in turn, has important consequences for the formal and computational treatment of argumentation, as well as some potential technological implications. The paper is organised as follows: in Section 2, we briefly discuss how relevance is to be understood in argumentation, and propose a distinction between two main types of argumentative relevance; then in Section 3 we ask whether these forms of relevance require trust, and

*Corresponding author. Email: fabio.paglieri@istc.cnr.it

of what kind, whereas Section 4 is devoted to understanding what reasons may justify our trust in argumentative relevance; Section 5 outlines the implications of this analysis for formal and computational models of argumentation, as well as potential technological applications; Section 6 concludes the paper, by summarising its main results and pointing to open-ended issues for future work. While throughout the paper we will often engage with the rich literature on trust and relevance, it is not our purpose to offer a comprehensive survey of either topic, and this paper should not be taken as such.

2. Two forms of relevance in argumentation

It is notoriously difficult to arrive at a comprehensive and satisfactory definition of relevance in information dynamics (on that problem, see Bremer & Cohnitz, 2004; Floridi, 2008): elsewhere (Paglieri & Castelfranchi, 2012) we have argued that such a definition will require making reference to the *goals* of the parties engaged in information exchange, and we also insisted that the connection between information and goals is made explicit by a belief-based model of goal processing, like the one detailed by Castelfranchi and Paglieri (2007). With respect to argumentation, the role of goals in defining relevance has long been recognised, thus there is no need to elaborate on it any further here. An early mention is found in the following passage by Sidgwick: ‘The more we understand that all thought is purposive the more we shall be able to discover the less obvious kinds of irrelevance in argument’ (1914, p. 149). More recently, Walton (2004) has explicitly tied the definition of argumentative relevance to the dialogical goals of the interaction in which the parties are engaged: in turn, such dialogical goals contribute to determine the type of dialogue, and thus the appropriateness of the arguers’ moves in it (as discussed in Walton, 1998). For instance, a deliberation dialogue aims at determining a course of action deemed acceptable by both parties, whereas in eristic confrontation (e.g. a quarrel) the goal is to vent one’s feelings, with little concern for any resolution in the underlying conflict of opinions. Due to this sharp difference in dialogical goals, what is relevant in one context may well not be so in the other.

It has also been often noted that relevance, in argumentation as well as in other forms of communication, is not a matter of topical overlapping (see, for instance, Walton, 2004, p. 10). This is easy to see. Imagine we are debating whether or not the moon has an impact on tides: then my claim that ‘Looking at a full moon over the sea at high tide is lovely’ is clearly irrelevant to the discussion, despite the significant thematic overlap. This has prompted some authors to elaborate a more specific notion of relevance. For instance, Walton proposes the following as a working definition of what he calls material relevance (as opposed to relevance in general):

> an argument (or other speech act) is *materially relevant* in a conversation if it bears directly or strongly on the issue so that it is worth prolonged or detailed consideration in relation to the specific problem or issue that the conversation is supposed to resolve. (2004, p. 14)

Once again, the reference to the dialogical goal of the interaction is central to the definition.

For the purpose of this paper, it is useful to distinguish two types of relevance that are important in argumentation: *external relevance* (or argument relevance) and *internal relevance* (or premise relevance). Let us first offer a working definition for these notions, and then see how they fit in the current literature on relevance in argumentation.

EXTERNAL RELEVANCE (or argument relevance): The relevance of an argument with respect to a certain topic of discussion and/or matter of concern in a prolonged dialogical exchange.

INTERNAL RELEVANCE (or premise relevance): The relevance of a premise to the conclusion it is meant to support or prove.

External relevance is akin to the sense in which relevance is defined with respect to information in general, as something that matters to present concerns. There is an important difference, though: while in most everyday transactions determining what matters requires some insights into the goals of the parties (Paglieri & Castelfranchi, 2012), in argumentation the cards are usually put on the table – that is, the topic and the commitments each party undertook in previous discussion are typically manifest and agreed upon, or at least open to public scrutiny. This provides an element of objectivity (or, more exactly, of intersubjectivity) to the evaluation of external relevance in argumentation. It also implies that arguments can be relevant to the discussion but unhelpful to one or more parties, and, conversely, they can be irrelevant to the dialogical interaction but extremely helpful for one or more parties – as is often the case with fallacious moves that fail to be detected by the counterpart.

While external relevance concerns the relationship between an argument and the dialogical context in which it is used, internal relevance refers to the proposed connection between the premises and the conclusion of an argument – hence the label ‘internal’, to emphasise that it is a property of the internal structure of an argument. Are the premises truly relevant to the conclusion they are supposed to support? This is the sense in which relevance is considered a key criterion for assessing the cogency of an argument (Govier, 1985; Vorobey, 2006), and the meaning of Blair’s (1992) notion of premissary relevance. It is also the notion of relevance implied by the characterisation of *ignoratio elenchi* given by Aristotle (1939) in his *Topics*:

> When the argument is a demonstration of something, but it is something irrelevant which has nothing to do with the conclusion, no inference will be drawn from it about the latter; if there appears to be such an inference, it will be a sophism not a demonstration. (*Topica* 162 a13–16)

A similar distinction is proposed by Jacobs and Jackson (1992), when they distinguish between the informational relevance of a premise to the truth of a conclusion (internal relevance) and the pragmatic relevance of argumentative acts to the overall aim of a dialogue (external relevance). An argument that lacks internal relevance, i.e. its premises do not really support its purported conclusion, cannot have external relevance, any more than a flawed brick can help in supporting the building where it is placed. But arguments can fail to be externally relevant, and yet retain internal relevance (Jacobs & Jackson, 1992). The following example, taken from Walton (2004, p. 13), illustrates the point:

> Mr Smith, who is a patient in an intensive care unit and has an unbeatable cancer of the pancreas, informs his physicians he wants to be taken off life-support systems. His wife disagrees and feels he should be kept in the intensive care unit. The physicians, staff, and family form a committee to discuss the problem, in order to try to arrive at a decision on how to proceed. Should Mr Smith be taken off the life-support systems or not? During the discussion, Dr Jones, one of the attending physicians, starts to discourse on the general economic questions posed by this type of patient, saying, ‘It costs, on average, over thirty thousand dollars to maintain this kind of patient in intensive care. We have a growing national debt in this country. Can we, as a nation, really afford to maintain this expensive kind of treatment, when the money could be more usefully spent on things like preventative medicine? I think we need to reconsider our priorities in health care spending.’ At this point in the meeting, another physician, Dr Brown replies, ‘Financial questions of this sort are not really all that relevant to our present discussion of the specific problem we face here. Let’s get back to the question of what we propose to do. Our problem is whether to keep Mr Smith in the unit, or to release him to a medical ward, according to Mr Smith’s own request.’

Here, Dr Jones’ argument is charged with being externally irrelevant, since, according to Dr Brown, whether or not current health care spending is efficient has no bearing on the fate of Mr Smith. Even if one agrees with that charge, this has nothing to do with the internal relevance of that argument: the costs of life-support systems for incurable patients, as opposed to alternative uses of those funds, are certainly pertinent to evaluate whether current health care spending makes sense (the immediate conclusion of the argument). Here, the premise offered by Dr Jones is relevant to the conclusion of his argument, even if that argument in turn is arguably immaterial to establishing the particular fate of Mr Smith.

The rationale of the distinction between internal and external relevance, if not the precise labels, is not novel in argumentation theory, and it even precedes the article by Jacobs and Jackson (1992) mentioned above. Similar categories can be found in the treatment of fallacies of relevance in logic textbooks from the first half of the twentieth century, such as Castell (1935) and Werkmeister (1948): these authors oppose the fallacy of irrelevant thesis (lack of external evidence) to the fallacy of irrelevant evidence (lack of internal relevance). Also, the distinction between red herring and the fallacy of irrelevant reason in Johnson and Blair (1983) closely mirrors the difference between external and internal relevance. Even more significantly, the same distinction is found, under different names, in the *Federal Rules of Evidence* (2011) of the USA. In particular, rule 401, 'Test of Relevant Evidence', reads as follows:

> Evidence is relevant if:
>
> (a) it has any tendency to make a fact more or less probable than it would be without the evidence and
> (b) the fact is of consequence in determining the action.

Interestingly, a slight paraphrase of the rule is sufficient to reveal its connection with the distinction between (a) internal and (b) external relevance:

> An argument is relevant if:
>
> (a) its premises have any tendency to make its conclusion more or less probable than it would be without them and
> (b) the conclusion is of consequence in advancing the debate.

Some authors have focused more on one or the other aspect of argumentative relevance. Govier's analysis of relevance (1985) in terms of probative weight is essentially concerned with internal relevance – in fact, it closely resembles clause (a) above, since it is a type of probative relevance (Walton, 1992; Woods, 2013). In contrast, the various relevance criticisms discussed by Krabbe (1992) mostly pertain to external relevance, and also Walton (2004) is definitely more focused on this type of relevance, when he repeatedly stresses the importance of the dialogical context in determining the relevance of an argument. As he notes:

> relevance is a dynamic concept that cannot be structured only by the model of an argument as a localized inference from a set of premises to a single conclusion. (…) Relevance is therefore a dialectical concept in that it relates to how a chain of argumentation is being directed toward some ultimate end point that is an issue of a discussion. (2004, p. 28)

While we do not disagree with Walton's dialectical model of (external) relevance, we think it is also important to account for the internal relevance of 'a localised inference', for reasons that will become apparent in the next section.

As for fallacies of relevance, in argumentation theory their treatment spans the divide between internal and external relevance, although it is not always easy to see on which side of that divide each individual fallacy lies. This largely depends on the fact that even the list of those fallacies is somewhat controversial (on this point, see Walton, 2004; Woods, 2013). The standard quartet includes all the '*ad* fallacies': *ad hominem* (personal attack), *ad populum* (appeal to popular opinion), *ad misericordiam* (appeal to pity), and *ad baculum* (use of threats). But the straw man, i.e. attacking a point that unfavourably misrepresents the counterpart's position, and the *ignoratio elenchi* or red herring are also often cited as fallacies of relevance, the latter being often used as a ragbag category, as Hamblin put it (1970, p. 31): whenever a fallacy cannot be classified otherwise, treat it as a red herring.[1] While nowadays these forms of reasoning are no longer considered as invariably fallacious, depending on the context, they still tend to be grouped under the heading 'fallacies of relevance' – so much so as to inspire Krabbe's (1992) metaphor of 'the sponge of relevance', to indicate the widespread habit of absorbing within this notion a variety of different phenomena.

While we do not aim to settle the issue here, it seems appropriate to further classify the four *ad* fallacies as problems of internal relevance: for instance, whenever the personal characteristics of an individual have no bearing on establishing whether s/he committed a certain action (e.g. murder) or possesses a specific quality (e.g. being a reliable witness), then those characteristics cannot be used as premises for concluding anything on that action or quality, and doing otherwise would be a case of *ad hominem* fallacy. Similarly for appeals to popular opinion, emotions, or the use of threats: whether or not they are used fallaciously hinges on whether or not they constitute relevant premises for the conclusion they are supposed to prove – that is, whether or not they have internal relevance for that argument. In contrast, straw-manning seems more a failure of external relevance: it refers to the practice of making an unhelpful contribution to the ongoing debate, precisely because the position of the counterpart is being misconstrued. Something similar can be said about the red herring (*ignoratio elenchi*), at least based on Walton's definition of that fallacy:

> The *ignoratio elenchi* fallacy is committed when the proponent's argument cannot be extrapolated forward by any plausible or convincing sequence of reasoning, so that it winds up at the conclusion to be proved by him in the discussion. (2004, p. 24)

Once again, the problem here is lack of external relevance: whatever the merits of the argument per se, it acts as a red herring whenever it has no bearing on the general issue under discussion. Seen in this light, Dr Jones' remark on health care costs, in the example discussed above, would have been charged with committing a red herring.

3. Trust in relevance: does it matter for argumentation?

While the centrality of relevance for the study of argumentation is well established, not much work has been devoted to discuss whether trust plays any significant role in assessing the relevance of arguments. In fact, the only exception we are aware of is a recent paper by Santibañez (2012), who proposes an evolutionary connection between relevance (in the sense of Sperber & Wilson, 1995) and trust, largely inspired by Tomasello's (2008) views on the evolution of language. Simply put, the idea is that the cooperative expectations on which human communication is grounded, including a presumption of relevance, could not have evolved in the absence of a certain degree of trust among members of the same community. While we do not find this evolutionary hypothesis implausible (on the contrary, it is consistent with the idea of a reciprocal altruism of knowledge proposed by Castelfranchi & Poggi, 1993), we believe the connection between trust and relevance to run deeper than that: in particular, trust in relevance is an essential ingredient of everyday communication right now, whether or not it played a role in the evolution of language. Appreciating that, however, requires an understanding of trust different from the one favoured by Santibañez, when he writes:

> The importance of the ability to trust, and the act of trusting, is that it enables an *energetic rest*; a rest from thinking and being aware that when an individual recognises another as a party for a collaborative exchange, the other is doing precisely the same. Trust is necessary for reciprocal altruism, an important source of our sociability. (2012, p. 490)

Although it is undeniable that a trusting attitude allows us to lower our defences[2] and thus entails the willingness to make us vulnerable towards others, there is a dangerous equivocation lurking behind these lines: the idea that trust is a form of epistemic laziness, or even epistemic desperation – a last resort when nothing else is available. This potential misconception is also present in the words of Tuomela and Tuomela (2005, p. 71) quoted by Santibañez: ‘When knowledge is lacking, trust is needed for cooperation’. If this is taken to mean that trust in *p* is required only when *p* is not self-evident, then we agree: you do not need to trust that the sun will come up tomorrow (except ironically), since you know that it will happen as a matter of absolute certainty. But if this is taken to imply that the act of trusting someone has no rational grounds, other than lack of better alternatives, then we emphatically disagree. This would reduce all forms of trust to blind trust, but such a reduction is not at all warranted: more often than not, the choice and the act of trusting others is based on reasons, both epistemic and practical – that is, trust is a critical attitude (so much so that it is enlisted among the cognitive tools we use to exert epistemic vigilance; see Sperber et al., 2010).

In fact, *trust is based on reasons* even when it does not seem so at first glance. Take the common occurrence of asking for directions from a complete stranger in the streets of a city you are visiting: you have no hard evidence of the competence, character, or motives of your informant, and you are unlikely to ever meet him/her again to sanction any misdirection s/he may offer – yet, you will often trust in his/her indications. Is that not an act of blind trust? Not at all: at the very least, your trust is based on the reasonable expectation that, lacking any motive to cause you harm or be less than cooperative, the informant in question (having already decided to *invest* some resources in his/her communicative act) will try to provide good information. This is demonstrated by the fact that your choice is not random: for instance, you will probably avoid bothering someone who is in an obvious hurry, or engaged in some other activity – that is, you will try to pick someone who is likely to be well disposed to answer your query. As for his/her competence, this is also factored in your choice, within the limits afforded by your current predicament: again, you will not select someone who is obviously a fellow tourist (e.g. someone taking pictures of the monuments), but rather someone who seem to be familiar with the surroundings (e.g. someone returning from a grocery shop). All of these are subtle indications of the reason-based nature of your act of trust, even when information on the object of that trust is very scant (another notable example being our frequent trust in Wikipedia entries, well analysed in Goodwin (2010)).

Turning back our attention to relevance in argumentation, it is worth asking whether trust, in the sense outlined above, is required in dealing with external relevance. In assessing whether an argument contributes or not to the matter under discussion, do we need to trust in the proponent of that argument? The answer to this question is in the negative, at least typically. In argumentation, external relevance is either self-evident, openly criticised, or mistakenly assumed, but in none of these cases does trust play a major role. Let us start from the most common situation, when external relevance is self-evident. The reasons behind this fact have been already hinted at: while arguing on a matter of joint interest, all interlocutors tend to know what they are talking about – not in the sense of being necessarily expert on the subject matter, but in the sense of having a clear and often complete view of what transpired so far and thus what are the public (manifest) positions of each party. This makes it relatively easy to directly assess whether or not a new argument is relevant to the matter under discussion, and immediately prompt criticism or demands for clarification if it is not obviously so. Thus, there is not much need for trust in relevance, nor an inclination to grant it, since the whole point of arguing about a certain topic is precisely to subject each position to careful scrutiny, instead of accepting entire arguments based on trust.

Notice that the point we are making is not that external relevance is uncontroversial in discussion (on the contrary, it can become the subject of heated debate), just that it is not typically a matter of trust. If the external relevance of an argument is in doubt, it will be challenged, and the challenge will become part of the discussion. This is the second instance we mentioned above. For instance, let us imagine, following Walton (2004, pp. 25–26), that it has been decided, at the early stage of a committee meeting, that the task at hand is to deliberate whether to take action A or B, and then at some point during the discussion a committee member questions this earlier decision, suggesting that the committee should also consider the possibility of refraining from either action for the time being. Is this motion to be ruled as irrelevant, or not? As Walton notes, this will depend on the circumstances of the case, but that is not our point. Whenever external relevance of arguments comes to be debated in dialogue, it still does not involve trust. The statement 'Trust me, this is relevant' is not really a suitable answer to a doubt of relevance: at most, it can be accepted provisionally, as a promise that the relevance of that particular argument will soon become apparent. This sometimes occurs in legal proceedings, whenever a judge, following an objection of relevance, allows one of the parties to proceed, with the admonition that they will have to 'get to the point' in short order. While similar cases underscore the provisional and procedural nature of relevance judgements (Walton, 2004), they do not constitute genuine instances of trust: in fact, the alleged relevance of the argument for the case is not accepted until it is openly demonstrated, and only a temporary suspension of judgement is granted.

Finally, it is undeniable that some arguments appear to be externally relevant and are accepted as such, even if they are not. Imagine, for instance, that the ethnic majority in a country is prompted by a particularly moving appeal to indignation to take action against a minority, even if there are no grounds to assume that indignation to be justified, and thus appealing to it should be irrelevant in deciding how to act towards that minority. This is a perfectly possible occurrence: after all, we have ample evidence of rhetorically successful fallacies of relevance. But once again their success does not involve trust: in similar cases, people are swayed by their own cognitive and emotional make up, rather than by any overarching trust in the speaker. They are being played, perhaps expertly, but the key to playing them is not trust.[3]

To sum up, external relevance does not typically involve trust. However, things are different with respect to internal relevance, that is, the relevance of premises for the conclusion they are intended to support. The mere fact of presenting something as a reason for something else presupposes, among other things, a claim on its relevance: when someone argues that '*p*, therefore *q*', the (internal) relevance of *p* to *q* is assumed – in fact, it is asserted, in the sense that the proponent commits to it. However, quite often such an assumption cannot be easily checked by the counterpart, thus making its acceptance (or not) a matter of trust, or lack thereof. Consider the following example:

(1) Most birds fly, and Tweety is a bird. Therefore, Tweety probably can fly.

This is a slight variation of Pollock's (1995) famous example of defeasible reasoning. Notoriously, the fact that Tweety is a penguin, or an ostrich, defeats the argument, and it does so also in this version: the fact that Tweety can fly is no longer probable, once you know that Tweety belongs to a non-flying species of birds. But the interesting twist is that, once the argument is so defeated, its original premises (most birds fly, and Tweety is a bird) are revealed as being inadequate not because they are false (on the contrary, they are indisputably true), but because they are completely irrelevant, and, in hindsight, always were. If Tweety belongs to a non-flying species, facts about its inclusion in the category of birds and the properties of flying birds species (to which Tweety does not belong) have no bearing on whether it can fly or not: that is, those premises are irrelevant to establishing that conclusion.

Yet we speculate that most arguers, lacking previous knowledge on the exact species of birds to which Tweety belongs, would be ready to accept (1) as a good presumptive argument – that is, they would be willing to trust in the relevance of its premises for the proposed conclusion. Moreover, we argue that they would usually be justified in doing so: that is, the fact that our trust in premise relevance can turn out to be misplaced does not imply that granting such trust is always a reasoning error. Whether or not the arguer is guilty of 'hasty trust' depends on several considerations, which we will discuss in the next section. For the moment, the important point is to have established that trust in internal relevance indeed occurs within argumentative exchanges, whether or not relying on it is normatively correct.

Incidentally, the fact that the relevance (or lack thereof) of a premise depends on other premises that are considered in conjunction with it has already been noted in the literature, for instance, by Johnson and Blair: 'Relevance (unlike truth) does not inhere in each individual premise but rather in an individual premise taken together with certain other items of evidence or information' (1983, p. 39).

Another interesting example of how trust in relevance matters for argumentation is given by the following enthymeme (taken from Paglieri & Woods, 2011a):

(2) All butterflies are short-lived; therefore, the *Psittacula krameri* is short-lived.

This argument is very naturally interpreted as implicitly assuming something like 'the *Psittacula krameri* is a butterfly'. The fact that this statement is false (it is a small parrot, not at all short-lived) is only meant to add further emphasis to the point, which is the following: what prompts us to favour that interpretation of the argument is not any encyclopaedic knowledge on birds and butterflies, but rather a presumption of relevance for the one premise which is stated in the argument – i.e. 'all butterflies are short-lived'. That statement would be completely irrelevant to determine the average life length of the *Psittacula krameri*, unless it was also assumed that it is a species of butterfly. In other words, the interpretation of the enthymeme here is driven by trust in internal relevance.

To sum up, trust in internal relevance can affect both argument acceptance, as in (1), and argument reconstruction, as in (2). Moreover, we suggest that this is not a rare occurrence in everyday argumentation. This is obviously an empirical claim, for which appropriate corroboration will have to be sought – something that is not within the scope of the present contribution. However, as *prima facie* evidence of the relative frequency of trust in premise relevance, consider the following: *a need for trust in relevance is created whenever the link between premises and conclusion is neither self-evident nor already known to the interpreter*. But arguments where that link is self-evident or already known are typically of little practical value, and thus rarely used, precisely because their inferential contribution is so obvious. Take the mother of all enthymemes, 'Socrates is a man, therefore Socrates is mortal': while this is certainly a fine illustration of a truncated syllogism, it is unlikely the argument will ever be used in a real debate, since everyone is already well-acquainted with the relationship between being a man and being mortal, as well as with the fact that Socrates is indeed a man. On the contrary, there is practical interest in using arguments in which the link between premises and conclusion is neither self-evident nor already known to the counterpart, since these are both more informative (if the situation is fully cooperative) and more useful in presenting the opponent with novel challenges (if the context is adversarial). Hence, we speculate that interpreting and assessing arguments in everyday dialogues will often involve trust or distrust in the relevance of their premise for the conclusion.

It is also worth noting that, in interpreting (1), we trust its premises to be not only relevant, but also *maximally* so – roughly speaking, we trust that they offer all the relevant evidences needed to support as strongly as possible the conclusion, to wit, whether Tweety can fly or not.[4] Hence (1) is interpreted by assuming, among other things, that the arguer does not know whether Tweety belongs to a non-flying species of bird or not. Indeed, we would feel cheated, if we were to find out later on that the arguer knew Tweety to be an eagle all along, since then it would have been maximally relevant to argue along the following lines:

(3) Eagles typically can fly, and Tweety is an eagle. Therefore, Tweety most probably can fly.

The premises in (3) have much greater presumptive weight for the disputed conclusion, thus they should be offered in place of those in (1), whenever possible. Sure, even (3) is defeasible: Tweety may happen to be an eagle with clipped wings, or so wizened by old age as to be unable to fly. Again, such added premises would make the original premises in (3) utterly irrelevant, and thus we would expect the arguer not to use (3), if s/he knew about these exceptional circumstances. In other words, in (3) too we trust the arguer to be maximally relevant, assuming s/he has no knowledge of any reason why Tweety may be a non-flying eagle.

Before proceeding any further, it is important to address a potential objection: how is the proposed analysis of trust in relevance any improvement on other, well-known ways of dealing with cooperative dynamics in meaning construction? After all, readers familiar with pragmatics and philosophy of language might be experiencing a strong déjà vu at this point: disambiguating (1) and (3) requires little more than a straightforward application of Grice's maxim of relevance, whereas enthymeme reconstruction in (2) seems to be guided by a relevance-based version of the principle of charity (for critical discussion of the role of charity in enthymeme interpretation, see Lewinski, 2012; Paglieri & Woods, 2011a, 2011b). Both are old friends of any student of argumentation. Not to mention the fact that 'trust in relevance' seems oddly interchangeable with 'presumption of relevance', and presumptions in turn have a long scholarly tradition in argumentation theory (Godden & Walton, 2007; Kauffeld, 1998, 2003; Macagno & Damele, 2013; Walton, 1996, 2014a). So why should we embark in this complex analysis of trust in relevance, when more familiar analytical tools are available?

We see at least three benefits in this theoretical move. First, trust is a key ingredient in any type of social interaction, whereas presumptions tend to be discussed in relation to specific phenomena (e.g. language use) or contexts of application (e.g. legal presumptions). Even accepting a certain overlapping between presumptions of relevance and trust in relevance, it would still be useful to clarify that accepting the relevance of an argument requires an attitude of trust, as so many other forms of social interchange do. While language is certainly a unique social phenomenon, its uniqueness does not depend on trust: what linguists and argumentation scholars are used to label presumptions are just acts and attitudes of trust, rather than some specific feature of language alone.

Second, analysing judgements of relevance in terms of trust emphasises the fact that they are not the undiscriminated output of some automatic process, as we shall discuss more thoroughly in the next section. While trust in relevance is often granted by default, this does not mean trusting without reason or justification. It suggests instead that reasons and justifications are needed to *suspend* the default of trust, rather than the opposite. Incidentally, the whole Gricean view of language is amenable to be reinterpreted in terms of default heuristics, both on the side of the speaker ('Unless you have specific incentives to deceive or misinform, provide relevant and true information') and on the side of the interpreter ('In the absence of specific reasons for diffidence and doubt, assume the message to be true and relevant'). As discussed elsewhere (Paglieri & Castelfranchi, 2010b; Paglieri & Woods, 2011a), such default trust attitudes are parsimonious without making us unduly vulnerable to deception.

This relates to the last and most decisive reason for favouring an analysis of relevance in terms of trust, as opposed to a more standard presumptive account. Trust is a stronger notion than presumption; or, to put it differently, trust is a highly specialised form of presumption, so the latter, broader notion cannot accurately describe the former. Trust, before becoming an act and a relationship, is an attitude towards the other, an attitude made of presumptions (or better, *evaluations*) on the qualities of someone or something else, and of presumptions (or better, *expectations*) about her/his behaviour. But a true expectation, as well as the trust than can be based on it, is more than a mere presumption, since it requires a positive conviction regarding a future state of affairs; a possibility is not enough (it might be true/relevant), we need at least a positive probability (it is likely to be true/relevant). Such probability has to be high enough for us to decide to trust as an action, that is, to make ourselves vulnerable, to take some risk. This is exactly the type of risk we expose ourselves to whenever we take at face value the qualities of an argument, including its relevance.

In argumentation theory, attempts to analyse the notion of presumption tend to reveal this natural link with the stronger notion of trust, albeit the label itself is rarely, if ever, used. For instance, Kauffeld (2003) describes ordinary presumptions of truthfulness as based on supposing the speaker to be unwilling to risk criticism, resentment, retribution, etc., for failing to be sincere and reasonably knowledgeable on what s/he is talking about. Albeit Kauffeld, to the best of our knowledge, does not discuss presumptions of relevance, similar considerations would easily apply. But seen in this light, presumption is just a technical term to talk about trust:[5] in particular, Kauffeld is suggesting that we trust arguers to be motivated to be truthful (and relevant, we hasten to add) by fear of negative repercussions, given current social conventions. If one is to accept this approach, then the notion of trust appears well suited to convey the high stakes involved even in the most basic act of reliance on others as veridical and relevant arguers. In other words, what grounds our linguistic presumptions is the trust we have reason to put in our fellow arguers, either due to general considerations (e.g. social conventions and principles of parsimony) or to specific motives (e.g. past experience with those arguers, knowledge of their incentives and expertise).

4. When is it OK to trust in premise relevance?

Even granting that we frequently rely on trust regarding the internal relevance of an argument, are we ever justified in doing so? Under what circumstances? And what kind of justification is involved – epistemic, pragmatic, or both? Before even beginning to answer these questions, one needs to distinguish between social expectations and rational justifications, since both are involved in trusting relevance. While these two dimensions are intertwined, they are not to be collapsed or confused. In fact, epistemic vigilance (in the sense of Sperber et al., 2010) and social competence are often at odds with each other: even when challenging or subjecting to critical scrutiny a certain argument is rationally justified, it could also be considered as socially inappropriate – and rightly so, depending on circumstances. Imagine a social dinner, with people engaged in a deliberation dialogue (Walton, 1998) on which wine to order, and then John puts forward the following argument from expert opinion: 'I know my wine, and this 2006 Barolo is sure to be good. We should go for it!'. To which, according to the standard treatment of this argument scheme, it would be perfectly rational for Lucy to reply: 'Do you really know what you are talking about?' Obviously, the most likely rejoinder to that critical question would be an angry look, not a well-argued reassurance on John's credentials as a wine expert, based on the fact that Lucy's comment would be taken as aggressive – which it is, given current social norms on conversation.

The independence between rational justification and social competence works both ways, of course: the fact that Lucy's comment breaches social conventions does not make it any more (or less) justified on rational grounds – after all, John may very well be utterly ignorant on wines, in spite of his boastful remarks. But this is exactly our point: rational justification and social competence are not one and the same.

This has one important implication for establishing the rational grounds for judgements of relevance: while considering under what conditions trust in premise relevance is warranted, we must acknowledge two aspects to this question. First, whether trusting in relevance is socially justified. Second, whether doing so is based on sound reasons, be they epistemic or practical. Let us recall the argument in (2), 'All butterflies are short-lived, therefore the *Psittacula krameri* is short-lived'. Imagine that a hypothetical receiver trusts in the relevance of its premises, thereby accepting something like 'The *Psittacula krameri* is a butterfly', and let us stipulate that doing so is rationally ill-advised in this situation – for instance, because there were reasons to suspect the speaker to be dishonest, incompetent, or both. Still, trust in relevance remains socially legitimate, insofar as it is demanded by the mere assertion of (2), as part of its pragmatic meaning: that is, understanding (2) includes assuming its stated premise to be relevant, hence whoever asserts (2) also puts forward such a relevance claim. Granted, taking at face value the argument as relevant may very well be a sign of epistemic laziness. But this does not make the proponent of the argument any less guilty of abusing the trust of the counterpart (or attempting to).

A simple analogy will make the point even clearer: imagine that a young girl decides to walk at night in a notoriously dangerous district of her hometown, and as a result ends up being gang-raped. It would be factually correct, albeit very cruel, to say that she brought such a tragedy upon herself by failing to exert even the obvious minimum of caution. In fact, if the girl somehow had avoided being raped and her parents had learned of her escapade, they would have severely scolded her for such a basic lack of judgement, and rightly so. But none of this makes gang rape, attempted or perpetrated, any less a crime, precisely because one is socially entitled to personal safety, regardless of any prudential consideration. The same applies to trust in relevance, and more generally to the role of trust in argumentation: whether or not we have good reasons for trusting in internal relevance, the mere fact that something is proposed as a premise for a conclusion authorises that trust, as far as basic principles of cooperation are concerned.

Again, this works both ways: precisely because the very functioning of communication requires *a basic level of cooperation*, a speaker is *entitled* to expect, and even *demand*, a certain *degree of trust* from the counterpart, unless there is previous evidence to the contrary.[6] This is why, in the dinner example, John would be entitled to resent Lucy's scepticism on his knowledge of wines, or, similarly, we often react poorly to any sign of distrust from a friend or a spouse, no matter how acceptable their doubts may be on rational grounds. 'You just have to trust me!' is a frequent refrain in conversation, one that usually is not used to deceive or avoid scrutiny, but rather to stress the kind of social relationship that has to be maintained throughout a civil debate.

It is easy to lose sight of this balancing act between rational questioning and social demands, especially in argumentation theory, where there is an undeniable tendency to favour purely logical and dialectical factors, as opposed to a broader attention to social conventions (on this point, see Gilbert, 1997; Santibañez, 2012). As a case in point, with respect to relevance, Walton proposes that 'the presumption is that [an argument] is irrelevant unless its proponent can show otherwise' (2004, p. 178). A moment of consideration will show that this principle, taken literally, is much too demanding to capture reasonable everyday interactions. Consider the following exchange, inspired by a real conversation between a father and his two-year-old son (indeed, a very typical interaction with children around that age):

(4) *Father:* Stop biting your brother!
Son: Why?
Father: Because you are hurting him, and that is a horrible thing to do!
Son: Why?
Father: Because he loves you, and he would never do the same to you!
Son: Why?

At this point, the father typically stops arguing and resort to better pedagogical means, e.g. threats and penalties – by the way, those who invariably consider the *ad baculum* a fallacy either have no children, or are not very good at educating them. However, each of the son's questions is a challenge of relevance, and the father is arguably failing to answer them: more precisely, he does answer the challenge on the first issue, only to face a similar challenge to his answer, and so on. So, if we take Walton literally, the son is being perfectly rational, whereas the father is not. But that literal interpretation cannot be correct, since we all perceive the son as being unhelpful and obdurate in this exchange (possibly without malice), and this perception is justified by the fact that arguers do not always have the obligation to 'show', in the sense of demonstrating, that their premises are relevant. At the more basic level, the (internal) relevance of a premise to a conclusion is a matter of self-evidence, or, as Blair put it, of *gestalt* (1992, p. 216): the facts that (i) one is hurting a sibling and (ii) doing so is a bad thing are manifestly relevant to conclude that (iii) one should stop, with no need for further probing. This is why we see the child's requests here as understandable but ultimately illegitimate: the father has an educational duty to instil certain basic principles in the child, not an obligation to rationally justify them against never-ending scepticism.[7]

None of this should be taken as suggesting that internal relevance is always to be trusted, with no questions asked. Even if social conventions place on us a demand for trust in each other's utterances, and critical probing on relevance (as on anything else) has to stop somewhere, trust in internal relevance can also be based on further epistemic foundations, and often is.[8] Conversely, each of these reasons for trust can also become a reason for distrust, if it appears clearly defective in the arguer and/or in the argument. Thus, the following also count as epistemic and pragmatic roots for trust in relevance:

- considerations pertaining to the arguer in general, such as competence (is s/he knowledgeable enough on the proposed connection between premises and conclusion?), motivation (does s/he have any reason to misled me?), and generic benevolence (is s/he a friend or an enemy?);
- considerations pertaining to the arguer in the context of the ongoing debate (is s/he committed? what is at stake? etc.), in order to assess both the degree of attention s/he is paying to the quality of his/her argument, and the level of scrutiny most appropriate for the interpreter;
- the interpreter's own position, given the dialogical situation (are we competing or cooperating?), to calibrate one's expectations on the counterpart's moves; and
- features of the argument itself, based on the interpreter's background knowledge: for instance, the internal relevance of (2) is more easily trusted due to the well-known convention of using Latin names for species, whereas trust would be harder to elicit (and much less justified) for something like 'All butterflies are short-lived, therefore the *Be-Pop-A-Lula* is short-lived'.

In a sense, one could interpret these as reasons to endorse a higher or lower degree of epistemic vigilance (Sperber et al., 2010), with respect to internal relevance. Moreover, it is worth noting that, in the absence of indications to the contrary, trusting in relevance without further scrutiny may very well be the default option, and rationally so. As has been argued elsewhere (Paglieri & Castelfranchi, 2010a, 2010b; Paglieri & Woods, 2011a, 2011b), there is a deep connection between basic cooperative principles and parsimony, that is, the efficient allocation of scant resources to the production and interpretation of arguments. With respect to relevance, this would imply that (again, in the absence of any sign of foul play) a climate of generalised trust in relevance is more convenient than one of widespread distrust or continuous critical scrutiny, thus justifying a default trusting attitude.

To sum up, it seems that trust in internal relevance is not only a necessary ingredient for argumentation, but also a justifiable attitude, on a variety of grounds: compliance with social conventions (including basic principles of communication), the necessary finite nature of sceptical probing, several considerations regarding both the arguer, his/her proposed argument and the context of the dialogue, as well as a tendency towards parsimony in argumentation. Once again, this is not at all intended to suggest that we are always right, morally or rationally, in trusting the relevance of each other's arguments. However, it certainly vindicates the opportunity for further study of the role of trust in argumentative relevance, and suggests that such study need not be confined to descriptive models, but could also encompass normative concerns.

5. Formal models and computational applications for trust in argumentative relevance

As noted in the Introduction, much attention has been paid to trust and argumentation in computer science: more recently, their interaction has also started to be explored in greater detail (for an up to date survey, see Koster, 2014). The most natural way of integrating these two aspects of information dynamics is by looking at how trust in information sources ought to affect the evaluation of their arguments. Somewhat ironically, this would have been traditionally regarded as the basic *ad hominem* fallacy, since of course the quality of the source is not necessarily a reliable indication of the quality of its message. Similarly, even the presence of a large number of independent sources converging on the same opinion would once have been considered an *ad populum* fallacy, based on the fact that all those sources could still be wrong. However, as we discussed above, it has long been recognised that features of the source (including convergence with other independent sources) may nonetheless provide robust presumptive reasons to accept his/her claim, in the appropriate context: for a recent application of this approach to arguments from expert opinion, see, for instance, Walton (2014b). More generally, computer science has always found the connection between source and information to be a central one, especially in distributed artificial intelligence, e.g. in multi-agent systems: when computer agents have to interact autonomously with each other, it becomes essential to reliably assess information based on the reputation of its source, as well as to form well-grounded judgements of its trustworthiness in current circumstances.

Similar considerations apply also to the exchange of arguments with or among artificial agents, thus existing models could be extended to deal with relevance as well. Here, we will briefly discuss three approaches suitable for such an extension, only to hint at how it might be achieved: the modelling of relevance as a feature of arguers, alongside competence and sincerity (Paglieri et al., in press), the use of modal logic to capture trust in relevance (Amgoud & Demolombe, 2014), and the application of meta-argumentation techniques (Boella, Gabbay, van der Torre, & Villata, 2009; Villata, Boella, Gabbay, & van der Torre, 2013) to express claims of relevance, as well as challenges to them, within an abstract argumentation framework (Dung, 1995).

Paglieri et al. (in press) focus on a relatively unexplored aspect of information dynamics: how the quality of a message should impact on the subsequent assessment of its source (*feedback dynamics*). The issue was discussed theoretically by Falcone and Castelfranchi (2004), but formal models tend instead to concentrate on how the assessment of the source influences the evaluation of the message. In contrast, Paglieri et al. emphasise the bilateral relationship between sources and messages. Their work, however, deals with informational messages in general, and analyses only two features of sources: their competence (i.e. whether or not they are likely to have accurate knowledge of a given domain) and their sincerity (i.e. whether or not they are likely to provide honest information, as opposed to lying). Both dimensions are captured using bipolar pairs of graded beliefs (based on Dubois & Prade, 2008), to allow for ignorance on the source features: that is, an agent may be in the dark as to whether source S is competent (sincere) or incompetent (insincere). Their model details how to integrate these dimensions to generate an expectation on the quality of information provided by the source, and how to differentially attribute praise (or blame) to each feature when that information turns out to be better (or worse) than expected.

While the specifics of their approach are beside the point here, relevance appears to be conspicuously absent from the features to be considered in assessing a source – a fact the authors acknowledge at the end of the paper (Paglieri et al., in press). In contrast, we obviously have a right to expect our sources to be not only knowledgeable and honest, but also capable of providing relevant information (Paglieri & Castelfranchi, 2012). The same applies to arguments, both with respect to external relevance (a perfectly valid argument on the literary value of Gothic novels is not much use in deliberating where to spend a family vacation) and with reference to internal relevance (an argument with irrelevant premises is necessarily of poor quality, even if it might seem otherwise in the case of fallacies). Thus, it would be useful to define relevance also as *a property of sources*: more precisely, as the property of being likely to provide relevant information. Then such property could be modelled with the same formal tools used for competence and sincerity, to study the interaction between these three features in shaping argument evaluation, as well as the feedback they receive when argument quality proves to differ significantly from one's expectations. Among other things, integrating relevance into the picture would allow capturing a special family of dialectical infractions: *omissions*. When a source is aware of an argument relevant to the matter under discussion, and yet fails to mention it, then that source can be blamed for omitting something, due to a lack of understanding or honesty. In other words, an omission is not simply a failure to mention any valid argument, but rather a failure to validly convey relevant information. Without a proper analysis of relevance, omissions cannot be identified, because one cannot distinguish between the harmless (indeed, beneficial) habit of passing under silence everything that does not matter, and the damaging policy of withholding valuable information, on purpose or by mistake (reticence).

The modal approach to trust in information sources pioneered by Demolombe (2004) distinguishes six properties of the sender that the receiver may trust in, based on the relationship between informing, believing, and truth. Later on (Amgoud & Demolombe, 2014), it has been proposed to use argumentation to allow reasoning about such properties. Four of the properties identified by Demolombe are listed below:

- trust in *sincerity*: the truster believes that if he is informed by the trustee about some proposition, then the trustee believes that this proposition is true
- trust in *competence*: the truster believes that if the trustee believes that some proposition is true, then this proposition is true
- trust in *vigilance*: the truster believes that if some proposition is true, then the trustee believes that this proposition is true
- trust in *cooperativity*: the truster believes that if the trustee believes that some proposition is true, then he is informed by the trustee about this proposition.[9]

As noted by Paglieri and Castelfranchi (2012), the fact that relevance is overlooked in this approach is problematic, since in any situation of minimal complexity, the omniscience required by vigilance (believing all that is true) is neither feasible nor necessary, and the over-sharing attitude preached by cooperativity (mentioning everything that one believes) is not even desirable. We certainly do not require good sources to know everything, and we most definitely do not want them to blurt out all their beliefs, even if they happen to be true – and the same applies to arguments, *mutatis mutandis*.[10] In contrast, what we expect from a good source (arguer) is to possess and share relevant knowledge (arguments). Moreover, it is easy to see how relevance projects its own 'varieties of trust', similarly to those proposed by Demolombe for truth. Thus we end up with the following taxonomy (taken from Paglieri & Castelfranchi, 2012):

- trust in *pertinence*: the truster believes that if s/he is informed by the trustee about some proposition, then the trustee believes that this proposition is relevant for him/her
- trust in *understanding*: the truster believes that if the trustee believes that some proposition is relevant for him/her, then this proposition is relevant for him/her
- trust in *knowledgeability*: the truster believes that if some proposition is relevant for him/her, then the trustee has some information about it

- trust in *sharing*: the truster believes that if the trustee has information about a relevant proposition, then s/he is informed by the trustee of it.

The same relationships identified by Demolombe apply equally well to truth and to relevance, and these new properties should be included in the assessment of source quality: that is, a source is of high quality only if its information turns out to be both true and relevant (an expansion of Demolombe's validity), and if it has access to enough true and relevant knowledge (an expansion of Demolombe's completeness). Even more radically, while correctness and relevance are both essential and yet independent constraints on source quality, completeness (vigilance plus cooperativity) should be more properly redefined in terms of relevance, as noted above.

The rationale of integrating relevance in this modal approach is quite straightforward, yet its formal implementation may be less obvious. The original approach (as detailed in Amgoud & Demolombe) includes two modal operators, one for belief and the other for information sharing, as follows:

$\mathrm{Bel}_i\phi$ the agent i believes that ϕ holds.

$\mathrm{Inf}_{j,i}\phi$ the agent j has informed the agent i about ϕ.

To capture relevance as well, one has to either introduce it as a primitive feature of information or extend the formalism with a modal operator for goals, in the vein of Cohen and Levesque (1990). The latter seems to us a more promising strategy: once goals are included (e.g. $\mathrm{Goal}_i\phi$: the agent i wants that ϕ holds), then relevance can be defined in terms of goals, including sub-goals, that is, means to an end. Thus, informally, being informed about ϕ is relevant for agent i if and only if ϕ is either a goal or a sub-goal of i. Interestingly, if the goal operator is interpreted as capturing objective interests (what the agent needs, as a matter of fact) and not subjective desires (what the agent wants, whether or not it is actually good for him/her), then it is also possible to distinguish between objective and subjective relevance (Paglieri & Castelfranchi, 2012). While the informal definition of relevance given above covers objective relevance, i.e. what the agent needs in fact to know, it is easy to define also subjective relevance, i.e. what the agent thinks (possibly mistakenly) that s/he needs to know. In particular, being informed about ϕ is subjectively relevant for agent i if and only if i believes ϕ to be either a goal or a sub-goal of i. Thus, objective relevance is measured against goals, whereas subjective relevance depends on beliefs about one's own goals.[11] Once the model is thus extended, argumentation can then be used to reason on trust in relevance, as well as trust in correctness and completeness. Or, more drastically, one might try to shift the focus from information sources to proper arguers, replacing propositions with arguments and truth with validity. Whether or not these extensions may prove fruitful, though, is a matter for future work.

Finally, an interesting aspect of relevance is that it cannot be directly captured in an abstract argumentation framework (Dung, 1995), one of the most popular methods to formalise argument structures. The reason behind that limitation is twofold: first, the approach focuses on relationships between arguments and not on the internal structure of each argument, making internal relevance impossible to capture; second, the attack relationship is defined between nodes (arguments), whereas a challenge of relevance is essentially an attack against an edge of the argumentation graph, in that it questions the fact that argument p has anything to do with argument q.[12] This is why capturing the notion of relevance in that context requires expanding Dung's framework with meta-argumentation (Boella et al., 2009; Villata et al., 2013), in which binary relationships between arguments, e.g. support and attack, can be explicitly represented and reasoned about at the meta-level. It remains unclear, though, how this might enable to express specific criticisms of relevance, as opposed to more general attacks against the validity of a proposed relationship between two arguments. After all, challenging the claim that 'p supports q' or 'p attacks q' does not specify the reason behind the challenge: it could be relevance, but it could also be an objection against sufficiency – e.g. p is acknowledged as relevant for q, but insufficient to presumptively support q. Further work on meta-argumentation will be needed, if this method is to be used to model trust in relevance, along the lines pioneered by Villata et al. (2013).

Even if many possibilities are open to the formal modelling of trust in argumentative relevance, one might remain sceptical on the technological implications of this line of research. Do argument-based software applications really need to incorporate trust in relevance? The answer is a resounding 'Yes!'. Not only because, quite obviously, the circulation of irrelevant chatter, be it among humans or artificial agents, is a capital sin in this age of information overload, and the ability to detect relevance a cardinal virtue (on that topic, it is always instructive to read Brin & Page, 1998, which illustrates the core ideas on relevance of what later became a small company known as Google). But also because users of argumentation technologies have the need for, and thus a justified expectation of, systems able to be *maximally relevant*, in the sense discussed in Section 3. Imagine a user who must decide whether to take medicine X to treat illness Y, with the help of an argument-based decision support system; there are no arguments against taking medicine X in the knowledge base of the system, while it contains the following three arguments in favour of that course of action:

(A) Clinical trials show that taking medicine X triples the life expectancy of patients with illness Y. Therefore, you should take medicine X.
(B) Medicine X is less painful than other available treatments for illness Y. Therefore, you should take medicine X.
(C) Medicine X is less expensive than other available treatments for illness Y. Therefore, you should take medicine X.

Obviously, the user would be extremely disappointed (and rightly so) if the system did not recognise argument (A) as being by far the most relevant, and thus the first or even only one to be offered in recommending use of medicine X. The difference between these arguments is precisely one of internal relevance – more exactly, what Walton (1992) calls probative relevance. In this context, presenting all three premises to support use of the medicine would make the argument redundant, since it would use more premises than those needed to prove the conclusion. As Vorobey (2006) noted, redundant arguments, while formally valid, are 'flawed as an exercise in rational persuasion' (p. 89). Once presented with (A), we no longer care for (B) and (C), thus insisting that presenting them would constitute a nuisance for the user, and a failure at being maximally relevant.[13] In argumentation as elsewhere, flogging a dead horse is not a welcome strategy, nor a very rational one. For similar reasons, Rubinelli, Wierda, Labrie, & O'Keefe (2011) advocate using pragma-dialectics in AI (Eemeren & Grootendorst, 2004), and in particular the notion of strategic manoeuvering (Eemeren, 2010), to model relevance with respect to the concerns and goals of the target audience, in the context of health promotion interventions.[14]

6. Conclusion and future work

In this paper, we tried to make a case for the study of trust in argumentative relevance: after a brief survey of the literature, we proposed a distinction between external and internal relevance of arguments, and contended that trust is especially central to the latter; then we discussed several rational grounds that may justify the act of trusting in the relevance of an argument, and sketched possible formal developments for the ideas presented here, as well as their import for argumentation technologies. While much more remains to be said on the connection between trust, relevance, and arguments, we hope to have persuaded readers that this is indeed a worthy line of inquiry, thereby luring others to join us in further exploring its ramifications.

Nonetheless, we would like to close this paper with a word of caution, regarding its technological implications. While the relevance of trust for establishing argumentative relevance (pun intended) is hardly in doubt, its actual implementation faces several practical difficulties. In that regard, we do not completely share Walton's optimism on the promises of heuristics (as those discussed in Pearl, 1984) for the automatic assessment of relevance (Walton, 2004, pp. 120–123). In fact, while Walton's idea of using backward and forward chaining of arguments as (part of) a method to establish relevance has merits, it is likely to be often quite impractical, or even impossible to implement – namely, whenever we lack a shared knowledge structure and inference rules with our interlocutors and/or end-users of argumentative technologies. Walton is aware of this difficulty, but seems to see it just as a minor inconvenience, whereas we consider it a major obstacle:

> Ideally, if an argument really is relevant in a dialogue, it will be connected to the issue of the dialogue by a chain of argumentation. But in reality, there may be a big gap to be filled between the argumentation in a given case and the ultimate thesis or goal at which argumentation is supposed to finally arrive. What has to be done is to try to chain the argumentation forward to see if it could really be extended to reach its supposed ultimate point. (2004, p. 177)

To be clear, we do not object to Walton's method in principle, yet we remain doubtful that the resources needed for its application are available for most real-life applications. Imagine for instance an argumentative exchange on a social network, e.g. in the comments to a Facebook post: the identification of both the ultimate goal, the background knowledge, and the shared inference rules pertinent to that context is likely to be problematic even for a human interpreter, which currently makes it nearly impossible for automated systems (regardless of the undeniable opportunities for argumentation research in social networks, as noted by Heras et al., 2013). Also, automated relevance assessment requires argument mining, i.e. the automatic detection of argument structures in natural language dialogues (Abbas & Sawamura, 2008; Mochales, Palau, & Moens, 2011; Wyner, Schneider, Atkinson, & Bench-Capon, 2012). This is in and by itself a formidable challenge for argumentation technologies – one that many scholars are addressing, and yet which remains far from being solved. None of this should suggest a defeatist attitude: on the contrary, the objective difficulties in the automatic detection of argumentative relevance combine with its importance in making it a priority for future research. Hopefully, a better understanding of the role of trust in human judgements of argumentative relevance will contribute to the success of this scientific enterprise.

Acknowledgements

This research was financially supported by the PON research project Interoperable Cloud Platforms for Smart Governments' (PRISMA), founded by the Italian Ministry for Education, University and Research (MIUR). We are grateful to Rino Falcone, John Woods, Floriana Grasso, Chris Reed, Emiliano Lorini, Serena Villata, Celia Costa Pereira, Andrea Tettamanzi, Andrew Koster, and the members of the Goal-Oriented Agents Lab (GOAL) at the ISTC-CNR for insightful discussion of many issues presented here, and to two anonymous reviewers for their excellent comments and criticisms of a previous version of this paper. All remaining mistakes and omissions are our sole responsibility.

Notes

1. In contrast, Hamblin interpreted *ignoratio elenchi* as the misconception of refutation, scoffing at the idea of identifying it with the fallacy of irrelevant conclusion (Hamblin, 1970, pp. 87–88).

2. Even more crucially, it also entails a significant reduction of cognitive and social costs, on many demanding activities like search and evaluation, decision-making, negotiation, monitoring, etc. (see Castelfranchi & Falcone, 2010).
3. To be clear, our claim is that assessing external relevance does not *typically* require trust: this clearly countenances some exceptions. One such exception is the case of an arguer straw-manning the position of a third party, rather than that of the counterpart in the ongoing argumentative exchange. When it is the opponent's position to be misrepresented, this violation would be rather self-evident (assuming the opponent at least understands his/her own case), thus involving little trust; but when the opponent has to judge whether the arguer's reconstruction of the standpoint of a third party is accurate, then trust in external relevance becomes a key concern. Although similar cases are certainly worthy of further analysis, this task will not be undertaken here.
4. In Section 5, we will discuss how eminently relevant reasons for a conclusion may well render irrelevant any additional reason that points to the same conclusion. Once you have an iron-clad case for *p*, any further argument for *p* becomes moot: more precisely, even if further arguments would be relevant under different circumstances (i.e. in the absence of another overarching reason for *p*), they are not relevant in the present context. This view is compatible with the notion of relevance favoured by Sperber and Wilson (1995): once your point is proven beyond doubt, further reasons for it add to communicative costs without producing any benefit, thus reducing relevance.
5. As Godden and Walton (2007) put it, what Kauffeld proposes is an expectation-based theory of presumptions, thus naturally akin with the socio-cognitive notion of trust (Castelfranchi & Falcone, 2010).
6. This demand for trust in relevance (more exactly, this expectation that the interpreter will assume the premises to be relevant to the conclusion) captures a key aspect of Grice's (1989) principle of cooperation, often misinterpreted. His maxim of relevance (as all the others, for that matter) does not maintain that speakers will always be relevant, nor that interpreters will always assume them to be. Rather, the maxim states that understanding an utterance (in this case, an argument) implies taking it as an attempt (successful or failed, by mistake or on purpose) to be maximally relevant: even if the premises of your argument are not really relevant to its conclusion, you are presenting them to me as if they were, and you want me to understand them as such; and even if I see through your fallacious move and recognise the argument as lacking internal relevance, I still understand your intent of making me assume the opposite – otherwise I am simply failing to recognise the communicative intention behind your act.
7. In a similar vein, Botting (2013) has recently argued that a satisfactory formal analysis of relevance is impossible, on pain of infinite regress – precisely because whatever reasons one might individuate to support a judgement of relevance, they will have in turn to demonstrate their own relevance for that purpose. In his own words:

   > when I speak of the irrelevance of relevance I do not wish to be taken as implying that it is not necessary to make relevance judgments; what is irrelevant is a theory or formal analysis of relevance. The idea that such judgments need a theory behind them to make them respectable is, I believe, entirely false. On the contrary, judgments intended to have a justifying function cannot be backed by such a theory without initiating a regress. (2013, p. 17)

   By way of alternative, Botting outlines instead a phenomenological account of relevance judgements.
8. Clearly, we are not suggesting that compliance with social conventions on trust in relevance is not, in and by itself, a pragmatic reason for granting such trust: it clearly is, for instance, to avoid becoming a social pariah in everyday conversation. All we want to stress is that social conventions (contra Jones, 2002) are not the *only* root of trust in relevance.
9. The remaining two properties, trust in validity (if the source informs me of something, then it is true) and trust in completeness (if something is true, then the source will inform me of it) can be obtained by combining two of the other four: sincerity and competence for validity, vigilance and cooperativity for completeness. Thus, here we will discuss only the four properties of trust listed above.
10. In fact, the ritual formulation 'the truth, the whole truth and nothing but the truth', frequently used in sworn testimony, is elliptic in that respect: obviously, it is to be understood as referring to the *relevant*

truth for the case under discussion – otherwise the second part of the obligation would be impossible to meet by any witness!

11. Incidentally, this also illustrates why objective relevance is still *subject-dependent*, as claimed by Paglieri and Castelfranchi (2012). The objective relevance of an information is a matter of fact, but the facts of the matter depends on the identity of the person who receives that information, since it is based on his/her goals.
12. This is linked to the fact that irrelevant premises are *fake* premises, that is, they purport to be premises for something, but in fact they are not. This is very different from being a legitimate premise that happens to be defeated, either because it is proved to be false, or because it is overruled by other considerations (which is the notion of attack captured by abstract argumentation).
13. The weaker relevance of B and C in the example refers to the standard case of a patient whose paramount concern is life expectancy, rather than avoiding pain and saving money. But of course one could think of instances in which these priorities change: e.g. someone terrified by any physical pain, or a very poor person looking for any cure s/he may afford. This does not change the main point: whenever the relevance of a premise is abysmal in comparison with that of other premises, insisting on mentioning it constitutes an annoyance for users.
14. A very early, yet still interesting attempt to give a Gricean account of how relevance operates in social contexts (activity types) and how this relates to AI modelling is found in Levinson (1979).

References

Abbas, S., & Sawamura, H. (2008). A first step towards argument mining and its use in arguing agents and ITS. In I. Lovrek, R. Howlett, & L. Jain (Eds.), *Knowledge-based intelligent information and engineering systems* (pp. 149–157). Berlin: Springer.

Amgoud, L., & and Demolombe, R. (2014). An argumentation-based approach for reasoning about trust in information sources. *Argument & Computation,* doi:10.1080/19462166.2014.881417.

Aristotle. (1939). *Topica*. (E. S. Forster, Trans.). Cambridge, MA: Harvard University Press.

Baier, A. (1995). *Moral prejudices*. Cambridge, MA: Harvard University Press.

Blair, A. (1992). Premissary relevance. *Argumentation, 6*, 203–217.

Boella, G., Gabbay, D., van der Torre, L., & Villata, S. (2009). Meta-argumentation modelling I: Methodology and techniques. *Studia Logica, 93*(1), 297–355.

Borg, E. (2005). Intention-based semantics. In E. LePore & B. Smith (Eds.), *The Oxford handbook of philosophy of language* (pp. 250–267). Oxford: Clarendon Press.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10), 913–925.

Botting, D. (2013). The irrelevance of relevance. *Informal Logic, 33*(1), 1–21.

Bremer, M., & Cohnitz, D. (2004). *Information and information flow – an introduction*. Frankfurt: Ontos Verlag.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In P. Enslow & E. Allen (Eds.), *Proceedings of the 7th World Wide Web conference* (pp. 107–117). Amsterdam: Elsevier.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model*. London: Wiley.

Castelfranchi, C., & Paglieri, F. (2007). The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. *Synthese, 155*, 237–263.

Castelfranchi, C., & Poggi, I. (1993). Lying as pretending to give information. In H. Parret (Ed.), *Pretending to communicate* (pp. 276–291). Berlin: de Gruyter.

Castell, A. (1935). *A college logic*. New York, NY: Macmillan.

Cohen, J. (1994). Some steps towards a general theory of relevance. *Synthese, 101*, 171–185.

Cohen, P., & Levesque, H. (1990). Intention is choice with commitment. *Artificial Intelligence, 42*, 213–261.

Cooper, W.S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*, 19–37.

Crestani, F., Lalmas, M., Van Rijsbergen, C.J., & Campbell, I. (1998). Is this document relevant?. . . Probably: A survey of probabilistic models in information retrieval. *ACM Computing Surveys, 30*(4), 528–552.

Demolombe, R. (2004). Reasoning about trust: A formal logical framework. In C. Jensen, S. Poslad, & T. Dimitrakos (Eds.), *Trust management: Proceedings of iTrust 2004* (pp. 291–303). Berlin: Springer.

Dubois, D., & Prade, H. (2008). An introduction to bipolar representations of information and preference. *International Journal of Intelligent Systems, 23*(8), 866–877.

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

Eemeren, F.H. van (2010). *Strategic maneuvering in argumentative discourse*. Amsterdam: John Benjamins.

Eemeren, F.H. van, & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge: Cambridge University Press.

Falcone, R., & Castelfranchi, C. (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. In *Proceedings of AAMAS 2004* (pp. 740–747). New York: ACM.

Federal Rules of Evidence. 2011. Retrieved from http://www.law.cornell.edu/rules/fre

Floridi, L. (2008). Understanding epistemic relevance. *Erkenntnis, 69*, 69–92.

Gambetta, D. (Ed.). (1988). *Trust: Making and breaking cooperative relations*. Oxford: Basil Blackwell.

Gilbert, M. (1997). *Coalescent argumentation*. Mahwah: Lawrence Erlbaum.

Godden, D., & Walton, D. (2007). A theory of presumption for everyday argumentation. *Pragmatics & Cognition, 15*(2), 313–346.

Goodwin, J. (2010). The authority of Wikipedia. In J. Ritola (Ed.), *Argument cultures. Proceedings of OSSA 2009,* CD-Rom (pp. 1–22). Windsor: OSSA.

Govier, T. (1985). *A practical study of argument*. Belmont, CA: Wadsworth.

Govier, T. (1997). *Social trust and human communities*. Montreal: McGill-Queen's University Press.

Govier, T. (1998). *Dilemmas of trust*. Montreal: McGill-Queen's University Press.

Grice, H.P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Falcone, R., & Castelfranchi, C. (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. In Proceedings of AAMAS 2004 (pp. 740–747). New York: ACM.

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

Heras, S., Atkinson, K., Botti, V., Grasso, F., Julián, V., & McBurney, P. (2013). Research opportunities for argumentation in social networks. *Artificial Intelligence Review, 39*, 39–62.

Jacobs, S., & Jackson, S. (1992). Relevance and digressions in argumentative discussion: A pragmatic approach. *Argumentation, 6*, 161–176.

Johnson, R., & Blair, A. (1983). *Logical self-defense* (2nd ed.). Toronto: McGraw-Hill Ryerson.

Jones, A. (2002). On the concept of trust. *Decision Support Systems, 33*, 225–232.

Kauffeld, F. (1998). Presumption and the distribution of argumentative burdens in acts of proposing and accusing. *Argumentation, 12*(2), 245–266.

Kauffeld, F. (2003). The ordinary practice of presuming and presumption with special attention to veracity and the burden of proof. In F. van Eemeren, A. Blair, C. Willard, & F. Snoeck Henkemans (Eds.), *Anyone with a view: Theoretical contributions to the study of argumentation* (pp. 133–146). Dordrecht: Kluwer Academic.

Koster, A. (2014). Trust and argumentation in multi-agent systems. *Argument & Computation*, doi:10.1080/19462166.2014.885083.

Krabbe, E. (1992). So what? Profiles for relevance criticism in persuasion dialogues. *Argumentation, 6*, 271–283.

Levinson, S. (1979). Activity types and language. *Linguistics, 17*, 365–399.

Lewinski, M. (2012). The paradox of charity. *Informal Logic, 32*(4), 403–439.

Macagno, F., & Damele, G. (2013). The dialogical force of implicit premises: Presumptions in enthymemes. *Informal Logic, 33*(3), 361–389.

Mochales Palau, R., & Moens, M.-F. (2011). Argumentation mining. *Artificial Intelligence and Law, 19*(1), 1–22.

O'Neill, O. (2002). *Autonomy and trust in bioethics*. Cambridge: Cambridge University Press.

Paglieri, F., & Castelfranchi, C. (2010a). Why argue? Towards a cost-benefit analysis of argumentation. *Argument & Computation, 1*, 71–91.

Paglieri, F., & Castelfranchi, C. (2010b). In parsimony we trust: Non-cooperative roots of linguistic cooperation. In A. Capone (Ed.), *Perspectives on language use and pragmatics* (pp. 99–117). Munich: Lincom Europa.

Paglieri, F., & Castelfranchi, C. (2012). Trust in relevance. In S. Ossowski, G. Vouros, & F. Toni (Eds.), *Proceedings of agreement technologies 2012* (pp. 332–346). Tilburg: CEUR-WS.org.

Paglieri, F., Castelfranchi, C., da Costa Pereira, C., Falcone, R., Tettamanzi, A., & Villata, S. (in press). Trusting the messenger because of the message: Feedback dynamics from information quality to source evaluation. *Computational and Mathematical Organization Theory*. doi:10.1007/s10588-013-9166-x

Paglieri, F., & Woods, J. (2011a). Enthymematic parsimony. *Synthese, 178*, 461–501.

Paglieri, F., & Woods, J. (2011b). Enthymemes: From reconstruction to understanding. *Argumentation, 25*(2), 127–139.

Pearl, J. (1984). *Heuristics: Intelligent search strategies for computer problem solving*. Reading, MA: Addison-Wesley.

Pollock, J. (1995). *Cognitive carpentry*. Cambridge, MA: MIT Press.

Ramchurn, S., Huynh, D., & Jennings, N. (2004). Trust in multi-agent systems. *The Knowledge Engineering Review, 19*(1), 1–25.

Rubinelli, S., Wierda, R., Labrie, N., & O'Keefe, D. (2011). The problem of premissary relevance. In N. Green, S. Rubinelli, & D. Scott (Eds.), *Artificial Intelligence and health communication: Papers from the AAAI Spring Symposium* (pp. 53–56). Menlo Park, CA: AAAI Press.

Santibañez, C. (2012). The principle of relevance in the light of cooperation and trust: Discussing Sperber and Wilson's theory. *Pragmatics & Cognition, 20*(3), 483–504.

Sidgwick, A. (1914). *Elementary logic*. Cambridge: Cambridge University Press.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language, 25*(4), 359–393.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Malden: Basil Blackwell.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tuomela, R., & Tuomela, M. (2005). Cooperation and trust in group context. *Mind & Society, 4*, 49–84.

Villata, S., Boella, G., Gabbay, D., & van der Torre, L. (2013). A socio-cognitive model of trust using argumentation theory. *International Journal of Approximate Reasoning, 54*, 541–559.

Vorobey, M. (2006). *A theory of argument*. Cambridge: Cambridge University Press.

Walton, D. (1992). Which of the fallacies are fallacies of relevance? *Argumentation, 6*, 237–250.

Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Walton, D. (1998). *The new dialectic: Conversational contexts of argument*. Toronto: University of Toronto Press.

Walton, D. (2004). *Relevance in argumentation*. Mahwah, NJ: Lawrence Erlbaum.

Walton, D. (2014a). *Burden of proof, presumption and argumentation*. Cambridge: Cambridge University Press.

Walton, D. (2014b). On a razor's edge: Evaluating arguments from expert opinion. *Argument & Computation*, doi:10.1080/19462166.2013.858183.

Werkmeister, W. (1948). *An introduction to critical thinking*. Lincoln, NE: Johnsen.

Wilson, D., & Sperber, D. (2004). Relevance theory. In L.R. Horn & G.L. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Malden, MA: Basil Blackwell.

Woods, J. (2013). *Errors of reasoning. Naturalizing the logic of inference*. London: College.

Wyner, A., Schneider, J., Atkinson, K., & Bench-Capon, T. (2012). Semi-automated argumentative analysis of online product reviews. In B. Verheij, S. Szeider, & S. Woltran (Eds.), *COMMA 2012 – proceedings of the 4th international conference on computational models of argument* (pp. 43–50). Amsterdam: IOS Press.