Elements of a Social Semantics for Argumentative Dialogue

Rodger Kibble[1]

Abstract. This paper represents an initial attempt to pull together some parallel strands in argumentation theory, multi-agent systems, and philosophy of language, centering around the notion of *commitment*. A framework for argumentative dialogue is proposed which is inspired by recent work in multi-agent communication, but is also intended to be broadly applicable to human argumentation. The notion of "commitment" is elaborated in terms of Brandom's notion of *entitlement*, which is itself fleshed out using Habermasian *validity claims*. The appeal to intersubjectively observable deontic statuses in place of mentalistic notions of belief and intention provides a social semantics in Singh's sense, and allows for a plausible account of the emergence of Walton's argumentation schemes as well as rhetorical devices such as RST's Evidence and Justify.

1 Argumentation-based communication

This paper represents an initial attempt to pull together some parallel strands in argumentation theory, multi-agent systems, and philosophy of language. The speech act theories of Austin [1] and Searle [14] have had unanticipated applications in agent technology, inspired initially by the demonstration in [4] that Searle's systematic analysis of speech acts such as promising, requesting, asserting in terms of preconditions and outcomes, and the beliefs and intentions of participants, could naturally be formalised in terms of planning operators. This has led to important developments in the field of multi-agent and human-computer communication.

However, agent design in terms of notions such as *belief* and *intention* faces the software engineering problem that it is not generally possible to identify data structures corresponding to beliefs and intentions in heterogenous agents [18], let alone a "theory of mind" enabling agents to reason about other agents' beliefs. This problem has been addressed by developing alternative semantics based on intersubjectively observable notions of commitments [15]. This development is in some ways anticipated by moves in the philosophy of language to eliminate or at least downgrade mentalistic notions in favour of social constructs, including Brandom's inferential semantics [2, 3] and Habermas's theory of communicative action [5, 6, 7]. A goal of the work reported in this paper is to utilise some of this analysis in order to develop a more fine-grained conceptualisation of notions like *commitment* and *challenge* in the context of computational modelling of argumentative dialogue.

As autonomous software agents play an increasing role in electronic commerce, with e-science and even e-government on the horizon, it will be essential for citizens, consumers and the intelligent agents themselves to be able to judge the trustworthiness and reliability of agents they encounter in the virtual world. An important recent development is the notion of argumentation-based communication, which insists that agents support messages with reasons why those messages are appropriate [13]. The work of Brandom and Habermas (op. cit.) suggests ways this approach can be generalised so that:

1. agents need not provide explicit argumentation but must be capable of providing justifications on demand, and of recursively supporting the justifications themselves;
2. communicative acts always involve certain implicit validity claims each of which potentially stands in need of recursive justification: the claims that a proposition is true (which may describe e.g. the content of an assertion, or the preconditions for an intended action); that the agent is truthful and reliable; and that any institutional or normative preconditions are satisfied (for instance, the agent is authorised to enter a contract or issue an instruction);
3. building on these assumptions, we can model the emergence of basic patterns of argumentation [16] and rhetorical structure [11] as strategies for pre-empting challenges of various types (cf [10]).

It is inherent in both Brandom's and Habermas's work that the integrity of communication relies on normative obligations to defend implicit entitlement or validity claims to the point where no further challenges are issued. These approaches are however non-foundationalist in that entitlement and validity are ultimately grounded in *rational consensus* rather than a demand for certainty: "doubts too sometimes need to be justified" [2, p. 177]. In particular, Brandom [2, pp. 174-6] outlines a "default and challenge" model which seems to finesse the **completeness problem** noted by [16]: entitlement to claims is often attributed by default, and justification is ultimately grounded in shared norms which habitually go unchallenged within a community. The **contestability semantics** of [12] appears somewhat less forgiving, being designed for the MAS world where we may expect an absence of shared socio-cultural norms among agents.

2 Entitlements, validity claims and challenges

The notion of *commitment* in argumentation goes back at least to [8]. More recently, Robert Brandom [2] has developed the normative dimension of *entitlement*: assertion is modelled as undertaking a commitment to defend a claim, to which the speaker may or may not be entitled on grounds of empirical evidence or inference. Entitlement may be provisional if supported by a defeasible inference: for example the inference from *this is coffee* to *this tastes good* can be voided if machine oil has been added to the beverage. It is important to note that challenges to entitlements themselves stand in need of entitlement.

[1] Goldsmiths College, University of London, UK email: r.kibble@gold.ac.uk

© 2004
CMNA IV, 4^{th} *Workshop on Computational Models of Natural Argument*
Edited by Floriana Grasso, Chris Reed and Giuseppe Carenini
ECAI 2004

I propose to add some structure to this notion by incorporating a variant of Habermas's threefold "validity claims" (*Geltungsansprüche*). (Cf also [17, 15, 12].) Under the formulation in [6] utterances raise three simultaneous claims, which the speaker undertakes to defend: they must be true (*wahr*), sincere or truthful (*wahrhaftig*) and "right" or appropriate to social norms (*richtig*). Joseph Heath [9] argues against the notion that every speech act raises all three claims, and proposes that Habermas's account can only be made coherent on the assumption that the validity claims are associated with different types of discourse: theoretical, practical and expressive. Space does not permit elaboration of this issue, though I will note that [7] introduces a distinction between "weak" and "strong" communicative rationality, whereby the former involves only the truth and sincerity claims. Most instances of multi-agent communication in the current state of the art would probably count as weakly rational in this sense.

For now I will adopt an approximation to Habermas' scheme whereby entitlement to doxastic commitments in persuasive argumentation can be challenged or defended under one of the following headings[2]:

Type 1. Content of an utterance can be challenged by asserting an incompatible proposition, **or** by asserting a proposition which is incompatible with a precondition or a consequence of the proposition. The latter strategy assumes the interlocutor will endorse the relevant inference as well as the content of the challenge. Defending the content of a propositional commitment may involve appeal to observations or to more "basic" commitments; both of these may of course be open to further challenge.

Type 2. Reliability (truthfulness) is claimed for the speaker and for the source of any commitments which are inherited by testimony. Reliability can be challenged by e.g. instancing occasions when the speaker has (wittingly or not) uttered falsehoods, by questioning their qualifications or by raising doubts over "normal input-output conditions" in Searle's sense. For example, "you couldn't have seen that, it was too dark/you're near-sighted ..." etc.

Type 3. Status: utterances may depend for their appropriateness on the speaker's social role: when an invigilator announces the start of an exam, or a football referee blows the final whistle, they are not simply signalling a state of affairs but bringing that state of affairs into existence. The status in question may be classified as *institutional* (deriving from an individual's formally-recognised role or position) or *conventional*.

In practice there can be some overlap between these categories: for instance "Trust me, I'm a doctor" can be glossed as either "My formal training and experience equip me to make reliable judgments" (Type 2) or "My professional status exempts me from scrutiny by laymen" (Type 3).

Parsons et al [13] offer a distinction between *credulous*, *cautious* and *skeptical* agents which ultimately relies on a stipulated preordering over knowledge bases in terms of *degrees of belief*. An analogous distinction might be made in terms of commitment and entitlement:

- a *credulous* agent will grant entitlement to commitments by default;
- a *cautious* agent will grant entitlement to commitments if not already committed to an incompatible claim;
- a *skeptical* agent will grant entitlement to a commitment only if it survives justified (entitled) challenges to entitlement.

Additionally, we may distinguish three orthogonal stances associated with the validity claims:

- a *rational-empiricist* agent will endorse only claims which the speaker is entitled to on grounds of *content*;
- a *social* agent will endorse claims which the speaker is entitled to on grounds of *reliability*;
- a *deferential* agent will endorse claims which the speaker is entitled to on grounds of *status*.

However, by contrast with [13] the framework outlined here inherits a weakness of [2, 3] in that there is no satisfactory account of degrees of belief.

3 Deontic scorekeeping

Following [2, pp. 185-6, 190-1] I assume that dialogue participants maintain a *deontic scoreboard* of the commitments and entitlements which each participant undertakes and discharges. In any multi-agent interaction, each agent A_n maintains a set of commitments for each agent A_i as follows:

- $C_{Ack}(A_i)$ Commitments A_i acknowledges
- $C_{Attr}(A_i)$ Commitments A_n attributes to A_i
- $E_{Cl}(A_i)$ Entitlements A_i claims
- $E_{Attr}(A_i)$ Entitlements A_n attributes to A_i

1. Commitments can be classified into *practical* (commitments to act, corresponding to *intentions* in mentalistic accounts) and *doxastic* (commitments to justify an assertion, corresponding to *beliefs*). In this paper we are mostly concerned with the latter.
2. Agents $(A_1, \ldots A_{i-1}, A_{i+1}, \ldots A_n)$ may have different views of A_i's commitments: some of them may have missed, misheard or misconstrued one or more of A_i's utterances. Searle's criterion that "normal input-output conditions obtain" [14] is an idealising presupposition which cannot be relied on in natural dialogue.
3. $C_{Ack}(A_i)$ may well be a proper subset of $C_{Attr}(A_i)$: commitment stores are not assumed to be closed under any notion of consequence, as it would be unrealistic to assume that agents will overtly acknowledge all the inferential consequences of those propositions to which they do acknowledge commitment, and interlocutors will differ in the extent to which they are able or care to work out the consequences of A_i's commitments. This approach avoids issues of *logical omniscience*.
4. On the other hand $E_{Cl}(A_i)$ may be a superset of $E_{Attr}(A_i)$ if A_n is disposed to dispute A_i's claims. A_i may seek to resolve the difference by challenging entitlements or the two may agree to disagree.
5. As Brandom [2, p. 196] observes: from time to time we undertake incompatible *practical* commitments, and it is equally possible to undertake and acknowledge incompatible *doxastic* commitments and for others to attribute them to us without risking incoherence. However, we cannot be *entitled* to incompatible commitments.
6. To *endorse* another agent's commitment is to adopt the commitment oneself and undertake to defend it against further challenges: it is effectively to acknowledge the asserted proposition as *true*.
7. The scoreboard must also include a **history** of challenges and justifications which may be consulted to resolve disputes over incompatible commitments.

[2] These headings cut across the three types of *grounding* (experiential, formal and social) proposed by [17].

The functions of natural argumentation thus include:

- Requiring agents to disclose and justify why they claim entitlement to assertions;
- Demonstrating the unacknowledged consequences that follow inferentially from acknowledged commitments;
- Inducing agents to abandon incompatible commitments and unjustified entitlement claims.

According to this framework **communication** is reflected in inter-speaker inheritance of commitments.

4 Argumentation patterns

The framework will include specifications for the following proto-speech acts among others:

assert: undertake commitment to justify a propositional claim
endorse: ascribe entitlement to a commitment, and undertake the commitment oneself
instruct: bestow a practical commitment on another agent
challenge: require agent to justify or abandon a commitment
respond to a challenge
retract an entitlement claim or commitment

Challenges can be informally specified as

challenge(Ty, Co, CCo) where

Ty: the type of challenge (1 | 2 | 3)
Co: the commitment(s) challenged (if unspecified, defaults to the most recent or salient)
CCo: a counter-commitment incompatible with Co

Any of the arguments may be left unspecified; some examples are

- *Why?* or *What?*
  challenge(-, -, -)
- *What gives you the right to say that?*
  challenge(3, Co, -)
- *You're wrong; Marx died in 1883, not 1893*
  challenge(1, Co , CCo)

Responses to challenges may consist of a *justification* for the challenged commitment, or a counter-challenge questioning the interlocutor's entitlement to the challenge:

respond(Ty, Ch, (Just | CCh))

Note that the value of Ty need not match the challenge: e.g. a challenge on the grounds of **content** may be met by a justification on the grounds of **reliability**: *I'm a pretty straight kind of guy; Our sources are impeccable,* etc.

4.1 Challenges and discourse structure

In this section I sketch how the framework can be used to reconstruct some standard accounts of persuasive dialogue [16] and monologue [11].

Example

(a) A: Take an umbrella. B: *Why?* A: It's going to rain.
A: instruct(ϕ); B: challenge(-, ϕ, -);
A: respond(1, challenge(-, ϕ, -), assert(ψ))

(b) A: You should take an umbrella. It's going to rain.
A: instruct(ϕ); assert(ψ)

(c) A: It's going to rain. You should take an umbrella.
A: assert(ψ); instruct(ϕ)

In the above scenario, suppose A has the goal that B undertake a practical commitment to carry an umbrella. Examples (a - c) illustrate three different strategies:

- Issue a bare instruction; offer justification only if challenged.
- Issue an instruction, followed by an assertion that **pre-empts** a potential challenge.
- **Obviate** the challenge by uttering the justification **before** the instruction.

Argumentation schemes Walton has proposed an inventory of argumentation schemes such as the **argument from position to know** (see e.g., [16]) which is discussed below:

Argument from Position to Know (Version 1)

Major Premise: Source **a** is in a position to know about things in a certain subject domain **S** containing proposition **A**.

Minor Premise: **a** asserts that **A** (in domain **S**) is true (false).

Conclusion: **A** is true (false).

Critical Questions

CQ1: Is **a** in a position to know whether **A** is true (false)?

CQ2: Is **a** an honest (trustworthy, reliable) source?

CQ3: Did **a** assert that **A** is true (false)?

This argument can be employed *in toto* on grounds of **rightness** to justify putting a question about **A** to **a**, or to respond to a challenge to **A** on the grounds of **truthfulness**. The proponent defends against the challenge by ascribing commitment to **A** to source **a** and so claims entitlement to commit to **A** by appeal to **testimony**. Alternatively, the minor premise may be offered as an initial response with the major premise held in reserve for a supplementary challenge. The critical questions represent different ways of challenging the major and minor premises: CQ1 and CQ2 challenge the major premise on grounds of truth and truthfulness respectively; CQ3 challenges the minor premise on grounds of truth. This list of questions does not exhaust the possibilities: the proponent's ascription of **A** to **a**, and of expertise in **S** could also be challenged on grounds of truthfulness. The proponent might not be sincere in claiming that **a** said **A**, or might not be qualified to assess **a**'s specialist knowledge.

Rhetorical structure The argumentation protocol sketched above can accommodate *persuasive monologue* if this is modelled as a dialogue with a "silent partner": the author anticipates possible challenges and generates appropriate responses. In natural argumentation both pre-empt and obviate strategies may be employed, and both can be applied *recursively* since justifications are also open to challenge.

Challenges may themselves be accompanied with material that pre-empts potential counter-challenges. This gives rise to a hierarchical discourse structure exhibiting RST [11] relations such as Motivate, Justify and Evidence (see [10] for a similar account using the framework of Update Semantics). According to RST, discourses can be analysed as tree structures where adjacent nodes are classified as Nucleus or Satellite depending on how central they are to the author's purpose. In the framework of this paper, Satellites can often be seen as pre-empting or obviating potential challenges to the Nucleus. The "umbrella" example illustrates the Motivate relation; the assertion *It's going to rain* could be defended on reliability grounds by citing the weather forecast (Evidence) or claiming personal expertise (Justify).

4.2 Complexity

Complexity considerations affect both speaker (S) and hearer (H) roles. Suppose S has the goal H that H endorse ϕ, S has to consider various scenarios, including but not limited to:

- H will endorse ϕ out of the blue
- H will endorse $\psi; \phi$ where ψ entitles commitment to ϕ and H is entitled to commit to ψ. (This may apply recursively, supposing H requires justification for ψ.)

H's options on hearing ϕ are: endorse; challenge; defer evaluation in case a justification is subsequently offered. Deferring evaluation involves augmenting the commitment store with a **stack** mechanism. The complexity costs can be ranked as follows:

- For S, the lowest complexity comes from uttering ϕ on its own, though this can turn out as more costly if there is a challenge. The highest complexity comes from *planning* to utter ψ before ϕ, since $\mathsf{utter}(\phi)$ has to be placed on a goal stack. The intermediate case is: $\mathsf{utter}(\phi)$; test for potential challenge; $\mathsf{utter}\psi$.
- For H, the least cost comes from $\psi; \phi$ as this does not require the *stack* mechanism for deferring ϕ, and ϕ will only be evaluated once.

Thus there is a conflict between S's and H's complexity costs; complexity of an utterance cannot be evaluated independently of speakers' and hearers' perspectives.

For persuasive monologue the situation is more difficult in that the author has to judge how much pre-emptive defence will be appropriate in order to anticipate objections from a typical reader.

5 Conclusions and future developments

Space has permitted only rather simple examples to be presented. I have concentrated on expounding some underlying ideas at the expense of formal rigour, and on decribing the data structures making up the "deontic scoreboard" rather than explaining how they are processed. I have outlined some techniques for modelling dialogue in terms of *commitments* and *entitlements, challenges* and *justification* according to *content, reliability* and *status*, and indicated some complexity issues. Taken together, these ingredients provide for fine-grained modelling of a variety of patterns and styles of argumentation including rhetorically complex utterances. Future research will concentrate on formalising and operationalising the notions listed above, modelling how higher-level rhetorical and argumentation patterns emerge from the interaction of these fundamental elements and in particular how complexity issues influence interlocutors' choice of strategy. An important development will be to incorporate a mechanism for degrees of belief (commitment).

The framework is largely modelled on Brandom's system [2, 3] supplemented with ideas from Habermas [5, 6, 7] and has somewhat glossed over significant differences between these two philosophers as well as the important critique of the latter in [9]. I intend to continue to explore the relevance of this body of work for multi-agent design and dialogue modelling, taking full account of other applications of Habermasian ideas such as [15, 12].

Acknowledgements

I am grateful to the anonymous reviewers for their helpful and pertinent comments.

REFERENCES

[1] J. L. Austin, *How to do things with words*, 1962.
[2] Robert Brandom, *Making it Explicit*, 1994.
[3] Robert Brandom, *Articulating Reasons: An Introduction to Inferentialism*, 2000.
[4] Philip Cohen and Raymond Perrault, 'Elements of a plan-based theory of speech acts', *Cognitive Science*, (1979).
[5] Jürgen Habermas, 'Intentionalistische Semantik (1975/6)', in *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, (1984).
[6] Jürgen Habermas, 'Was heißt Universalpragmatik? (1976)', in *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, (1984).
[7] Jürgen Habermas, 'Some further clarifications of the concept of communicative rationality', in *On the Pragmatics of Communication*, ed., Maeve Cooke, (1998).
[8] Charles Hamblin, *Fallacies*, 1970.
[9] Joseph Heath, *Communicative Action and Rational Choice*, 2003.
[10] Rodger Kibble, 'Inducing rhetorical structure via nested update semantics', in *Proceedings of the Fourth International Workshop on Computational Semantics*, University of Tilburg, The Netherlands, (2001).
[11] William C. Mann and Sandra A. Thompson, 'Rhetorical structure theory: A theory of text organization', Technical report, Marina del Rey, CA: Information Sciences Institute, (1987).
[12] Peter McBurney and Simon Parsons, 'Engineering democracy in open agent systems', in *Engineering Societies in the Agents World (ESAW-2003)*, (2003).
[13] Simon Parsons and Peter McBurney, 'Argumentation-based communication between agents', in *Communication in Multiagent Systems*, ed., Marc-Philippe Huget, (2003).
[14] John Searle, 'What is a speech act?', in *Philosophy in America*, ed., M. Black, (1965).
[15] Munindar Singh, 'A social semantics for agent communication languages', in *Proc. IJCAI'99 Workshop on Agent Communication Languages*, pp. 75–88, (1999).
[16] Doug Walton and Chris Reed, 'Argumentation schemes and defeasible inferences', in *Procs of Workshop on Computational Models of Natural Argument*, (2002).
[17] Terry Winograd and Fernando Flores, *Understanding Computers and Cognition*, 1986.
[18] Michael Wooldridge, 'Semantic issues in the verification of agent communication languages', *Journal of Autonomous Agents and Multi-Agent Systems. 3(1)*, 9–31, (2000).