To appear in the Online proceedings of the AISB50 Convention, Goldsmith's College London, April 2014

Towards a Computational Account of Inferentialist Meaning

Paul Piwek[1]

Abstract. Both in formal and computational natural language semantics, the classical correspondence view of meaning – and, more specifically, the view that the meaning of a declarative sentence coincides with its truth conditions – is widely held. Truth (in the world or a situation) plays the role of the given, and meaning is analysed in terms of it. Both language and the world feature in this perspective on meaning, but language users are conspicuously absent. In contrast, the inferentialist semantics that Robert Brandom proposes in his magisterial book 'Making It Explicit' puts the language user centre stage. According to his theory of meaning, the utterance of a sentence is meaningful in as far as it is a move by a language user in a game of giving and asking for reasons (with reasons underwritten by a notion of good inferences). In this paper, I propose a proof-theoretic formalisation of the game of giving and asking for reasons that lends itself to computer implementation. In the current proposal, I flesh out an account of defeasible inferences, a variety of inferences which play a pivotal role in ordinary (and scientific) language use.

1 INTRODUCTION

Formal semantics emerged as a field in the wake of the seminal work by Richard Montague [22]. Montague developed a formal semantics for fragments of English, where his point of departure was the classical truth-conditional view of meaning: meaning as analysed in terms of truth (in a model), without further analysis of the concept of truth itself.

Though Montague's original framework was modified and refined in numerous ways, its truth-conditional foundation has remained largely unshaken. Witness, for example, Kamp & Reyle's 'From discourse to logic':

> 'Since truth and falsity are of such paramount importance, and since it is in virtue of their meaning that thoughts and utterances can be distinguished into those that are true and those that are false, it is natural to see the world-directed, truth-value determining aspect of meaning as central; and, consequently, to see it as one of the central obligations of a theory of meaning to explain how meaning manifests itself in the determination of truth and falsity.' Page 11 of Kamp & Reyle [15]

In 'From discourse to logic', Hans Kamp (who studied with Montague) and Uwe Reyle present Discourse Representation Theory (DRT), a detailed account of natural language interpretation that describes how a representation of the meaning of a discourse can be computed incrementally from the contributions of individual sentences. This took formal semantics beyond the boundary of the sentence into the domain of extended discourse and intersentential anaphora. Each sentence is viewed as giving rise to an update of the meaning representation of the discourse so far. As a result, the field of formal semantics took a dynamic turn. The truth-conditional foundations were however retained. For instance, in DRT, the meaning of the representation of a discourse is essentially truth-conditional – it is given in terms of an embedding relation between that representation and a model.[2]

The truth-conditional approach has been very fruitful. Over the past fourty years, it has led to numerous insights into natural language semantics, ranging from the interpretation of intra- and intersentential anaphora, tense, aspect and intensional contexts to plurality and generalised quantifiers. It has also been developed further in some of the most advanced work in computational semantics (e.g., [4]). Though the dynamic turn has blurred the boundaries between semantics and pragmatics, on the whole, the traditional separation of linguistic studies into syntax, semantics and pragmatics is still in force. The dynamic turn did stimulate studies into construction of semantic representations with reference to language users, but the semantic import of these representations themselves is still analysed in terms of a correspondence relation between the representation and models. In other words, the way the meaningfulness of these representations is accounted for still leaves the language user out of the picture.

From the point of view of Artificial Intelligence (and, more generally, the Cognitive Sciences), an account of meaning that has little to say about *how* the capacity to produce meaningful utterances is part of a language user's capacities for interacting with others and the world (through communication, perception and action) is rather unsatisfactory. From this point of view, the work by the philosopher Robert Brandom provides a promising alternative theory of meaning. Brandom [6] puts forward an analysis of meaning in terms of a game of giving and asking for reasons. In this game, the language user is central: sentences acquire meaning by virtue of their use in such a game. From the point of view of the programme of Artificial Intelligence, this game provides a way to link meaning with an agent's capacities for inference, action and perception.

In this paper I aim to show how playing a game of giving and asking for reasons involves the language user's capacity for inference,

[1] The Open University, United Kingdom, email: paul.piwek@open.ac.uk

[2] The algorithmic sentence-by-sentence construction of representations of discourse meaning has been taken to a new level in Dynamic Syntax (DS) [16], which models the construction of these representations as an incremental left-to-right *word-by-word* process. However, also in Dynamic Syntax, the emphasis is primarily on how the representations are build from linguistic input and less on how these representations acquire semantic import. The current proposal can be thought of as a step towards a non-truthconditional inferentialist semantics for incremental theories of interpretation such as DRT and DS.

perception and action. Each of these capacities is taken to be computational in nature. Thus we arrive at a theory of meaning that is decidedly computational - eschewing reference to non-computational notions such as truth in a model.

The remainder of this paper is organised as follows. In the next section, I examine two puzzles associated with the truth-conditional perspective. These puzzles emerge when we ask how language users deploy meaning in perception and conversation. I provide a sketch of how each of the two puzzles can be solved whilst sticking with a truth-conditional concept of meaning. I then, however, proceed to show how these solutions (especially when they are taken jointly) give rise to further problems. I draw the conclusion that exploration of an alternative non-truth-conditional account of meaning is warranted.

In Section 3, I present the outline of such an account based on Brandom's work. Though it is inspired by Brandom's proposals, it is not intended as a complete or even literalist exegesis: I only cover a subset of Brandom's intricate system, and there are points where I deviate from the original. After briefly indicating which ingredients of Brandom's work I take to be of principal importance, I also draw attention to some differences between the current proposal and that of Kibble [17].

Section 4 presents my attempt to take some of Brandom's key insights and turn them into a computational theory of meaning. The game of giving and asking for reasons is formalised and shown to be playable by an agent with certain (computationally grounded) capacities for inference, action and perception. A pivotal role is played by a proof-theoretic account of defeasible inference. This account allows us to define the interface between reasoning on the one hand and action and perception on the other, without falling into the empiricist trap – i.e., what Sellars dubs the myth of the given [31]: the myth that observation sentences are meaningful outside of the rich web of inferential connections between the sentences of a language.

Some aspect of the formalisation are then illustrated in Section 5. There I analyse certain features of a short dialogue fragment in terms of the game of giving and asking for reasons.

Finally, Section 6 reviews the current proposal and points forward to further work.

2 PUZZLES AND PROBLEMS

2.1 The Puzzle of the Use of Meaning in Perception

Firstly, there is *the puzzle of the practical use of meaning in perception*. Let us assume, for the moment, that the truth-conditional account of meaning is correct. Let us also assume that understanding a sentence amounts to grasping its meaning, i.e. in this case its truth conditions. We will explore how, on these assumptions, a language user can deploy the meaning of a sentence to make judgements about its truth in a concrete situations.

Take for instance the sentence 'John has measles'. Grasping its meaning would involve the ability to distinguish between situations in which 'John has measles' is true and those in which it is false (additionally, as in Situation Semantics [2], we may also need to allow for the sentence to be neither true nor false in those situations where John isn't present). Now, human language users can't apply meanings directly to situations which they find themselves in, if we conceive of meaning in this way. In practice a language user isn't always able to tell correctly whether the situation itself is one in which 'John has measles'. If they could do so, they would be perfect truth tellers. Sometimes they get deceived by appearances, even when they have taken all the available evidence on board. For example, they may decide that John has measles because he shows all the symptoms of measles including what appears to be a rash. Their conclusion may nevertheless be wrong, because the 'rash' really is only the result of liberally applying a red marker pen.[3]

The lesson I draw from this example is that in practice a language user will never be entirely certain that they have looked at a situation carefully enough or in sufficient detail. So, they are never quite sure which situation out of many possible ones, they find themselves in. Therefore, they are never able to apply meanings (conceived as truth conditions) directly to the situation at hand. Clearly, we need a more elaborate story to explain how meaning as truth conditions is applied in perception.

One possible story goes as follows. We need to relax the idea that in perception a language user tests whether the meaning of a sentence applies to the actual situation at hand. Rather, they somehow make an informed guess about which class of situations they currently find themselves in and then check whether these are a subset of the set of situations in which the sentence is true (i.e., they compare it against the meaning of the sentence). This shifts the problem to how we, as finite human beings, can somehow store such large, possibly infinite, sets of situations and compare them with each other in a reasonable amount of time. It also doesn't give us a mechanism for making informed guesses about the class of situations that we find ourselves in. Part of the purpose of this paper is to show that once we take inference rather than truth as our primitive, a solution to this puzzle is available.

2.2 The Puzzle of Truth-conditionally Equivalent Sentences

Secondly, there is *the puzzle of truth-conditionally equivalent sentences*. This one is particularly grating in the case of mathematical knowledge. If one subscribes to the view that true mathematical statements are true in every possible situation or world, then the truth-conditional conception of meaning entails that any two true mathematical statements have the same meaning. So, Gödel's incompleteness theorems mean the same as, for instance, $1+1 = 2$. This doesn't seem to sit well with our everyday use of the notion of meaning.[4] If the meanings are the same, how can we nevertheless avoid the undesirable consequence that when they are put to practical use they become indistinguishable? A possible answer lies in the observation that although two statements may have the same meaning, it may not be trivial to establish this: whether two mathematical functions (representing truth conditions) are one and the same may in practice not be decidable (it would involve comparing a potentially infinite number of input–output pairs).

The *puzzle of truth-conditionally equivalent sentences* surfaces also for statements about the observable world. Consider the following variation on Frege's puzzle. The statements 'Achilles looks

[3] One may argue that they've simply not investigated the situation carefully enough. They should have checked that the red marks on John's body were indeed part of a rash, rather than ink marks. But now consider the following situation. John really has a rash but rather deviously: 1) concealed the red patches with mascara and 2) used a felt tip to insert red dots in the clear areas. Having discounted the observable red marks, our language user may conclude that John doesn't have rash. But then, some further investigation would tell them that the rash is hidden and John has measles after all. In principle, there is no end to the possibility of such double, triple, quadruple, and so on, deceptions.

[4] Claiming that one understands the meaning of $1 + 1 = 2$ doesn't commit one to also understanding the meaning of Gödel's first (or second) incompleteness theorem.

at the morning star' and 'Achilles looks at the evening star' are true in exactly the same situations, since 'the morning star' and 'the evening star' both refer to Venus. Therefore, according to the truth-conditional account, these statements mean the same thing. Now, the ancient Greeks didn't know that both names refer to Venus. An ancient Greek wouldn't affirm that 'Achilles looks at the morning star' follows from 'Achilles looks at the evening star'. On the truth-conditional account they should however do so. Somehow the ancient Greeks seem to have only had partial knowledge of the meaning of each of these sentences. When we learned that both 'the morning star' and 'the evening star' refer to Venus, we somehow extended our partial knowledge of the meaning. Such an answer suggest that we need to look for a notion of partial knowledge of truth-conditional meanings. If we formalise meaning as a function from situations to truth values, this suggest that partial knowledge involves restricting the set of situations in which the ancient Greeks were able to apply the meanings (technically, we narrow the domain of the function, e.g. such that they can only apply one of the sentences in the morning and the other in the evening).

2.3 From Puzzles to Problems

The second puzzle suggests that we need, in addition to the notion of truth-conditional meaning, a further notion of partial knowledge of meanings. Once we make the move to (partial) knowledge of meaning, one can ask whether truth-conditional meaning itself, as a opposed to any partial knowledge of it, is required at all. Is it possible to let go of fully fledged truth-conditional meanings in favour of partial representations of meanings, that are not dependent on some underlying complete, as opposed to partially known, meaning? Is it perhaps reasonable to apply Occam's razor and get rid of these complete meanings? There is some evidence that this is not such a bad move. If we stick to the idea that complete truth-conditional meaning is the foundation for any notion of meaning or partial knowledge of meaning, we rule out making sense of the verbal practices that involve concepts, such as phlogiston, that at a later date turned out not to refer at all. We now know in 2014 that in all possible situations, the sentence 'phlogiston is present' is false. Does it follow that therefore its meaning has been void all along (and speaking of partial knowledge of that meaning is meaningless as well)? If so, how could this meaning have governed the use of the term prior to the renunciation of the term? Were the interlocutors that used this term prior to its rejection, merely making meaningless noises?

Let me outline one further argument. Gödel's incompleteness results have been used in support of the claim that the human mind cannot be adequately modelled as an algorithm or formal system [18]. Gödel sentences demonstrate that, given any sufficiently complex formal system, there will be sentences about the system which are true but not provable within the system. The unrestricted truth-conditional view of meaning seems to however endow us with an absolute ability to determine whether a sentence is true, once we have grasped its (complete) meaning. I would like to suggest that we can reverse the argument and use this powerful property of truth-conditional meaning to argue against it. Is it really sensible to assume that we, as human beings, can decide for any sentence (even about ourselves) whether it is true, once we have grasped its meaning? Is it really impossible that each of us has their own Gödel sentences. In this case, the truth-conditional view of meaning may incorrectly endow us with powers that it is impossible for us to have.

In this section, I have drawn attention to three problems with the truth-conditional view of meaning. Firstly, we need to tell a supplementary story to show how this approach can be made to address the puzzles about perception and equivalence: there is a gap between the idea of meaning as function from situations to truth values, and the everyday use of meanings in perception and conversation. Secondly, such a story seems to introduce a notion of partial knowledge or representations of meaning. Once this move is made, it is however not evident why we can't start from such partial knowledge or representations in the first place, rather than retain the practically inactive truth-conditional meanings. Thirdly, though Gödel's incompleteness results are normally used to argue that human minds aren't mere computers, one can just as well argue that, on the assumption that human minds are computers, the notion of truth-conditional meaning doesn't fit with the sort of meanings that we actually are able to grasp and deploy.

3 SKETCH OF AN INFERENTIALIST APPROACH

We have seen that several puzzles and problems emerge from the assumption that meaning equals truth conditions. This not to say that there is no way to solve these puzzles and problems satisfactorily whilst sticking to the truth-conditional view. We could try to formulate a notion of knowledge of the truth-conditional meaning of a sentence, which allows for partial knowledge of meanings and introduces limits to computations that are afforded by this knowledge. An alternative strategy, which we explore here is to start with an inferentialist account of meaning that acknowledges from the outset the computational nature of meaning.

The inferentialist conception of meaning goes back to Gentzen's work on natural deduction [11]. Gentzen suggests that in natural deduction introduction rules for logical connectives define the meaning of these connectives. Additionally, it also has roots in the Brouwer–Heyting–Kolmogorov explication of intuitionistic truth in terms of proof construction. The use of natural deduction to get a handle on meaning has been advanced, most prominently, by Prawitz (e.g. [28]) and Dummett (e.g. [10]). The Brouwer–Heyting–Kolmogorov idea of proof construction has culminated, via the Curry–Howard–de Bruijn correspondence between logic and type theory, in Martin-Löf's Constructive Type Theory [19].

The main focus of all the aforementioned efforts has been on modelling the meaning of mathematical statements, though there have also been some recent efforts to apply Martin-Löf style semantics to natural language (e.g. Ranta [29], Ahn [1], Piwek & Krahmer [27]). The main pay-off of this strand of inferentialism has been a finer-grained notion of meaning in terms of proof rather than truth conditions. This allows one to address, among other things, the truth-conditional equivalence problem (as described above). It also also provides the means for a powerful analysis of anaphora and presupposition (that, arguably, improves on Discourse Representation Theory [15]).

There are however some limitations to this strand of work. Firstly, its primarily mathematical orientation means that defeasible inferences, which are rife in everyday and also scientific reasoning, are glossed over (but see the proposal on pages 20–26 of Piwek [23], which we develop further in Section 4.9). Secondly, the intuitionistic foundations of this work have given it the reputation of being revisionist in orientation (aiming to critique and revise existing linguistic practices, rather than model those practices as we find them). Thirdly, the idea that meaning corresponds with proof conditions (as in the work of Martin-Löf), is one-sided (and reminiscent of the earlier verificationis programme) – Dummett has, however, drawn attention to this issue and has suggested that the meaning of a sentence consists of two components: its justifications (i.e. proofs) and its consequences. Finally, proof-theoretic work on meaning has concentrated on rules for the meaning of the logical connectives. The meaning of non-logical vocabulary is left unanalysed and its inferential potential is only explored when its inferential relations are made explicit using logical vocabulary. This suggest a dependence of non-logical vocabulary on the logical vocabulary. The proposal that follows rejects the idea that logical vocabulary is needed to specify a semantics for non-logical vocabulary.

A comprehensive inferentialist alternative is offered by Brandom. A condensed description of this alternative can be found in Brandom's [7] and also Wanderer [33], whereas the full account is laid out by Brandom in [6]. What follows is a summary of the key tenets.

Brandom brings language users into the picture from the start, asking the question: What makes us treat a person's utterances as meaningful? Brandom's answer is that we ascribe meaning provided that we can understand this person's actions as moves in a certain kind of dialogue game. This approach is reminiscent of Dennett's intentional stance and Turing's test for intelligence: we are not required to examine the inner life of the meaning maker, merely whether their behaviour can be profitably described in a certain way. However, in contrast with Turing, who proposed a practical test for intelligence, Brandom is interested in giving a theoretical account. This account consists of a specification of the dialogue game in question. This game is referred to as the game of giving and asking for reasons. Its description is normative, i.e. in terms of rules for how to correctly play the game – Brandom does not attempt a reduction of the game to non-normative language and seems to be non-committal about whether such a reduction is possible.

We briefly outline the game's key features, which we formalise in the next section. This is to be understood as our interpretation of the game, rather than a entirely faithful rendering of Brandom's proposal. In particular, we will gloss over Brandom's distinction between endorsement and attribution.

- Participants can make one of five moves: 1) assert a sentence, 2) challenge a sentence, 3) retract a sentence, 4) make an observation, and 5) perform an action.
- For each participant, we keep track of the sentences they have asserted, i.e. their non-inferential commitments. We call this the participant's commitment store.
- A participant is consequentially committed to a sentence if it follows from their commitments via commitment preserving inferences.
- A participant is entitled to sentence if the entitlement to the sentence is not blocked through an incompatibility inference and a) the sentence hasn't been challenged or b) a challenge is addressed by: 1) asserting another sentence, which the participant is entitled to, and from which the earlier sentence follows, or 2) asserting that the sentence is inherited from the entitlements of another participant.[5]
- A sentence follows from another one if it follows through a single or combination of any of the entitlement preserving, commitment preserving and incompatibility inferences.
- An incompatibility inference is an inference where a commitment to a sentence prohibits entitlement to another sentence. Entitlement preserving inferences are defeasible: an entitlement preserving inference can be blocked by an incompatibility inference.
- Observations can entitle one to certain sentences and sentential entitlements can entitle one to certain actions via dialogue game entry and exit rules, respectively.
- Entitlements can be inherited from other participants. In particular, one can justify entitlement to a claim by referring the challenger to someone else who made that claim, relying on that person's ability to furnish the reasons, if challenged.[6]
- All sentences in the most basic version of this type of game are non-logical, i.e. free from logical vocabulary. Extending the game with logical vocabulary and appropriate inferences is viewed as a adding a new layer to a game. This layer allows one to play a game in which the inferences of the original game can be expressed explicitly. Brandom refers to this conception of logical vocabulary as logical expressivism. He generalises this in [8] where elaborates on the idea that one vocabulary can make explicit the practices in which another vocabulary is used.

4 FORMALISATION OF THE INFERENTIALIST APPROACH

The proposal in this section continues a long-running programme to develop a formal inferentialist semantics as a viable alternative to traditional truth-conditional semantics (Piwek [23]; Piwek & Krahmer [27]; Piwek [25]; Piwek [26]). It engages with the proof-theoretic tradition, in contrast with Brandom and Aker's work on formalising his insights.

One of the few other formalisations of (part of) Brandom's work can be found in Kibble [17]. There are a number of ways in which the current proposal differs from Kibble's. The main one is that Kibble does not elaborate on how interlocutors draw or compute inferences, including defeasible ones. Also, in contrast with Kibble, the current proposal defines a game that lacks logical vocabulary (such as $\rightarrow$ and $\neg$). In the current game, background knowledge about meaning is modelled in terms of inference rules, rather than explicit (logically complex) formulae in the communication language. This is more in the spirit of Brandom's layer cake model of language games.[7]

4.1 Commitment Stores and Commitments

Each interlocutor X has a positive and negative commitment store: Γ_X^+ and Γ_X^-. New sentences are introduced by X into these commitment stores through assertion and denial dialogue acts, respectively

[5] The key idea here is that one is only entitled to a commitment in so far as one has defended it against any challenges. Challenging another person's commitments is one way in which the meaning of sentences (in terms of its inferential relations with other sentences) is put to work, and sustained. I would like to point out that this view is foreshadowed in Mill's defence of the freedom of speech [21], where he argues that an uncontested but true common opinion is in danger of losing its very meaning; see, in particular, the summary of his argument at the end of Chapter II: 'Thirdly, even if the received opinion be not only true, but the whole truth; unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or feeling of its rational grounds. And not only this, but, fourthly, the meaning of the doctrine itself will be in danger of being lost, or enfeebled, and deprived of its vital effect on the character and conduct: the dogma becoming a mere formal profession, inefficacious for good, but cumbering the ground, and preventing the growth of any real and heartfelt conviction, from reason or personal experience.'

[6] The formalisation of inheritance of entitlement is beyond the scope of the current paper.

[7] A further difference is that Kibble includes a mechanism for getting participants to agree on commitments, i.e. to build up a common ground. This takes us beyond the game of giving and asking for reasons as specified by Brandom and, in our view, is orthogonal to the concerns of Brandom's game.

(Section 4.7). We use Γ as a shorthand for $\langle\Gamma^+, \Gamma^-\rangle$, with subscripts for agents omitted where this doesn't result in ambiguity.

The (positive/negative) commitments of an interlocutor consist of those commitments that have been directly asserted or denied and also any commitments that follow (monotonically) from these. This results in the following definition:

(Positive and negative commitment) An interlocutor's positive/negative commitments consist of those sentences that are monotonically affirmed/refuted by their commitment store $\langle\Gamma^+, \Gamma^-\rangle$.

Affirmation and refutation are defined below (Section 4.4).

4.2 Challenges

Each interlocutor X has a set of positive and negative challenges: Ξ^+_X and Ξ^-_X. These consist of sets of sentence whose assertion or denial by X has been challenged.

4.3 Open Challenges

Each interlocutor X has a set of positive and negative open challenges: Ω^+_X and Ω^-_X. These consist of sentences ψ such that their assertion or denial has been challenged, i.e. a reason has been asked for the asserted or denied sentence *and* ψ is not non-monotonically affirmed ('$\vdash_{NM}$') or refuted ('$\dashv_{NM}$') by the commitments of X (after ψ itself has been removed – ψ does not count as a reason for itself). The notions of non-monotonic affirmation and refutation are defined in Section 4.10.

(Positive open challenges) $\psi \in \Omega^+_X$ if and only if (a) $\psi \in \Xi^+_X$, (b) $\psi \in \Gamma^+_X$ and (c) not $\langle\Gamma^+_X - \{\psi\}, \Gamma^-_X\rangle_{NM} \vdash \psi$

(Negative open challenges) $\psi \in \Omega^-_X$ if and only if (a) $\psi \in \Xi^-_X$, (b) $\psi \in \Gamma^-_X$ and (c) not $\langle\Gamma^+_X, \Gamma^-_X - \{\psi\}\rangle \dashv_{NM} \psi$

Open challenges can be addressed in two ways: one can provide a reason for the challenged sentence (by asserting or denying another sentence), but it is also possible to retract the sentence. The dialogue acts for doing so are specified in Section 4.7.

4.4 Judgements: Affirmations and Refutations

Given a set of positive and negative commitments $\Gamma = \langle\Gamma^+, \Gamma^-\rangle$, we write:

(Affirmation) $\Gamma \vdash \phi$ for ϕ is monotonically affirmed by Γ.

(Refutation) $\Gamma \dashv \phi$ for ϕ is monotonically refuted by Γ

We define judgements in terms of affirmations and refutations:

(Judgement) A judgement is either an affirmation or a refutation.

The inferential role of mononotic affirmation and refutation is defined in Section 4.9.

4.5 Inconsistency

The commitment store $\Gamma = \langle\Gamma^+, \Gamma^-\rangle$ is inconsistent if there is a sentence ϕ that is *both* affirmed (signified by '$\Gamma \vdash \phi$') and refuted monotonically (signified by '$\Gamma \dashv \phi$').

(Inconsistency) $\langle\Gamma^+, \Gamma^-\rangle$ is inconsistent if and only if for some ϕ: $\langle\Gamma^+, \Gamma^-\rangle \vdash \phi$ and $\langle\Gamma^+, \Gamma^-\rangle \dashv \phi$

4.6 Entitlements

(Entitlement) Interlocutor X is positively/negatively entitled to ϕ *iff* ϕ is non-monotonically affirmed/refuted by Γ_X and for any open challenge ψ, the sentence ϕ is non-monotonically affirmed or refuted by $\Gamma_X - \{\psi\}$.

In words, one is only entitled to those sentences that follow non-monotonically from one's commitments and which will still hold, even if one needs to retract sentences that are currently open challenges. We presuppose that Γ_X is consistent. Note also that, if ϕ is monotonically affirmed/refuted by Γ, then by definition (see Section 4.10) it is also non-monotonically affirmed/refuted.

4.7 Dialogue Acts

We assume that a contribution to the dialogue game is the utterance of a sentence ϕ with a certain force by a speaker (S) to an addressee (A). Contributions are mapped to one of four types of dialogue acts: assertion, denial, asking for a reason (of an assertion or denial) and retraction (of an assertion or denial).

The inclusion of denial is unconventional. As put nicely by Smiley, in modern logic, '[l]ike the grey squirrel and red squirrel, assertion and negation have all but driven out rejection' [32]. Smiley observes that natural languages do, however, provide us with the means to express denial (or in his words, rejection) directly without resorting to assertion and negation, as in the answer 'No' to a polar question 'P?'

Apart from the fact that denial is a part of everyday language use, there are theoretical grounds for its inclusion. It is possible to extend my interpretation [26] of Brandom's logical expressivism to negation, once we have adopted denial as the counterpart of assertion. This gives us a system[8] in which negation (similar to implication) can be conceived of as a means for making an underlying practice explicit; in this case the practice of denial. Additionally, the resulting system is classical (and harmonious in Dummett's sense), rather than intuitionistic (i.e., $\Gamma \vdash \phi$ is derivable from $\Gamma \vdash \neg\neg\phi$). This allows us to address the common unease with proof-theoretic accounts of meaning as a result of their supposedly non-classical (i.e. intuitionistic) conception of logic.

We require, see below, that both the assertion and denial are informative. The preconditions for assertion and denial ensure that the asserted/denied sentence conveys new information relative to the speaker's (public) commitment store. Thus, it amounts to undertaking a commitment.

1. S ASSERTS sentence ϕ to A

   Precondition *not* $\Gamma_S \vdash \phi$

   Postcondition Γ^+_S is set to $\Gamma^+_S \cup \{\phi\}$

2. S DENIES sentence ϕ to A

   Preconditions *not* $\Gamma_S \dashv \phi$

   Postconditions

   Γ^-_S is set to $\Gamma^-_S \cup \{\phi\}$

3. S ASKS REASON FOR ASSERTION ϕ to A

   Preconditions (a) $\phi \in \Gamma^+_A$ and (b) not $\langle\Gamma^+_A - \{\phi\}, \Gamma^-_A\rangle \vdash_{NM} \phi$

[8] This system is obtained by addition of the following four rules for negation to [26], along the lines of Rumfitt's [30] bilateral logic: (1) From $\Gamma \vdash \neg\phi$ to $\Gamma \dashv \phi$, (2) from $\Gamma \dashv \neg\phi$ to $\Gamma \vdash \phi$, (3) from $\Gamma \dashv \phi$ to $\Gamma \vdash \neg\phi$ and (4) from $\Gamma \vdash \phi$ to $\Gamma \dashv \neg\phi$.

Postconditions Ξ_A^+ is set to $\Xi_A^+ \cup \{\phi\}$

4. S ASKS REASON FOR DENIAL ϕ to A

   Preconditions (a) $\phi \in \Gamma_A^-$ and (b) not $\langle \Gamma_A^+, \Gamma_A^- - \{\phi\}\rangle \dashv_{NM} \phi$

   Postconditions Ξ_A^- is set to $\Xi_A^- \cup \{\phi\}$

5. S RETRACTS ASSERTION OF ϕ

   Precondition $\phi \in \Gamma_A^+$

   Postcondition Γ_A^+ is set to $\Gamma_A^+ - \{\phi\}$

6. S RETRACTS DENIAL OF ϕ

   Preconditions $\phi \in \Gamma_A^-$

   Postcondition Γ_A^- is set to $\Gamma_A^- - \{\phi\}$

4.8 Sanctionable Behaviour

An interlocutor's behaviour in a dialogue or subdialogue is sanctionable, if at the end of the (sub)dialogue, the interlocutor's set of positive and/or negative open challenges (Ω^+ and/or Ω^-) is non-empty and/or their commitment store (consisting of positive and negative commitments Γ^+ and Γ^-) is inconsistent.[9]

4.9 The Inferential Background

We says that a judgement J holds relative to a set **B** of rules (i.e., the inferential background which underwrites meaning-giving inferences) if J can be derived using **B**.

We discern three types of rules: entry, inference and exit rules. An entry rule is of the form:

(1) $\dfrac{Test}{Judgement}$

In particular, we have:

(2) $\dfrac{\phi \in \Gamma^+}{\langle \Gamma^+, \Gamma^- \rangle \vdash \phi}$

and

(3) $\dfrac{\phi \in \Gamma^-}{\langle \Gamma^+, \Gamma^- \rangle \dashv \phi}$

These rules say that a sentence ϕ is (monotonically) affirmed/refuted if it is a member of the agent's positive/negative commitments. In the context of the game, this means that the agent must have asserted or denied ϕ explicitly (and not retracted it subsequently). We assume that at the outset of a game, the positive and negative commitments of an agent are empty, i.e. $\Gamma = \langle \emptyset, \emptyset \rangle$

There are also entry rules which allow us to introduce information from observations:

(4) $\dfrac{Observation^+(\phi)}{\langle \Gamma^+, \Gamma^- \rangle \vdash \phi}$

(5) $\dfrac{Observation^-(\phi)}{\langle \Gamma^+, \Gamma^- \rangle \dashv \phi}$

In these cases, the sentence ϕ will correspond, in everyday vernacular, with an expression of the form 'It looks F to me' or 'It looks like an F to me' (e.g., 'It looks red to me' or 'It looks like a bird to me'). A specific instance would be:

(6) $\dfrac{Observation^+(look\_penguin\_tweety)}{\langle \Gamma^+, \Gamma^- \rangle \vdash look\_penguin\_tweety}$

We can understand $Observation^+(look\_penguin\_tweety)$ as a simple classification device, which returns 'yes' if Tweety looks like a penguin.[10] Of course, the fact that Tweety has been classified as looking like a penguin, doesn't necessarily mean that he is a penguin; we return to this issue below.

Next, we have inferential rules of the form:

(7) $\dfrac{Judgement_1 \; \ldots \; Judgement_n}{Judgement}$ with $n \geq 1$

These include, among other things, monotonic material inferences such as the inference from Tweety is a penguin to it's a bird:

(8) $\dfrac{\Gamma \vdash penguin\_tweety}{\Gamma \vdash bird\_tweety}$

They can also express incompatibilities, e.g., if Tweety is bird then it isn't a mammal:

(9) $\dfrac{\Gamma \vdash bird\_tweety}{\Gamma \dashv mammal\_tweety}$

Importantly, the conditional part of a rule (above the line) can include judgements which express that the rule has a limited scope; i.e. that it is only applicable in certain situations. These conditions refer to the rule itself. For example, we may have:

(10) $\dfrac{\Gamma \vdash bird\_tweety \quad \Gamma \vdash scope\_bird\_tweety\_fly}{\Gamma \vdash fly\_tweety}$

These scope conditions for rules play a special role in non-monotonic inferences which we explain further on. Yet another rule that we will make use of is:

(11) $\dfrac{\Gamma \vdash penguin\_tweety \quad \Gamma \vdash scope\_penguin\_tweety\_fly}{\Gamma \dashv fly\_tweety}$

This can be paraphrased as: Provided the scope condition of this rule is satisfied (i.e., there is no reason to think that the rule doesn't apply), we conclude from Tweety being a penguin the denial of Tweety being able to fly.

Additionally, we use what are essentially 'blocking' rules. The following rule tells us that if Tweety is a penguin, then the situation is beyond the scope of rule (10) about birds flyings.

[9] This notion is intended only as a first approximation and would need to be made relative to the inferential capabilities of the members of the community in which the sanctions are issued. Without such a refinement, interlocutors would be expected to have access to all the inferential consequences of a dialogue act, both to be able to comply with the game and to judge others' compliance. This seems rather unrealistic. Additionally, one is, arguably, more blameworthy for a direct contradiction, then a contradiction that arises via a significant number of inference steps, suggesting that the severity of sanctions may depend on this factor.

[10] In practice, this would be a more general classifier that can distinguish between situations in which a penguin is present and those in which there is no penguin. Note that labelling the sentences that this classifier deals with as 'X looks like an F' only makes sense given the context where they are then inferentially linked to sentences that express 'X is an F'. Here, the account is in agreement with Sellars's [31] point that being able to make claims of the form 'X looks like an F' is dependent on the ability to make claims of the form 'X is an F'.

$$\frac{\Gamma \vdash penguin\_tweety}{\Gamma \dashv scope\_bird\_tweety\_fly} \quad (12)$$

We can also use scope conditions in the formulation of rules which takes us from what things look like to actual claims about what they are. Such inferences are non-monotonic (e.g. it is possible that though Tweety looks like a penguin, it is a different type of bird that has been painted to look like a penguin).

$$\frac{\Gamma \vdash look\_penguin\_tweety \quad \Gamma \vdash scope\_look\_penguin\_tweety}{\Gamma \vdash penguin\_tweety} \quad (13)$$

Finally, we have exit rules to actions of the form:

$$\frac{Judgement_1 \ \ldots \ Judgement_n}{Action} \text{ with } n \geq 1 \quad (14)$$

Such rules will typically be defeasible and can be used to justify entitlement to an action. We mention these for the sake of completeness, but won't use them in what follows.

4.10 Non-monotonic Inference

We define sc(**B**) as the set of sentences which occur in **B** and which express scope conditions of rules. These are atomic sentences. Their meaning derives purely from our definitions of non-monotonic affirmation/refutation and the role they play in the inference rules: they can be used to indicate the scope of a rule, and have inferential relations with other sentences, as in rule (12), which allows a scope condition to be denied under certain circumstances and, consequently, the corresponding rule to be blocked.[11]

Non-monotonic affirmation ('$\vdash_{NM}$') and non-monotonic refutation ('$\rangle \dashv_{NM}$') are defined as follows:[12]

(Non-monotonic affirmation) $\langle\Gamma^+, \Gamma^-\rangle \vdash_{NM} \phi$ *iff* there is a subset SC of sc(**B**) such that (a) $\langle\Gamma^+ \cup SC, \Gamma^-\rangle \vdash \phi$, (b) $\langle\Gamma^+ \cup SC, \Gamma^-\rangle$ is consistent, and (c) not $\langle\Gamma^+, \Gamma^-\rangle \dashv_{NM} \phi$

(Non-monotonic refutation) $\langle\Gamma^+, \Gamma^-\rangle \dashv_{NM} \phi$ *iff* there is a subset SC of sc(**B**) such that (a) $\langle\Gamma^+ \cup SC, \Gamma^-\rangle \dashv \phi$, (b) $\langle\Gamma^+ \cup SC, \Gamma^-\rangle$ is consistent, and (c) not $\langle\Gamma^+, \Gamma^-\rangle \vdash_{NM} \phi$

[11] As pointed out by one of the reviewers of this paper, the sentences in sc(**B**) could be viewed as part of the logical vocabulary, thereby invalidating the claim that our game is pre-logical. However, note that the sentences in question operate as tacit assumptions: the interlocutors do not use them in contributions to the game; in other words, these sentences do not need to be part of their communication language. More generally, one may object that the current proposal requires a logically expressive meta-language in which the inference rules and predicates, such as $Observation^+$, are stated. This challenge can be addressed by reflecting on the role that the meta-language plays here. This meta-language, with which we describe the game of giving and asking for reasons, is *not* a communication language. It is better thought of as part of a programming language which allows us to express the practices of the (pre-logical) game in procedural/computational terms. Ultimately, it is therefore intended as a recipe for physically implementing the ability to play this game (i.e. as a computational device).

[12] Non-monotonic affirmation and refutation are defined simultaneously. This does, however, not involve an infinite regress. We may for example establish that a sentence is non-monotonically affirmed by establishing conditions (a) and (b) and establishing (c) by showing that for no subset SC' the conditions (a) and (b) for non-monotonic refutation hold (which do not refer back to non-monotonic affirmation).
The reason for including clause (c) is that it allows us to deal with the Nixon Diamond problem of defeasible logic – a further detailed discussion is beyond the scope of this paper.

Non-monotonic affirmation/refutation allows us to jump to a conclusion. We don't need to have positive evidence that a scope condition holds. The absence of information to the contrary is sufficient for it to be used. We can assume that a scope condition for a rule holds, as long as this doesn't give rise to an inconsistency (in combination with our other commitments). This means that we can apply the rule (provided its other conditions hold as well).

Note that (in)consistency is itself defined using the monotonic variants of affirmation and refutation (see Section 4.5).

5 Giving and Asking for Reasons: An Example

This section sketches an analysis of a dialogue fragment, consisting of 8 contributions, in terms of the game of giving and asking for reasons that we defined in the previous section.

c1.	John:	Tweety flies.	ASSERT fly_tweety
c2.	Mary:	Why?	ASK REASON fly_tweety
c3.	John:	It is a bird.	ASSERT $bird\_tweety$
c4.	John:	Wait no, I take it back.	RETRACT and DENY fly_tweety
c5.	Mary:	Why?	ASK REASON DENIAL fly_tweety
c6.	John:	It is a penguin.	ASSERT $penguin\_tweety$
c7.	Mary:	Why?	ASK REASON $penguin\_tweety$
c8.	John:	It looks like a penguin.	ASSERT $look\_penguin\_tweety$

After c1, $\Gamma^+_{John} = \{fly\_tweety\}$. c2 results in $\Xi_{John} = \Omega_{John} = \{fly\_tweety\}$; the sentence fly_tweety is now part of the open challenges Ω_{John}. John is obliged to answer by providing a reason, otherwise his behaviour becomes sanctionable. In c3 he provides $bird\_tweety$. So now, $\Gamma^+_{John} = \{fly\_tweety, bird\_tweety\}$. We can remove fly_tweety from John's open challenges Ω_{John}, since $\langle\{bird\_tweety\}, \emptyset\rangle \vdash_{NM} fly\_tweety$. The inference goes through via rule (10) and the fact that we can assume, without endangering consistency, that $scope\_bird\_tweety\_fly$.

In c4, John both retracts fly_tweety asserted at line c1, and denies that fly_tweety. Again (c5) Mary asks for a justification. In c6, John responds with $penguin\_tweety$. With this new commitment it is non-monotonically refuted that fly_tweety, based on Rule (11). Note that after $penguin\_tweety$ has been added to the commitments, it is no longer possible to consistently add $scope\_bird\_tweety\_fly$. Rule (12) prevents this. This rule can be thought of as saying that the general rule about birds no longer applies once we have the information that the bird under consideration is a penguin. Penguins are beyond the scope of the general rule about birds and flying (and consequently we no longer can derive that Tweety can fly).

Mary also asks, in c7, for the justification of $penguin\_tweety$ and this is again supported through a non-monotonic inference. In this case, the reply, c8, is warranted by Rule (13), which takes us from evidence for Tweety looking like a penguin to the (defeasible) conclusion that Tweety is a penguin.

6 CONCLUDING REMARKS

This paper started with a number of puzzles and problems for the classical truth-conditional conception of meaning. I argued that together they suggest that it may be profitable to explore an alternative computational/inferential approach along the lines of Brandom's 'Making It Explicit' [6].

I presented a proof-theoretic formalisation of such an approach in which assertion and denial are on an equal footing – as briefly discussed, the resulting system is classical rather than intuitionistic. The other major departure from existing mainstream proof-theoretic accounts of meaning, especially of mathematical statements, is the formalisation of defeasible inferences (using rules that can be unblocked and blocked by means of scope conditions).

Defeasible inference plays a key role in addressing the puzzle of the use of meaning in perception. I show how defeasible inference allows us to loosen the connection between perception and inference (as required by the puzzle), whilst still explaining how observations can lead to the entry of sentences into an inference.

The puzzle of truth-conditionally equivalent sentences is addressed by taking inferential steps, rather than truth, as our primitive. Working out the justifications and consequences of a sentence (its meaning) against the body of background rules **B** requires computational effort, making visible differences in meaning that are not apparent from a truth-conditional point of view.[13] By grounding meaning in inferential practices, it also becomes possible to get a handle on meaning change, by modelling these as changes to inferential practices.This includes the possibility of modelling defective practices: when interlocutors encounter inconsistency, the blame doesn't necessarily need to be assigned to their commitments; it might be that the inferential practices themselves are flawed and need revision.

The game of giving and asking for reasons that is studied in this paper brackets out concerns with how interlocutors arrive at a common ground (i.e. manage to acquire common or shared commitments). In this respect, the current proposal is complementary to formalisations in logic and formal linguistics of dialogue state dynamics (e.g. [14, 13, 9, 20, 3, 12]) and further work is needed to combine these (e.g. along the lines proposed by Kibble [17]).

The main challenge I intend to address in future work is that of building a computer implementation of the current proposal, including the account of defeasible inference.

ACKNOWLEDGEMENTS

I would like to thank the two anonymous reviewers for their helpful feedback and suggestions.

REFERENCES

[1] R. Ahn, *Agents, Objects and Events: A computational approach to knowledge, observation and communication*, Ph.D. dissertation, Eindhoven University, 2001.
[2] J. Barwise and J. Perry, *Situations and Attitudes*, Bradford Books, Cambridge, MA, 1983.
[3] R.J. Beun, 'On the Generation of Coherent Dialogue: A Computational Approach', *Pragmatics & Cognition*, **9**(1), (2001).
[4] P. Blackburn and J. Bos, *Representation and Inference for Natural Language: A First Course in Computational Semantics*, University of Chicago Press, Chicago, 2005.
[5] T. Borghuis and R. Nederpelt, 'Belief Revision with Explicit Justifications: An Exploration in Type Theory', CS-Report 00-17, Eindhoven University of Technology, (2000).

[13] One way of introducing a measure of effort into inferences is the use of a search horizon (as in [5] and [24]). This has an interesting consequence in that explicit assertions and denials of already derivable sentences may reveal inconsistencies that were previously inaccessible to an agent, because they were beyond their search horizon (but with the explicit information added have now come to fall just within it). Thus, the foundations are laid for an account of how making implicit inferences explicit has genuine epistemic benefits.

[6] R. Brandom, *Making It Explicit: reasoning, representing, and discursive commitment*, Harvard University Press, Cambridge, Mass., 1994.
[7] R. Brandom, *Articulating Reasons: An Introduction to Inferentialism*, Harvard University Press, Cambridge, Massachusetts, 2000.
[8] R. Brandom, *Between Saying and Doing*, Oxford University Press, Oxford, 2008.
[9] H. Bunt, 'Dialogue pragmatics and context specification', in *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, eds., H. Bunt and W. Black, volume 1 of *Natural Language Processing*, 81–150, John Benjamins, (2000).
[10] M. Dummett, *The Logical Basis of Metaphysics*, Duckworth, London, 1991.
[11] G. Gentzen, 'Untersuchungen über das logische Schliessen', *Mathematische Zeitschrift*, **39**, 176–210, 405–431, (1934).
[12] J. Ginzburg, *The Interactive Stance: Meaning for Conversation*, Oxford University Press, Oxford, 2012.
[13] C.L. Hamblin, *Fallacies*, Methuen & Co Ltd., 1970. [Reprinted by Vale Press, Newport New, Virginia, 1993].
[14] C.L. Hamblin, 'Mathematical Models of Dialogue', *Theoria*, **37**, 130–155, (1971).
[15] H. Kamp and U. Reyle, *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer, Dordrecht, 1993.
[16] R. Kempson, W. Meyer-Viol, and D. Gabbay, *Dynamic Syntax: The Flow of Language Understanding*, Wiley-Blackwell, Oxford, 2000.
[17] R. Kibble, 'Reasoning about propositional commitments in dialogue', *Research on Language and Computation*, **4**(2-3), 179–202, (2006).
[18] J.R. Lucas, *The Freedom of the Will*, Oxford University Press, Oxford, 1991.
[19] P. Martin-Löf, *Intuitionistic Type Theory*, Bibliopolis, Naples, 1984.
[20] C. Matheson, M. Poesio, and D. Traum, 'Modelling grounding and discourse obligations using update rules', in *Proceedings of the 1st Annual Meeting of the North American Association for Computational Linguistics (NAACL2000)*, (May 2000).
[21] J.S. Mill, *On Liberty*, Longman, Roberts & Green, London, 1859.
[22] R. Montague, *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven, 1974. [Edited by Richmond H. Thomason].
[23] P. Piwek, *Logic, Information and Conversation*, Ph.D. dissertation, Eindhoven University, The Netherlands, 1998.
[24] P. Piwek, 'The ALLIGATOR Theorem Prover for Dependent Type Systems: Description and Proof Sample', in *Proceedings of the Inference in Computational Semantics Workshop (ICoS-5)*, Buxton, UK, (2006).
[25] P. Piwek, 'Meaning and dialogue coherence: a proof-theoretic investigation', *Journal of Logic, Language and Information*, **16**(4), 403–421, (2007).
[26] P. Piwek, 'Dialogue structure and logical expressivism', *Synthese*, **183**(1), 33–58, (2011).
[27] P. Piwek and E. Krahmer, 'Presuppositions in Context: Constructing Bridges', in *Formal Aspects of Context*, eds., P. Bonzon, M. Cavalcanti, and R. Nossum, volume 20 of *Applied Logic Series*, 85–106, Kluwer Academic Publishers, Dordrecht, (2000).
[28] D. Prawitz, *Natural Deduction: A proof-theoretical study*, Stockholm Studies in Philosophy, Almqvist and Wiksell, Stockholm, 1965.
[29] A. Ranta, *Type-Theoretical Grammar*, Clarendon Press, Oxford, 1994.
[30] I. Rumfitt, 'Yes and No', *Mind*, **109**(436), 781–823, (2000).
[31] W. Sellars, *Empiricism and the Philosophy of Mind*, Harvard University Press, 1997. [Originally published in Minnesota Studies in the Philosophy of Science, vol. 1, Minneapolis: University of Minnosota Press, 1956].
[32] T. Smiley, 'Rejection', *Analysis*, **56**(1), 1–9, (January 1996).
[33] J. Wanderer, *Robert Brandom*, Philosophy Now, Acumen Publishing Ltd, Stocksfield, 2008.