Relevance Theory, Pragmatic Inference and Cognitive Architecture

Wen Yuan[a]*, Francis, Y. Lin[b] and Richard P. Cooper[c]

[a]*School of Foreign Languages, Beihang University, Beijing, China*
[b]*School of English, Beijing International Studies University, Beijing, China*
[c]*Department of Psychological Sciences, Birkbeck, University of London, London, UK*

Relevance Theory (RT: Sperber & Wilson, 1986) argues that human language comprehension processes tend to maximize "relevance", and postulates that there is a relevance-based procedure that a hearer follows when trying to understand an utterance. Despite being highly influential, RT has been criticized for its failure to explain how speaker-related information, either the speaker's abilities or her/his preferences, is incorporated into the hearer's inferential, pragmatic process. An alternative proposal is that speaker-related information gains prominence due to representation of the speaker within higher-level goal-directed schemata. Yet the goal-based account is still unable to explain clearly how cross-domain information, for example linguistic meaning and speaker-related knowledge, is integrated within a modular system. On the basis of RT's cognitive requirements, together with contemporary cognitive theory, we argue that this integration is realized by utilizing working memory and that there exist conversational constraints with which the constructed utterance interpretation should be consistent. We illustrate our arguments with a computational implementation of the proposed processes within a general cognitive architecture.

Keywords: pragmatics; relevance theory; relevance-based comprehension procedure; spreading activation; monitoring

1. Introduction

Relevance Theory (RT) was developed by Sperber and Wilson (1986) in an attempt to capture general principles that govern pragmatic interpretation. Sperber and Wilson argued that the four maxims posited by Grice (1975) in his theory of conversational implicature (*quantity*, *quality*, *relation*, and *manner*) overlap with each other, and developed RT by building on just one Gricean maxim, that of *relation* (i.e., "be relevant": Grice, 1975, p. 46). RT is based on two principles: the cognitive principle of relevance and the communicative principle of relevance. The former specifies that human cognition tends to maximize relevance while according to the latter every ostensive stimulus conveys to the hearer a presumption of its optimal relevance. Sperber and Wilson (2002) subsequently developed the relevance-based comprehension procedure (RBCP) – a procedure that a hearer is held to follow when trying to understand an utterance. The first part of the RBCP postulates that the hearer takes the path of least effort in searching for an interpretation, while the second part tells the hearer to terminate the search process when the interpretive hypothesis is optimally relevant, or specifically, "the most relevant one compatible with the communicator's abilities and preferences" (Sperber & Wilson, 1995, p. 270).

Since its formulation, RT has been highly influential. It has been applied widely in discourse analysis (e.g., Tendahl & Gibbs, 2008; Schourup, 2011; Ifantidou, 2014; Yus Ramos 2016) as well as translation studies (e.g., Gutt, 2004; Díaz-Pérez, 2015). RT has also been the topic of many research treatises and anthologies (e.g., Blakemore, 1987; Carston &

Uchida, 1997; Rouchota & Jucker, 1998; Walaszewska & Piskorska, 2012; Clark, 2013; Padilla Cruz, 2016). And yet RT has never been free from criticisms. While most criticisms have focused on the vagueness of the principles and the inability of the theory to be falsified (e.g., Bach 1994; Levinson, 1989, 2000; Burton-Roberts, 2007; Soria & Romero, 2010; Davis, 2014; Mazzone, 2015), specific aspects of the RBCP, such as how the path of least effort is taken, or how speaker-related information is incorporated into the pragmatic process, have received relatively little critical evaluation.

In fact, such questions arose from the debates between Carston (2007), herself a relevance theorist, and Recanati (2002, 2004), who proposed dividing the unitary inferential pragmatic process (as posited by RT) into two processes: an accessibility-based primary process and a Gricean inferential secondary process. According to Recanati, explicature ("what is said" in Recanati's sense) is derived through the primary process by selecting the most activated interpretive candidate for a given constituent when semantic disambiguation or referent resolution is required. This process is argued to be independent of the speaker's mental states or beliefs. In contrast, Carston (2007) argued that when a sentential component has multiple potential semantic values, the most accessible need not be the one that should be chosen, because the speaker's knowledge, such as his or her beliefs or desires, has to be taken into consideration. She described the following scenario:

> My student Sarah is walking along with me in the department of linguistics. Suddenly Sarah says to me "Neil has broken his leg."

"Neil" might refer to either Carston's son $Neil_1$ or one of her colleagues $Neil_2$. So in this case, both individuals named "Neil" are activated. If Carston's son frequently gets into trouble and she has been constantly worried about him, $Neil_1$ will be chosen as the referent because he is more accessible. However, considering that Sarah does not know that Carston's son is named "Neil", and that $Neil_2$ is Sarah's syntax teacher, Carston instead selects $Neil_2$ to be the correct (i.e., intended) interpretation of "Neil".

Carston points out that according to Recanati's account, $Neil_1$ should be selected as the referent as it is more accessible, but it is the wrong interpretation. At the same time, Carston argues that this case can be easily explained from the relevance-theoretic perspective. According to the RBCP, the more accessible candidate, $Neil_1$, is chosen as the referent and checked against the speaker's abilities and preferences. If they are found to be inconsistent, the comprehension process will continue since the chosen referent $Neil_1$ fails to arrive at optimal relevance. The alternative option $Neil_2$ is then selected and the checking process repeats to ensure that it matches the speaker's knowledge about the referent.

In his response to Carston, Recanati (2007) restates that his proposed primary process includes two stages.[1] In the first stage the candidate of higher accessibility is chosen while in the second "accessibility shift" occurs (or specifically, the initially less accessible candidate $Neil_2$ becomes more activated later). He further suggests the second stage is triggered by the speaker's specific knowledge about the selected referent. But this proposal contradicts his account of the primary pragmatic process as one that does not taken into consideration the speaker's mental states. Moreover, it does not specify why the speaker's knowledge about a specific individual should become active.

Mazzone (2013) further indicates that, apart from Recanati's accessibility-based primary pragmatic process, RT faces a similar problem concerning the interpretation of "Neil" (and

[1] Recanati (2004) previously mentioned this point.

referent resolution more generally). Optimal relevance requires compatibility of the interpretive hypothesis with the speaker's abilities and preferences, but it fails to illustrate why only the specific piece of information (e.g., Sarah does not know $Neil_1$) is retrieved and integrated into the cognitive process as a contextual assumption.

Mazzone attempts to explain how speaker-related information enters the pragmatic interpretation by following the association-based account of Recanati. He proposes that activation of speaker-related information is enhanced due to the representation of the speaker in the hearer's working memory (2013, p. 112). But this proposal doesn't differ much from the cognitive principle of relevance. The latter enables the hearer to allocate cognitive resources in an optimal way to "bring about the greatest contribution to the mind's general cognitive goals at the smallest processing cost" (Sperber & Wilson, 1995, p. 48). The hearer's goal is to understand the speaker's communicative intention, so it is necessary to pay enough attention to the speaker to maximize relevance, which naturally leads to the representation of the speaker gaining access to working memory. It seems that Mazzone's solution just translates the cognitive principle into associative terms. Mazzone (2015) further proposes that pragmatic processing is both governed and driven by higher-level goal-directed schemata and that speaker-related information receives prominence through top-down processing.

Seen from the above, both Mazzone and relevance theorists support the view that pragmatic understanding constitutes a unitary cognitive process. The former suggests that associative processes (effected by spreading activation within hierarchical schemata) can realize the inferences posited by RT, while the latter focuses on the inferential process of seeking communicative intention attributed to the speaker by the hearer.

In summary, RT neither specifies in detail how the least-effort path is achieved in searching for optimal relevance nor describes how speaker-related information might be incorporated into the pragmatic process of utterance comprehension at the architectural level. While Mazzone argued for a goal-based account of the Neil problem, his solution is not grounded in a complete and coherent model of cognition.

In this paper we argue that Mazzone's account is not incompatible with the assumptions held by RT. More specifically, we demonstrate that cases such as the Neil example may be addressed within a modified version of RT in which the basic theory is complimented with a goal-based view that includes production-like if/then rules. Furthermore, we suggest that this integration is realized by utilizing working memory and that there exist conversational constraints (or pragmatic schemata), which might be described in the form of production rules, with which a constructed utterance interpretation should be consistent. In order to support our argument, we present a computational model of utterance interpretative processes embedded within a contemporary cognitive architecture. The model illustrates two critical points: 1) that the accessibility order of different interpretive hypothesis of an utterance – an important step in the least-effort path – can be determined based on spreading activation; and 2) that the RBCP can be complemented with conversational constraints to specify how speaker-related knowledge may enter the process of utterance interpretation and thereby implement the mechanisms required for pragmatic reasoning.

2. Towards an Architecture for Pragmatic Interpretation

Following the concerns raised in the introduction about, for example, the role of speaker related knowledge (and world knowledge more generally) in pragmatic interpretation, we

take the view that the processes involved in pragmatic interpretation cannot be understood in isolation from the rest of the cognitive system. In order to relate these processes to wider cognition we therefore consider how they may be embedded within a cognitive architecture, i.e., within a theory of the functional components of the mind and their interaction (Newell, 1990). Several such theories have been proposed over the last 30 years, but a number of requirements of pragmatic interpretation appear to be met by the Contention Scheduling / Supervisory System (CS/SS) architecture, originally proposed by Norman and Shallice (1986). CS/SS proposes that routine or over-learned tasks are performed by an automatic system (CS), in which hierarchically-organised units compete for control of thought and behaviour through activation-based processing. In non-routine situations or where deliberate/intentional control is required, the operation of CS may be biased by SS, which may selectively excite or inhibit units within CS so as to achieve specific goals. SS consists of several functional subsystems. One of these monitors the processing of other subsystems to ensure that processing coheres with on-going goals and expectations, while another performs goal-directed problem solving and planning (see Shallice & Burgess, 1996, for a preliminary general decomposition of SS, and Sood & Cooper, 2013, and Sexton & Cooper, 2014, for more recent computational implementations in specific tasks).

Although the CS/SS architecture was originally proposed to explain cognitive control of external actions, Norman & Shallice (1986) point it out that this architecture can "apply to internal actions—actions that only involve the cognitive processing mechanisms" (p. 1). The specific analysis in Norman & Shallice's work also indicates the CS/SS architecture allows for the involvement of additional special-purpose cognitive subsystems (such as those concerned, for example, with language or number processing). Such features of CS/SS, combined with its successful implementations in different specific tasks, make it possible to apply this cognitive theory to pragmatic interpretation.

2.1 The High-Level Organization of Cognitive Processes for Pragmatic Interpretation

In order to apply the CS/SS architecture to the task of pragmatic interpretation, it is necessary to specify which aspects of pragmatic interpretation fall within the operation of CS and, for those aspects that do not, what supervisory processes are required. Consider first the processes and knowledge stores that must contribute to the pragmatic interpretation process, such that it might be influenced by speaker-related knowledge. We assume that the interpretation process contributes to or constructs a mental model of the topic of conversation (i.e., a situational model), and that this model is accessible to other cognitive processes. Within the CS/SS theory, the most parsimonious approach is to assume that the situational model is maintained in a task-general working memory.

The process of generating an interpretation for an utterance needs access to one's knowledge of the world (e.g., of people named "Neil"). We assume that the process of generating an initial or preliminary interpretation is routinized and automatic (and either performed by CS or by a special purpose linguistic subsystem operating along similar lines). However, as noted above, a key element of SS is a monitoring process, and such a process satisfies the functional requirement of ensuring that the situational model constructed in working memory during utterance interpretation is coherent. We suggest that this monitoring process also serves to ensure that any preliminary interpretation of an utterance is consistent with the rules or constraints of pragmatic interpretation (i.e., in the language of RT, that the interpretation matches the speaker's abilities and preferences). Where an interpretation fails to satisfy pragmatic conventions, the monitoring process must reject the interpretation. We

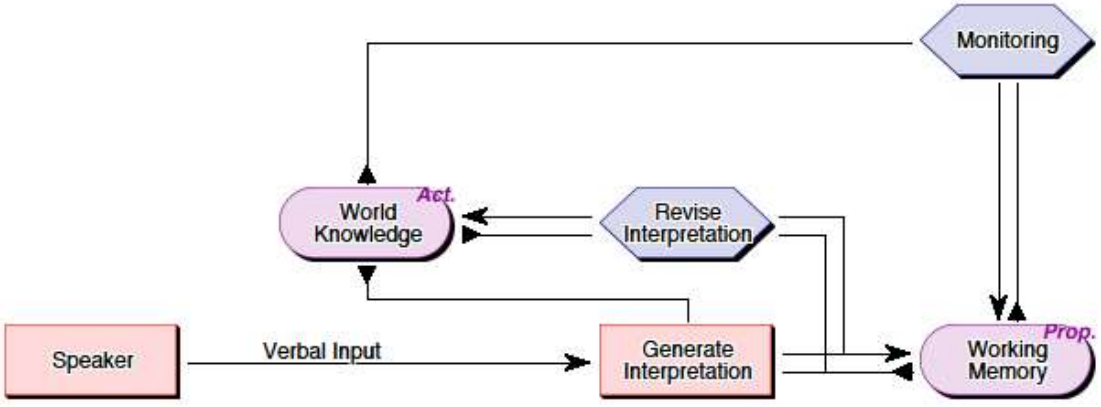

Figure 1. Critical architectural processes in the model of pragmatic interpretation. Rounded rectangular boxes depict buffers that store information or maintain representations. Hexagonal boxes depict processes that manipulate or transform those representations. Rectangular boxes represent compound components with internal structure. Arrows represent inter-component communication channels, with standard arrowheads indicating the sending of information and inverted triangular arrowheads indicating the reading of information. The diagram should be read as a circuit diagram, not a flow chart. In principle, each component functions in parallel.

assume that rejection of an interpretation results in the search for and subsequent generation of an alternative interpretation. This process is goal-directed and non-automatic, and must therefore recruit processes beyond CS (i.e., we envisage the process of revising an interpretation to draw on supervisory processes).

The relevant processes and knowledge stores, with their interconnections based on this discussion, are shown in Figure 1. It should be stressed that this figure shows the hypothesised information processing components and the channels between them that licence the exchange of information. It is more akin to a circuit diagram than a flow chart. The following sections further motivate and clarify the function of each of the components in the figure. Additional details of the operation of critical components are provided in the appendix.

2.2 The Generation of Interpretations

As indicated in Figure 1, we assume that separate stages are involved in the generation of an initial interpretation and, where that interpretation is found to violate pragmatic conventions, the generation of a revised interpretation. For expository purposes we treat these as distinct and separable cognitive processes. The initial process takes verbal input and incrementally builds an interpretation drawing on salient world knowledge. This interpretation is added to *Working Memory* in piecemeal fashion (i.e., as each syntactic sub-constituent is parsed), where it is assumed to augment the situational model being constructed in response to the current conversation. As noted above, we assume that this initial process, which includes parsing the verbal input, is either effected by the Contention Scheduling system or performed by a special-purpose subsystem that is governed by the same principles. That is, we assume it is automatic, activation-based, and driven by the most salient representations in the world knowledge store (Cooper & Shallice, 2000).

Units within CS encode cognitive or motor actions in terms of hierarchical relations and "triggering conditions". For example, in previous work on action control, units have encoded action associations such as "if the higher-level context is to add milk to a beverage and a non-empty milk container is held then it may be poured into the beverage container". Analogous structures may be used to encode cognitive operations such as "if the context requires interpretation of a name and a person by that name is known, then the name refers to that person". It is assumed that a large number of units coexist and that each unit has an associated activation value which receives excitation to the degree that its triggering conditions are met by sensory input, representations in *Working Memory* and/or salient representations in *World Knowledge*. The binding of object representations to a unit's argument roles (e.g., which container should the milk be poured into, or which individual named Neil should serve as the referent of "Neil") is determined by the salience or activation of object representations in the representation of the world. Contention Scheduling concepts therefore fulfil the requirements of an initial interpretative process.

A subsequent stage is assumed to be triggered when *Monitoring* detects that the situational model maintained in *Working Memory* is inconsistent (either internally or with the contents of *World Knowledge*). The *Revise Interpretation* process of Figure 1 performs this subsequent stage. It requires access to broad knowledge of the world (i.e., knowledge beyond that recruited by the immediate interpretation of the current conversation), and must be capable of performing arbitrarily complex reasoning. These properties are characteristic of the Supervisory System.

2.3 The Role of Working Memory

Any account of sentence interpretation is likely to require some kind of storage system to maintain the outcome of the interpretation process. We assume that this outcome is maintained in a general purpose working memory. We moreover assume as noted above that the outcome of interpretation is some kind of mental model of the situation being discussed, in which individuals and their properties and inter-relations are represented. The model must be veridical (i.e., its truth or falsity can in principle be ascertained) and so we assume that the component parts of the model are definite (i.e., held to be either true or false, and not probabilistic). At the same time there is no requirement that the model be internally consistent – ensuring consistency is the role of *Monitoring*, which as discussed below, is potentially unreliable. Lastly, the situational model must be accessible to processes beyond the initial interpretative process, so as a) to allow checking by the monitoring process, b) to feed into the generation of alternative interpretations if needed, and c) to be accessible to wider cognition in order to support action based on the interpretation (e.g., coherent engagement in an on-going conversation).

2.4 The Role of World Knowledge

We assume that world knowledge may be brought to bear in utterance interpretation and that this knowledge is represented within an activation-based store in which activation spreads between related concepts. Those concepts that are most active are most likely to play a role in the interpretation process. Thus, in the case of the Neil example, we assume that both individuals named "Neil" are represented in the hearer's store of world knowledge, but that the hearer's son is initially the more active of the two due to the hearer's personal associations and recent cognitive history (e.g., she has recently been thinking about her son).

The CS/SS architecture assumes that relevant aspects of world knowledge are represented in activation-based terms (as required by our model), but it does not provide a detailed account of activation within this knowledge store. Other cognitive architectures, in particular ACT-R, however, do. Thus ACT-R assumes that concepts (or "chunks" in ACT-R terminology) have an activation level that determines their accessibility in declarative memory and that a concept's activation level is determined by its history of use (which includes a recency bias) and its associations with related active concepts (cf. Anderson, 2007). We assume a similar mechanism operates within the CS/SS architecture, and with this, the activation of the two individuals named "Neil" may be used such that the one with the greatest activation is more accessible to the hearer. This accounts for the first part of the RBCP: that the hearer generates and tests the interpretive hypotheses in order of accessibility.

Accessing the second possible referent for "Neil" requires that the rank order of activations of the two individuals named "Neil" be reversed. This is achieved indirectly by the *Revise Interpretation* process, which may function either by deliberately inhibiting the representation of Neil-the-son and/or by deliberately exciting the representation of others known by the same name.

2.5 Monitoring and Epistemic Vigilance

The existence of monitoring as a task-general process is well supported by neuropsychological evidence (see Shallice & Cooper, 2011, for a review). *Monitoring* must have access to information from multiple sources (including *World Knowledge* and *Working Memory*), but it may be computationally relatively simple in that it detects (but does not attempt to resolve) inconsistencies in the situational model or mismatches between expectations and that model. It may be specified in terms of acquired goal-specific schemas that encode (amongst other things) pragmatic constraints (e.g., that the speaker has knowledge of a referent).

Psychological evidence (e.g., concerning action slips and lapses in everyday behaviour: Norman, 1981; Reason, 1984) suggests that monitoring is an imperfect process. It is subject to attentional failure and can be misled by confirmation bias (i.e., mismatches may not be detected if they are consistent with strongly held beliefs). Moreover, it is not in itself a conscious process, though the outcome of monitoring mismatch detection generally will be, because detection of such failures will typically trigger processes needed to resolve those mismatches (e.g., triggering search for an alternative pragmatically appropriate interpretation).

There is a close relation between monitoring and the concept of epistemic vigilance as introduced by Sperber et al. (2010) and extended by Mazzarella (2013, 2015a). Sperber et al. (2010) argue that epistemic vigilance and mind reading abilities develop simultaneously (so one can represent the speaker's beliefs and detect the mismatch between one's interpretation and those beliefs) and that the former is utilised to check the reliability of an utterance's source and the believability of the content. Mazzarella (2015a) extends the function of epistemic vigilance from discovering inconsistency to modulating the expectations of optimal relevance that drive the RBCP and determine its stopping point. Thus, Mazzarella proposed three types of expected optimal relevance corresponding to the three comprehension strategies of Sperber (1994):

(1) Naive optimists presuppose that the speaker is honest and that his or her utterance has an evidential basis. The hearer who adopts this strategy will regard the first interpretation that is relevant enough to be the meaning of the utterance.

Contents of World Knowledge (at cycle 14):

```
 0:  knows(sarah,neil_colleague,true).        0.4382
 0:  knows(sarah,neil_son,false).             0.5537
 0:  name(neil_son,neil).                     0.4883
 0:  name(neil_colleague,neil).               0.1979
```

Contents of Working Memory (at cycle 14):

```
14:  metadata(refers_to,utterance_2,ind_1).
14:  name(ind_1,neil).
14:  entity(ind_1).
 2:  metadata(speaker,utterance_2,sarah).
```

Figure 2. Contents of World Knowledge and Working Memory after initial processing of the word "Neil". Note that processing is cyclic, and the number at the left of each line indicates the cycle on which the element was added to the store. Thus, all elements in World Knowledge were present at initialisation (cycle 0), while the elements in Working Memory were added as processing progressed. Elements in World Knowledge have associated activation values, and these are shown to the right of each element. These were initialised to 0.5, 0.5, 0.5 and 0.2 respectively, but are subject to noise and hence vary slightly from cycle to cycle.

(2) Cautious optimists also believe in the honesty of the speaker, but they might doubt that the speaker correctly estimates what the hearer knows, such as what is the most accessible or relevant to the hearer under the given circumstance.
(3) Sophisticated interpreters, in contrast, drop the assumptions that the speaker is benevolent and competent. They select the interpretation that seems adequately relevant to the speaker.

In other words, epistemic vigilance mechanisms, according to Mazzarella (2015a), can check not only the constructed utterance meaning against the communicator's competence and benevolence, but also determine what version of the RBCP to deploy. The relation between epistemic vigilance and monitoring is further examined in Section 4.2.

3. A Worked Example: Implementation of the Neil Problem

In this section we demonstrate the model by showing its operation when processing Carston's critical sentence "Neil has broken his leg". The model takes as input one word at a time, and builds semantic units which are placed in *Working Memory* bit by bit to form a situation model. Binding of referents is also done bit by bit, as and when sufficient information is available to identify potential referents. We consider the functioning of each of the five boxes in Figure 1 throughout processing of the target sentence.

The hearer's *World Knowledge* is assumed to contain, amongst many other items, knowledge (or belief) that:

a) Neil-the-son is named "Neil"
b) Neil-the-colleague is named "Neil"
c) Neil-the-son is not known to Sarah
d) Neil-the-colleague is known to Sarah

Contents of World Knowledge (at cycle 15):

0:	`knows(sarah,neil_colleague,true).`	0.4308
0:	`knows(sarah,neil_son,false).`	0.5480
0:	`name(neil_son,neil).`	0.4780
0:	`name(neil_colleague,neil).`	0.2416

Contents of Working Memory (at cycle 15):

```
15: identity(ind_1, neil_son).
14: metadata(refers_to,utterance_2,ind_1).
14: name(ind_1,neil).
14: entity(ind_1).
 2: metadata(speaker,utterance_2,sarah).
```

Figure 3. Contents of World Knowledge and Working Memory after initial resolution of the referent of the word "Neil". On cycle 15, the individual in the situational model introduced by the proper noun "Neil" is identified with Neil-the-son.

It is moreover assumed that, when Sarah begins her utterance, the activation of element a) is higher than that of element b), reflecting the fact that Carston (the hearer) is particularly concerned about Neil-the-son. (Presumably all of Carston's beliefs about Neil-the-son have elevated activation, due to this concern.) For expository purposes we assume an initial activation of 0.50 for element a) and of 0.20 for element b).

When the speaker (Sarah) utters the first word of the phrase ("Neil"), subprocesses of *Generate Interpretation* parse the word and recognise it as a proper name. This results in the construction, in *Working Memory*, of the beginnings of the situational model. Figure 2 shows the contents of the two key memory stores, *World Knowledge* and *Working Memory* at this stage.

The presence of an unbound entity in the situational model immediately results in an attempt by *Generate Interpretation* to bind that entity to a known individual. Neil-the-son is selected as the corresponding individual due to the high activation of element a) compared to element b) in *World Knowledge* (cf. Figure 2). The result is shown in Figure 3.

A pragmatic schema within *Monitoring* then detects that Neil-the-son is believed to be not known by Sarah (the speaker), and so the assignment of the new entity to Neil-the-son is retracted. The resultant contents of the key memory stores are similar to those in Figure 2. The lack of a valid assignment of an individual to the referent of "Neil" then results in *Revise Interpretation* searching for an alternative interpretation. This results in excitation of knowledge of other individuals named Neil (and/or inhibition of knowledge related to Neil-the-son within *World Knowledge*). Consequently, knowledge that "Neil" may refer to Neil-the-colleague becomes more active than the corresponding knowledge that it may refer to Neil-the-son, as shown in Figure 4.

The mechanism for assigning known individuals to entities in the situational model (which is technically part of the *Generate Interpretation* process) then retrieves Neil-the-colleague as an alternative referent for "Neil".

Contents of World Knowledge (at cycle 22):

```
 0:  knows(sarah,neil_colleague,true).        0.4391
 0:  knows(sarah,neil_son,false).             0.6052
 0:  name(neil_son,neil).                     0.4900
 0:  name(neil_colleague,neil).               0.5116
```

Contents of Working Memory (at cycle 22):

```
14:  metadata(refers_to,utterance_2,ind_1).
14:  name(ind_1,neil).
14:  entity(ind_1).
 2:  metadata(speaker,utterance_2,sarah).
```

Figure 4. Contents of World Knowledge and Working Memory after rejection of the initial solution of the referent of the word "Neil" and subsequent excitation of the other possible referent. On cycle 16, the identification of the individual in the situational model introduced by the proper noun "Neil" with Neil-the-son is retracted and knowledge related to any other individuals known by the name of Neil begins to be deliberately excited in the search for an alternative referent. By cycle 22, the activation of the critical knowledge linking the name to Neil-the-colleague exceeds the corresponding information related to Neil-the-son. This will allow Neil-the-colleague to be retrieved on cycle 23.

While all of this is happening, further words are being uttered by the speaker and being integrated into the hearer's situational model. Figure 5 shows the contents of the key memory stores once the full utterance is processed.

The above is a detailed explanation of how the explicature is derived through the interaction of various cognitive mechanisms. The binding between the semantic unit of the proper noun Neil and the referent is based on activation strength of the candidates, and this process well instantiates how interpretive hypotheses are acquired in order of accessibility. The checking of compatibility of the interpretative hypothesis against speaker's abilities is accomplished by constraints in the monitoring system.

4. Discussion

We have elaborated the buffers and processes involved in the architecture of pragmatic interpretation, with the aim of grounding pragmatic interpretation in a theory of the cognitive architecture and illustrating in concrete terms how the theory solves referent resolution (and in particular, the Neil problem). Within our account and implementation, humans construct an interpretive hypothesis based on the activation strength of different meanings, but a monitoring mechanism functions to detect any inconsistency between the interpretation and the information in the working memory or world knowledge of the hearer. This detection process is realized by checking acquired conversational constraints, and ensuring that the situational model is consistent with those constraints, triggering reinterpretation if a constraint is found to have been violated. This explains in computational terms how the shift from the preliminary interpretation of "Neil" to the subsequent one occurs.

Contents of World Knowledge (at cycle 116):

```
  0:  knows(sarah,neil_colleague,true).        0.5513
  0:  knows(sarah,neil_son,false).             0.5279
  0:  name(neil_son,neil).                     0.4653
  0:  name(neil_colleague,neil).               0.5185
```

Contents of Working Memory (at cycle 116):

```
116:  is_male(ind_4)
116:  entity(ind_4).
116:  belongs_to(ind_3,ind_4).
116:  entity(ind_3).
116:  object(ev_2,ind_3).
116:  perfect_aspect(ev_1,ev_2).
116:  agent(ev_1,ind_1).
110:  metadata(refers_to,utterance_2,ind_3).
110:  isa(ind_3,leg).
 62:  metadata(describes,utterance_2,ev_2).
 62:  time(ev_2,past).
 62:  type(ev_2,breaking).
 38:  metadata(describes,utterance_2,ev_1).
 23:  identity(ind_1, neil_colleague).
 14:  metadata(refers_to,utterance_2,ind_1).
 14:  name(ind_1,neil).
 14:  entity(ind_1).
  2:  metadata(speaker,utterance_2,sarah).
```

Figure 5. Contents of World Knowledge and Working Memory after processing the entire utterance ("Neil has broken his leg"). Note that on cycle 23 ind_1 is identified (correctly) with Neil-the-colleague. Processing beyond cycle 23 integrates the remaining words into the situational model.

4.1 Reconsidering the Debate between Relevance Theory and Association-based Accounts

As described in the introduction, both relevance theorists and Recanati mention the shift from the preliminary interpretation to the alternative one in the Neil case. The former resorts to optimal relevance that requires the compatibility of the interpretive hypothesis with the speaker's abilities and preferences, while the latter attributes it to the dynamic change of the two interpretive candidates' activations in the second stage of the primary pragmatic process. Despite using different terms, relevance theorists' account and Recanati's explanation are essentially the same, as both concern a cognitive process that constructs an interpretive hypothesis and a confirmation process that finds the preliminary interpretation wrong and triggers a problem-solving process of searching for the alternative. However, both accounts lack elaboration of the mechanisms that implement the processes of deriving an interpretive hypothesis and evaluating the hypothesis. Our account bridges this gap by implementing the two component processes of "hypothesis formation" (based on the salience of each interpretive candidate) and "hypothesis confirmation" (in relation to world knowledge).[2]

[2] The two terms—"hypothesis formation" and "hypothesis confirmation" are borrowed from Mazzarella (2014).

Our account differs from RT in terms of the specific cognitive mechanisms that implement the above two component processes. As seen from the implementation section, the two processes are effected by the interaction of different cognitive mechanisms within a broader architecture of mind. In contrast, relevance theorists take them to be performed by an autonomous, dedicated module (e.g., Sperber & Wilson, 2002; Wilson & Sperber 2004). Whether such a module exists is highly controversial (see Bloom, 2002, Woodward & Cowie, 2004, and Cummings, 2015, for opposite views; see Sperber, 2000, 2005, for supportive arguments). Our account is therefore consistent with RT in terms of its theoretical proposal of the RBCP, but not its suggestion of a specialized module that completes this task at the level of implementation.

Our account implements the two stages of the primary pragmatic process proposed by Recanati, but it is in no way equivalent to saying we support his dual-process account of pragmatic interpretation. Instead, the secondary process in our view is implemented by the same mechanisms as the primary process. Relevance theorists (e.g., Mazzarella, 2014) have emphasized that association-based processes lack the mechanism for hypothesis evaluation. Thus Mazzone puts forward the idea of attentional processes for "active maintenance and conscious monitoring of information" (2013, p. 113). However, he fails to connect his idea to a coherent cognitive framework to implement such associative processes. Our model, composed of components which are constructed based on a single cognitive theory—CS/SS, successfully implements what Recanati and Mazzone take associative processes to do (e.g., the functions of attentional processes mentioned above, corresponding respectively to the functions of *Working Memory* and *Monitoring* in our model).

In this sense, our research provides a unitary account of utterance interpretation within the adapted cognitive framework of CS/SS, and shows that both the theoretical proposal of the RBCP and the association-based accounts may be implemented by the same cognitive modules, thereby effectively resolving the debate concerning the Neil example.

4.2 Monitoring and Epistemic Vigilance Reconsidered

As seen from the elaboration of *Monitoring* in Section 2.5, epistemic vigilance and *Monitoring* are different in terms of the specific tasks they serve. Epistemic vigilance, as described by Sperber et al. (2010) and Mazzarella (2015b), is specialized for verbal communication, but it has also been pointed out (as indicated in the footnote of Mazzarella, 2014, p. 93) that the proposal of an epistemic vigilance module seems to be inconsistent with relevance theorists' argument that the RBCP works as an autonomous module for inferential communication. Thus, *Monitoring* as we conceive of it, is a task-general process that encapsulates constraints or processing expectations (cf. Shallice & Cooper, 2011, pp. 367-371). These constraints are goal-specific (so interpretative pragmatic constraints apply during utterance interpretation, but not when, for example, playing a musical instrument), and assumed to be acquired through experience.

Despite the above difference, they also share some similarities. Firstly, *Monitoring* is functionally similar to epistemic vigilance. As mentioned earlier, epistemic vigilance tells us when to terminate the interpretation process by modulating the expected type of optimal relevance. The monitoring system can do a similar job through processing expectations for reinterpretation. For example, if the hearer expects to receive a specific type of answer (e.g., yes or no) but the speaker's utterance literally does not meet her/his expectation, the monitoring process, which detects this failure, will enable the hearer to choose a further

interpretation based on utterance meaning and contextual information (e.g., to derive implicature).

Secondly, epistemic vigilance and *Monitoring* are governed by similar constraints. According to Sperber et al. (2010) and Mazzarella (2015b), the activation of epistemic vigilance mechanisms depends on the relevance of the communicated content and the allocation of cognitive resources. As mentioned earlier, *Monitoring* is also subject to distraction. In this sense, attentional failure affects the performance of both epistemic vigilance and *Monitoring*. Thus, when the hearer is not fully attending in a conversation, s/he might have difficulty concentrating on the utterance the speaker says. This would prevent her/him from detecting errors or mistakes of the speaker or her/his own interpretive mistakes (e.g., accepting Neil_1 to be the referent in our example).

In the Neil case, the hearer Carston might initially take the name "Neil" to be her son, and then only on her way back home might she realize that it refers to her colleague. The selection of the wrong referent might be due to failure to detect the inconsistency or inability to solve the problem. The former might be caused by the hearer's lack of attentional resources or the failure of information retrieval to allow detection of the inconsistency (i.e., a failure of *Monitoring*). Consider the following example,

A: So, is this your first film?
B: No, it's my twenty-second.
A: Any favourites among the twenty-two?
B: Working with Leonardo.
A: da Vinci?
B: DiCaprio.
A: Of course. And is he your favourite Italian director?

(Richard Curtis, *Notting Hill*, 1999)

This conversation is extracted from an interview with the movie star B. Apparently B presupposes that A knows that Leonardo refers to the famous actor Leonardo DiCaprio. But judging from what A says, we infer that A mistakenly took Leonardo to be Leonardo da Vinci. But A is not sure about his understanding of the referent, so A replies with a question. Of course, the conversation, when placed within the broader context (e.g. a comedy film), is designed by the author of the film with the aim of yielding some humorous cognitive effects. Those effects are only achieved because the conversation is plausible or realistic. Presumably the reason for A getting the wrong referent is that Leonardo di Vinci is the most accessible person named "Leonardo" to A (Leonardo da Vinci is extremely famous). His failure to detect the inconsistency – that B could not have worked with da Vinci – might be due to A paying inadequate attention to the topic or the context or to the ignorance of A to the fact that da Vinci has nothing to do with acting.

The above analysis is consistent with the process of *Monitoring* as included in our model. It is not always reliable, and is constrained by attentional resources and cognitive control. Our model has the interpretive power to elaborate how correct utterance interpretation is arrived at, but also to indicate which factors lead to misunderstanding. As the specific operation of *Monitoring* depends upon learned pragmatic schemata, these will be explored further in the next section.

4.3 On the Nature of Pragmatic Schemata

The *Monitoring* process ensures that the developing situational model is coherent by applying a set of condition/action rules or pragmatic schemata. The general form of these schemata is that the conditions specify limitations on the contents of *Working Memory* (i.e., on the situational model) and the actions specify strategies for addressing these limitations. Thus, the specific pragmatic schema required for the Neil example (see the rule for *Monitoring* in the appendix) specifies that if the suspected referent of a referring expression is not known to the speaker then that suspected referent should be disqualified as the referent. When this rule fires it results in reinterpretation of the referent. One might ask where such schematic rules come from? Our view is that they are learned in very early childhood through observation and practice of verbal communication – either successful communication or misunderstanding of others.

As suggested by Recanati (2007), the Neil example involves the ability to metarepresent the speaker (e.g., the speaker is saying something). Such metapresentational ability has close links with children's general competence of attributing thoughts to others so as to predict or explain their behaviour. Evidence from developmental psychology reflects children's early use of such competence. For example, infants can identify what object an adult is labelling by referring to the adult's gaze (Baldwin, 1991), pick out the object that causes an adult's disgust towards it (Baldwin & Moses, 1994), and even learn the meaning of names by relying on some intuitive biases that derive from the ability to attribute thoughts to others (Bloom, 2002).

Compared with the above tasks, pragmatic interpretation is much more complex as it involves processing different types of information, such as linguistic, perceptual, and conceptual information, and it also requires metarepresentational abilities (Zufferey, 2015). Yet "children progressively become more skilled with the attribution of intentions in the context of verbal communication" (Zufferey, 2015, p. 91). In the transitional period, as children become progressively more skilled, abstract concepts (e.g., "goal", "intention", "know", etc.) are assumed to be grasped through an interaction between early mind-reading abilities and language development (de Villiers, 2007). Metarepresentational abilities are closely associated with the attribution of intentions, so such abilities also develop out of children's constant practice in the domain of verbal communication, which requires understanding of the speaker's communicative intention behind the utterance. On this account, those who have not acquired relevant abstract concepts or developed metarepresentational abilities might be unable to deliver the right interpretation of the name Neil. But when these conditions are satisfied, it is reasonable to assume that at first cognitive effort is required to deal with such situations so as to infer the speaker's intended referent. General problem solving may be required. However, as the appropriate strategy becomes proceduralized through constant use, the associated schema becomes established within *Monitoring*.

4.4 Comparison with Other Cognitive Architectures and Computational Approaches

We have presented our account within the context of a cognitive architecture based on the CS/SS framework of Norman and Shallice (1986), but many other cognitive architectures have been proposed. Of these other architectures, ACT-R (Anderson, 2007) is the most widely cited. It proposes that the cognitive system comprises a number of special purpose modules whose interactions are coordinated through a central production system. ACT-R modules correspond to computational functions, rather than psychological faculties. They include input systems (e.g., visual and auditory modules), output systems (e.g., the motor

module) and central modules (e.g., the intentional, declarative and imaginal modules). Each special-purpose module operates on its own content, and potentially in parallel. Furthermore, each module interfaces with the central production system through its own dedicated buffer. The central production system can read from and write to each of these buffers, thereby responding to input, triggering module-specific processing, and effecting motor activity.

ACT-R is of specific interest in the current context because a) its extensible nature allows further modules to be added to the architecture; and b) ACT-R's declarative memory is activation-based, with mechanisms supporting spreading activation and with the level of an item's activation determining its accessibility (as discussed above). With respect to the former, Lewis and Vasishth (2005) have exploited ACT-R's extensibility in their work on language comprehension and one could, for example, envisage a special purpose pragmatic module, which would be consistent with proposals of Sperber and Wilson (2002). With respect to the latter, ACT-R's activation-based mechanism offers a straightforward account as to why, in the Neil problem, the representation of Carston's son would be highly active and hence likely to be selected in the absence of a pragmatic violation.

Despite these strengths, ACT-R was not adopted here for two reasons. First, it does not distinguish between automatic and controlled processing, and second it does not treat monitoring as a distinct, separable cognitive process. Thus, while our account might be implemented within ACT-R, the ACT-R architecture would serve only as an implementation medium, and not as a background cognitive theory. Our adoption of the CS/SS architecture reflects a commitment to the psychological reality of the functional processes of this architecture within pragmatic interpretation.

More broadly, Poznański (1992) is one of the few previous researchers who has adopted a computational approach to implement the proposals of RT in utterance processing. Crystal, constructed by Poznański (1992), is a relevance-based utterance processing model that aims to provide a fine-grained computational account of the mechanisms involved and the interaction between them for achieving maximal relevance. Within Crystal, information from general input systems competes for access to a deductive device, and semantic segments of a given utterance from the language module are added incrementally to this device. Long-term knowledge in the form of concept-indexed chunks, each of which contains encyclopaedic, logical and lexical information, is stored in a conceptual memory. The three types of information of active chunks are copied respectively to the deductive memory, logical (meta-) rules and the language module. The deductive memory also receives source-labelled contextual implications produced by the inference process. Reason maintenance ensures logical consistency of the content in the deductive memory.

There are similarities between Crystal and our approach. For example, the deductive memory in Crystal is functionally similar to *Working Memory* in our model, as both can store information from encyclopaedic entries (world knowledge) and input from conceptual memory (lexical representations) and output of the inference process (contextual implications). The reason maintenance system of Crystal also shares some similarity with our *Monitoring* process, though as stressed above in our model *Monitoring* is a task-general cognitive subprocess, rather than a process specifically related to utterance interpretation. However, Crystal lacks a mechanism for spreading activation. Crystal's approach to semantic disambiguation instead relies on Sperber and Wilson's (1986) definition of relevance that equates mental effort with the number of assumptions used. This approach is inadequate in the case of the Neil example. Arguably, Crystal's inadequacy in resolving ambiguity might

be attributed to the fact that the work was accomplished before the addition of the communicative principle of relevance, which requires the interpretive hypothesis to be the most relevant, and to also match the speaker's abilities and preferences.

5. Conclusion

This paper has presented a solution for the Neil problem with a detailed computational account of pragmatic processing of referent resolution in utterance interpretation. The implementation of this example has shown how the hearer takes the path of least effort in searching for the interpretation of "Neil" and also how the speaker's personal knowledge about the referent is integrated with an evolving situational model in working memory. In order to support our proposed solution we have implemented two component processes of "hypothesis formation" and "hypothesis confirmation" within a general architecture of mind. Our account, to some extent, resolves the debate between the relevance-theoretic and association-based accounts of the Neil example.

As noted above, a key claim of RT is that the optimally relevant interpretation should match the speaker's abilities and preferences. The example discussed in this paper concerns the speaker's knowledge of an individual hypothesized as the referent. While we have focused on Carston's Neil problem, this problem is an instance of a more general pragmatic phenomenon where the hearer's assumptions of the speaker's knowledge are involved in the referent resolution process. We also need to explore how other kinds of speaker-related information might influence utterance interpretation. This question is closely related to the range of conversational constraints embedded within *Monitoring* that function to detect pragmatic violations by referring to world knowledge. Additional work is required both to elaborate further constraints and to evaluate how performance factors, such as cognitive load, affect the functioning of the architectural components.

Disclosure statement

No potential conflict of interest was reported by authors.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Bach, K. (1994). Conversational implicature. *Mind & Language, 9*, 124-162.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875-890.

Baldwin, D. A., & Moses, L. M. (1994): Early understanding of referential intent and attentional focus: Evidence from language and emotion. In C. Lewis & P. Mitchell (Eds.), *Children's Early Understanding of Mind: Origins and Development* (pp. 133-156). Hillsdale, NJ: Erlbaum.

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind and Language, 17*(1&2), 37-54.

Burton-Roberts, N. (2007). *Pragmatics*. Basingstoke: Palgrave.

Carston, R. (2007). How many pragmatic systems are there? In M. J. Frapolli (Ed.), *Saying, Meaning, Referring: Essays on the Philosophy of Francois Recanati* (pp. 18-48). New York: Palgrave Macmillan.

Carston, R., & Uchida, S. (1997). *Relevance Theory: Applications and Implications*. Philadelphia: John Benjamins Publishing Company.

Clark, B. (2013). *Relevance Theory*. New York: Cambridge University Press.

Cooper, R. P. (2002). *Modelling High-Level Cognitive Processes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cooper, R. P., & Fox, J. (1998). COGENT: A visual design environment for cognitive modeling. *Behavior Research Methods*, *30*(4), 553-564.

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297-338.

Cummings, L. (2015). Theory of mind in utterance interpretation: the case from clinical pragmatics. *Frontiers in Psychology*, *6*, 1286.

Davis, W. (2014). Implicature. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2014 Edition). Retrieved from http://plato.stanford.edu/archives/fall2014/entries/implicature/

de Villiers, J. G. (2007). The interface of language and Theory of Mind. *Lingua*, *117*, 1858-1878.

Díaz-Pérez, F. J. (2015). From the other side of the looking glass: A cognitive-pragmatic account of translating Lewis Carroll. In J. Romero-Trillo (Ed.) *Yearbook of Corpus linguistics and Pragmatics: Current Approaches to Discourse and Translation Studies* (pp. 163-194). Heidelberg: Springer.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41-58). New York: Academic Press.

Gutt, E. A. (2004). Challenges of metarepresentation to translation competence. In E. Fleishcmann, P. A. Schimitt & G. Wotjak (Eds.), *Translationskompetenz: Proceedings of LICTRA 2001: VII* (pp.77-89). Leipziger Internationale Konferenz zu Grundfragen der Translatologie. Tübingen: Stauffenburg.

Ifantidou, E. (2014). *Pragmatic Competence and Relevance*. Amsterdam: John Benjamins.

Levinson, S. C. (1989). Review article: Sperber and Wilson's relevance. *Journal of Linguistics*, *25*, 455-472.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375-419.

Mazzarella, D. (2013). 'Optimal relevance' as a pragmatic criterion: the role of epistemic vigilance. *UCL Working Papers in Linguistics*, *25*, 20-45.

Mazzarella, D. (2014). Is inference necessary to pragmatics? *Belgian Journal of Linguistics*, *28*, 71-45.

Mazzarella, D. (2015a). Pragmatics and epistemic vigilance: The deployment of sophisticated interpretative strategies. *Croatian Journal of Philosophy*, *44*, 183-199.

Mazzarella, D. (2015b). *Inferential Pragmatics and Epistemic Vigilance* (Unpublished doctoral dissertation thesis). University College London, London.

Mazzone, M. (2013). Attention to the speaker. The conscious assessment of utterance interpretations in working memory. *Language & Communication, 33*, 106-114.

Mazzone, M. (2015). Constructing the context through goals and schemata: top-down processes in comprehension and beyond. *Frontiers in Psychology, 6*, 1-13.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*, 1-15.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwarz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (Vol. 4, pp. 1-18). New York, NY: Plenum.

Padilla Cruz, M. (2016). *Relevance Theory: Recent Developments, Current Challenges and Future Directions*. Amsterdam: John Benjamins.

Poznański, V. (1992). *A relevance-based utterance processing system* (Report No. 246). University of Cambridge Computer Laboratory, Cambridge.

Reason, J. T. (1984). Lapses of attention. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of Attention* (pp. 29-61). Orlando, FL: Academic Press.

Recanati, F. (2002). Does linguistic communication rest on inference? *Mind & Language, 17*(1&2), 105-126.

Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, F. (2007). Reply to Carston. In M. J. Frapolli (Ed.), *Saying, Meaning, Referring: Essays on the Philosophy of Francois Recanati* (pp. 49-54). New York: Palgrave Macmillan.

Rouchota, V., & Jucker, A. H. (1998). *Current Issues in Relevance Theory*. Philadelphia: John Benjamins Publishing Company.

Schourup, L. (2011). The discourse marker now: A relevance-theoretic approach. *Journal of Pragmatics, 43,* 2110-2129.

Sexton, N., & Cooper. R. (2014). An architecturally constrained model of random number generation and its application to modeling the effect of generation rate. *Frontiers in Psychology, 5,* 1-14.

Shallice, T., & Burgess, P. W. (1996). The domain of supervisory processes and temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London, B351,* 1405-1412.

Shallice, T., & Cooper, R. P. (2011). *The Organisation of Mind*. Oxford: Oxford University Press.

Sood, M., & Cooper, R. P. (2013). Modelling the Supervisory System and frontal dysfunction: An architecturally grounded model of the Wisconsin Card Sorting Task. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1354-1359). Cognitive Science Society Incorporated, Berlin, Germany.

Soria, B., & Romero, E. (2010). *Explicit Communication: Robyn Carston's Pragmatics*. Basingstoke: Palgrave.

Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 179-198). Cambridge: Cambridge University Press.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 117-137). New York: Oxford University Press.

Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind: Structure and Contents* (pp. 53-68). Oxford: Oxford University Press.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, U., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 24*(4), 359-393.

Sperber, D. & Wilson, D. (1986/95). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D. & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language, 17,* 3-23.

Tendahl, M., & Gibbs, R. W. Jr. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics, 40,* 1823-1864.

Walaszewska, E., & Piskorska, A. (2012). *Relevance Theory: More than Understanding*. New Castle upon Tyne: Cambridge Scholars Publishing.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.

Woodward, J. F. & Cowie, F. (2004). The mind is not (just) a system of modules shaped (just) by natural selection. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 312-334). Malden MA: Blackwell Publishing.

Yus Ramos, F. (2016). *Humor and Relevance*. Amsterdam: John Benjamins.

Zufferey, S. (2015). *Acquiring Pragmatics: Social and Cognitive Pragmatics*. London and New York: Routledge.

Appendix: Model Details

The model described above is implemented in the COGENT cognitive modelling environment (Cooper & Fox, 1998; Cooper, 2002). COGENT is a graphical programming environment that allows cognitive models to be specified in terms of processes and storage buffers (as in Figure 1), and with the operation of processes defined in terms of if/then production-like rules. The following sections provide the precise rules used to implement the key processes in the model. The implementation also includes an incremental chart-parsing module adapted from chapter 7 of Cooper (2002).

Maintain Situational Model:

The process that maintains the situational model contains rules which copy semantic content produced by the parsing process into *Working Memory*, as well as a rule for binding discourse referents (e.g., entities introduced by proper names) with known individuals. The critical rule for the latter is as follows:

IF: `entity(X)` is in *Working Memory*
`identity(X, _)` is not in *Working Memory*
`name(X, Name)` is in *Working Memory*
`name(Individual, Name)` is in *World Knowledge*
THEN: add `identity(X, Individual)` to *Working Memory*

After parsing of the word "Neil", its semantic unit — `entity(X)` — is added to *Working Memory* (by other rules associated with maintaining the situational model). At this stage `X` has not been identified in *Working Memory* with a known individual since it is the first time `X` occurs in the conversation. `X`'s name is `Name`, which in the current example will be instantiated as "Neil". This information will also have been added to *Working Memory* by other rules associated with maintaining the situational model. The above rule consults the hearer's *World Knowledge* for known individuals whose name is `Name`. When the conditions in if-part are all satisfied, the action will be initiated to add the information that effectively binds the entity `X` with the individual whose name is `Name`. As introduced in section 2.2, the most accessible referent will be bound with the corresponding semantic unit if the latter has more than one potential referent within *World Knowledge*.

Monitoring:

In a fuller model, *Monitoring* will contain many schemata that implement goal-specific processing and pragmatic constraints. These schemata will also have associated activation values. For current purposes, we ignore the goal-specific and activation-based aspects of schemata and consider just the one schema critical to this form of referent resolution:

IF: `identity(X, Individual)` is in *Working Memory*
`metadata(speaker, Utterance, Speaker)` is in *Working Memory*
`metadata(refers_to, Utterance, X)` is in *Working Memory*
`knows(Speaker, Individual, false)` is in *World Knowledge*
THEN: delete `identity(X, Individual)` from *Working Memory*

With `X` identified as a specific individual, and attributing the utterance to the speaker, this conversational constraint (that the speaker has knowledge of the referent) functions to

retrieve the information that "it is false the speaker knows the individual" (the fourth condition of the rule). Hence the entity `X` is found to be assigned an invalid identity because it violates the conversational constraint. The preliminary mapping between `X` and the individual should therefore be deleted from *Working Memory*.

Revise Interpretation:

Removal of the preliminary mapping of a referent to an individual triggers the *Revise Interpretation* process. While in a more complete architecture, the function of *Revise Interpretation* would be fulfilled by a general problem solving process, just one rule is critical for the current example:

IF: `name(X, Name)` is in *Working Memory*
`identity(X, AnyIndividual)` is not in *Working Memory*
`metadata(speaker, U, S)` is in *Working Memory*
`knows(S, Individual, true)` is in *World Knowledge*
`name(Individual, Name)` is in *World Knowledge*
THEN: excite `name(Individual, Name)` in *World Knowledge* by `0.0500`

When *Working Memory* is found to contain a named referent `X` that has not been identified with an individual, but where an individual with the referent's name is believed by the hearer to be known by the speaker, then the representation of that individual in *World Knowledge* should be excited, so as to aid retrieval by the above rule in *Maintain Situational Model*. This rule will fire repeatedly until an appropriate individual is identified with the referent (i.e., until the second condition no longer holds).