THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2008, 61 (11), 1741–1760

Psychology Press
Taylor & Francis Group

Making disjunctions exclusive

Coralie Chevallier, Ira A. Noveck, and Tatjana Nazir
Institut des Sciences Cognitives, Lyon, France

Lewis Bott
Cardiff University, Cardiff, UK

Valentina Lanzetti and Dan Sperber
Institut Jean Nicod, Paris, France

This work examines how people interpret the sentential connective "or", which can be viewed either inclusively (A or B or both) or exclusively (A or B but not both). Following up on prior work concerning quantifiers (Bott & Noveck, 2004; Noveck, 2001; Noveck & Posada, 2003), which shows that the common pragmatic interpretation of "some", *some but not all*, is conveyed as part of an effortful step, we investigate how extra effort applied to disjunctive statements leads to a pragmatic interpretation of "or", *or but not both*. Experiment 1 compelled participants to wait for three seconds before answering, hence giving them the opportunity to process the utterance more deeply. Experiments 2 and 3 emphasized "or", either by visual means ("OR") or by prosodic means (contrastive stress) as another way to encourage participants to apply more effort. Following a relevance-theoretic line of argument, we hypothesized that conditions encouraging more processing effort would give rise to more pragmatic inferences and hence to more exclusive interpretations of the disjunction. This prediction was confirmed in the three experiments.

Keywords: Disjunction; Implicature; Pragmatics; Prosody; Relevance theory; Scalar inference.

Imagine a scenario in which Mary tells her boyfriend Paul what she would like for her birthday:

1. I'd like flowers or champagne.

Upon hearing (1), one could assume that Mary would be more than happy if her boyfriend gave her both flowers *and* champagne. This would mean that the inclusive interpretation of "or" applies. However, Mary could very well intend her boyfriend to buy her flowers or champagne, *but not both*. Most competent conversationalists would agree that this interpretation is derivable. In fact, the utterance is more informative with an exclusive interpretation of "or" because it is associated with fewer possible true cases than is the inclusive interpretation of "or" (see Table 1). Although each of these disjunctive interpretations arises with apparent ease, the process by which they are arrived at is far from obvious.

Does "or" represent a case of lexical ambiguity? Are the inclusive and exclusive interpretations of "or" both represented in the mental lexicon of English speakers (as are, for instance, the two senses of "cat" as a Felidae and as a domestic cat)? A consensus among semanticists, pragmaticians, and psycholinguists rules out this possibility.

Correspondence should be addressed to Ira Noveck, L2C2-ISC, 67 blvd. Pinel, 69675 Bron, France. E-mail: inoveck@isc.cnrs.fr

http://www.psypress.com/qjep
DOI:10.1080/17470210701712960

Table 1. *The truth conditions for an inclusive "or", an exclusive "or", and the conjunction, "and"*

P	*Q*	*or*	*or–exc*	*and*
T	T	T	F	T
T	F	T	T	F
F	T	T	T	F
F	F	F	F	F

Note: exc = exclusive "or".

Among the arguments against seeing "or" as linguistically ambiguous is that if it were ambiguous, one would expect its two senses to be expressed by two different words in many languages, just as are the two senses of "cat".[1] However, as Horn (1985) has argued, this is not the case for "or". Linguistic investigations reveal that there is no well-attested case of a word meaning "or" just in the exclusive sense. It is thus generally agreed that the linguistic meaning of "or" corresponds only to the inclusive sense (compatible with both disjuncts being true) and that the exclusive interpretation (incompatible with both disjuncts being true) results, not from mere linguistic decoding of the word "or" but from a pragmatic process of enrichment. In other words, the inclusive reading of "or" corresponds to a *literal* interpretation of the word whereas its exclusive reading corresponds to an *enriched* interpretation. Uncertainties and disagreement concern the way pragmatic enrichment works partly because the topic has been discussed mainly by philosophers of language, linguists, and pragmaticians who, in general, are less interested in processes, let alone in experimentally testable hypotheses about them.

In the remainder of the Introduction, we describe in detail two pragmatic theories, Stephen Levinson's theory of generalized conversational implicatures (GCI; Levinson, 2000) and relevance theory (Sperber & Wilson, 1986/1995), and show how the debate between them has already provided the cognitive literature with a body of experimental evidence that is generally favourable to relevance theory (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Noveck & Posada, 2003; also see, Noveck, 2001). This is followed by a description of *effort*, which is one of the factors considered critical to utterance interpretation by relevance theory, along with an explanation of how it can be especially influential in accessing the exclusive interpretation of the word "or". Finally, we introduce a set of three experiments testing our hypothesis regarding effort, while using an experimental design that is comparable to those of prior studies investigating similar logical lexical items such as "some" and "might".

Linguistic-pragmatic background

Both GCI theory and relevance theory draw on the work of the philosopher Paul Grice (1989). In particular, Grice wanted to show that ordinary language connectives such as "and" and "or" have a semantics that is equivalent to their logical meaning despite the fact that they can be used to convey richer and more diverse interpretations. For instance, "A and B" is logically equivalent to "B and A" but it can be used to convey an enriched "and then" (or "and thus") meaning, as in (2). Similarly, "A or B" is logically equivalent to "A or B or both" but it can be used to convey an enriched "A or B but not both" meaning, as in (3).

2. I dropped the vase and it broke.

3. I could invite John or Sarah.

In order to make his point, Grice drew on a fundamental distinction between sentence meaning and speaker's meaning. According to him, decoding the literal meaning of an utterance is only a means towards the real goal of comprehension, which is to arrive at an interpretation matching the speaker's intended meaning. Interlocutors in a conversation are guided, Grice argued, by expectations of informativeness, truthfulness, relevance, and so on. When the literal

[1] For example, the two senses of "cat" in French are expressed as "félin" and "chat"; a French speaker would never call a tiger a "chat".

meaning of an utterance does not suffice to fulfil these expectations, hearers are entitled to assume that the speaker intended them to inferentially enrich it. The distinction between the literal inclusive meaning of "or" and its frequent exclusive interpretation is a case in point.

To illustrate, consider Mary's case in (1) again. According to Grice, the semantic meaning of "or" in (1) is inclusive. Now, imagine there are conversational constraints that allow for refinement. For example, she knows that her boyfriend is short of money. She could still use the same utterance (1) since Paul would probably enrich the literal meaning of "or" on the basis of the specific conversational context (i.e., Paul has no money, and Mary knows it). Paul would then arrive at the enriched interpretation. This means that Mary's saying "flowers or champagne" could (potentially) indicate that she intends Paul not to buy both flowers and champagne although she did not *say* "but not both".

In the following section, we briefly present GCI theory and relevance theory and show how they make different claims about the way a hearer arrives at the exclusive interpretation of "or". In a nutshell, the former suggests that the inclusive interpretation arises after the default exclusive interpretation has been cancelled whereas the latter argues that the inclusive interpretation does not involve such a default to start with.

GCI and relevance theory

Both GCI theory and relevance theory agree that the inclusive meaning of "or" corresponds to its literal meaning and that the exclusive interpretation results from an inference. Both agree that this inference is common. Both agree that what makes the pragmatic inference justifiable is the fact that it provides an interpretation that meets the kind of expectations that interlocutors are entitled to have towards one another. The two theories differ about the role that context plays in drawing these inferences.[2]

For GCI theory, the inference from a literal inclusive interpretation of an utterance to an enriched exclusive interpretation is justified by the following consideration: The listener expects the speaker to be cooperative and well informed. Consequently, the speaker can be expected to utter the most cooperative and informative word possible. For the present case, "and" is more informative than "or" since "p and q" entails "p or q". Therefore, if a cooperative speaker chooses to utter a disjunctive sentence (i.e., "p or q"), the listener should take it to suggest either that the speaker has no evidence that "p and q" holds or that perhaps the speaker has evidence that it does not hold. Following the expectations that she has about the speaker, the listener should thus infer that *it is not the case that "p and q"*, thereby interpreting the disjunction as exclusive. For example, if a speaker chooses to say that "James or Peter will come to the party", the listener will have reasons to infer that "Not both James and Peter will come to the party". Because they are, in general, pragmatically valid, such inferences are made by default. That is, they are automatically triggered by the presence of the word "or" in the utterance, independently of the context.[3]

This is why such inferences are called *generalized* conversional implicatures (and are contrasted with *particularized* conversional implicatures, which are directly dependent on context). Of course, there are contexts where these inferences lead to an inappropriate interpretation (e.g., when an employer requires that applicants have "four years of experience or an MBA"). What happens, then, is that a second inference is drawn, based this time on contextual premises. This inference cancels the GCI (an applicant

[2] The two theories also differ about the exact content of these expectations and about the status of the inferred enrichment (implicature or aspect of the explicit meaning) but we are not concerned with these differences here.

[3] As we pointed out earlier, "or" is just one example of a class of terms that have often been called scalars. Other examples of scalar expressions that prompt such inferences include the quantifier "some", which excludes all, and the modal "might", which excludes must.

would recognize that an employer would want two beneficial attributes rather than one). The rationale for GCIs is that they speed up verbal communication by allowing fast default inferences that are pragmatically valid in most cases. Of course, in those cases (considered to be rare) where the default inference must be cancelled, the process should require distinctly more processing than it would have if the inference had not been drawn in the first place.

For relevance theory, there are no GCIs—that is, no default inferences in comprehension. In principle, listeners do not draw an inference unless it is contextually justified. In particular, relevance theorists expect listeners to adopt an enriched exclusive interpretation of "or" only when they have contextual reasons to do so. Without such contextual reasons a literal interpretation will do. A clear prediction of this view is that interpreting "or" exclusively involves deeper processing than interpreting it inclusively. This, of course, is just the opposite prediction to that of GCI theory.

In the following section, we summarize prior experimental work that addresses the issue of the mechanisms underlying pragmatic enrichment. To anticipate, we show that results in the experimental literature converge to show that pragmatic enrichment occurs if it is warranted by the context, as opposed to being generated automatically by a GCI. We thus argue that this body of data is generally more compatible with relevance theory. We then go on to explore some more specific claims made by this theory.

Experimental evidence

How does one test for the way an enriched interpretation is produced? As far as the disjunction is concerned, the only way to determine how people spontaneously treat "or" (as inclusive or exclusive) is to present a situation in which both disjuncts are true (Paris, 1973; Pelletier, 1977). That is, assuming that "p" and "q" are both true, an inclusive interpretation of "or" would lead to an evaluation of "p or q" as true, and an exclusive interpretation would lead to an evaluation of "p or q" as false; the other lines of the truth table remain unchanged (see Table 1).

Several classic studies investigated such cases using verbal descriptions like "The boy is riding a bicycle or the dog is lying down" when slides showed both a boy riding a bicycle and a dog lying down (Paris, 1973) or rules like "Either there is a P or a 4" and exemplars with both a P and a 4 (Evans & Newstead, 1980). Although the goal of these studies was not to study the actual process underlying the interpretation of disjunctions (e.g., the influence of deeper processing), the results indicate that the basic interpretation of "or" tends to be inclusive or at best equivocal. Below, we describe studies with other "scalar" terms where underlying interpretational processes were directly investigated.

Investigations into the processing of scalar terms have been generated by experiments designed to test the alternative views described earlier (Bezuidenhout & Cutting, 2002; Bezuidenhout & Morris, 2004; Bott & Noveck, 2004; Breheny et al., 2006; Noveck & Posada, 2003; also see, Noveck, 2001). Most of this work has focused on the time course of pragmatic enrichment and has progressed most intensively over the quantifier "some". This is why we now turn to findings related to the existential quantifier.

Consider how a participant can evaluate a sentence such as (4).

4. Some monkeys are mammals.

This utterance can be considered true if "some" is understood in its weaker form (which can be glossed as "Some and possibly all monkeys are mammals"), but it could also be considered false if it is enriched to "Some but not all" (yielding "Some but not all monkeys are mammals"). Although many have the intuition that the enriched interpretation is ubiquitous and perhaps primary, many data convincingly deny that. In Bott and Noveck (2004, Exp. 3), for example, participants answered equivocally to such underinformative sentences (indicating that roughly half draw the "not all" inference). Interestingly, the mean

reaction times of responses indicative of enrichments (i.e., answering "false" to statements such as 4) were extraordinarily long when compared to responses indicative of literal interpretations.[4] Data such as these were taken to indicate that the word "some" is optionally enriched through the computation of a pragmatic inference and that such an inference is not steadfastly linked to the word in the form of a default.

Even though such an experiment provided evidence against a default view, it brought no direct evidence in favour of relevance theory. Another experiment (Bott & Noveck, 2004, Exp. 4) was therefore designed to specifically test predictions from relevance theory concerning the processing of "some". As described earlier, according to relevance theorists, scalar inferences are not actually given by default but are computed, which involves a certain depth of processing. Consequently, if not enough time is given to participants to process an utterance, they should be less likely to draw scalar inferences. Following that prediction, Bott and Noveck (2004) manipulated the time available for the response: One group of participants had a relatively long time to respond (3,000 ms), and the other was given a short time to respond (900 ms). The results showed that while the rate of correct performance among the control sentences either improved or remained constant when the allotted response times increased, responses were such that there were significantly more literal responses to sentences like (4) in the *short-lag* condition than in the *long-lag* condition (72% and 56%, respectively). The experiments conducted by Bott and Noveck thus show that if people do not have enough time to answer, they are less likely to draw pragmatic inferences, an effect that is in line with the predictions made by relevance theory. Building on this body of work, which favours contextualist accounts over default accounts, we further explore some of the specific predictions made by relevance theory (for a review, see Noveck & Sperber, in press).

The analysis and predictions concerning "some" ought to apply similarly to "or". In this paper, we extend our investigation into enrichments by specifically testing the way depth of processing can influence the interpretation of "or" sentences. In the following section, we present some of the theoretical tools provided by relevance theory to study this issue as well as the ways we intend to test its specific predictions concerning disjunctions.

Testing relevance theory's predictions with disjunctions

At the core of relevance theory is the idea that hearers expect utterances to be relevant enough to be worth their attention and as relevant as is compatible with the speaker's abilities and goals. There are two factors that can influence the relevance of an utterance: effect and effort. On the one hand, the greater the "cognitive effects" (the consequences that can be derived from an utterance), the greater the relevance. On the other hand, the greater the processing effort[5] needed to achieve these effects, the lesser the relevance. The two factors work in tandem. An utterance will prove to be optimally relevant if its interpretation provides sufficient contextual effects while requiring only minimal justifiable processing costs.

If this view of the role of relevance in comprehension is correct, it suggests that manipulating the "effort" factor ought to affect interpretation. More specifically, causing listeners to expend more effort in processing some linguistic items ought to be a sufficient condition for enriching interpretations. Indeed, given that a hearer expects utterances to be relevant, she will be inclined to compensate any extra effort imposed

[4] Note, however, that the mean reaction times of responses reflecting literal interpretations was comparable to both true and false control items (such as "All cows are mammals" or "Some bees are mammals").

[5] What relevance theorists mean by "effort" corresponds better to what is meant in cognitive psychology by "depth of processing". In this article, therefore, we talk of "depth of processing", which is—like "effort"—a comparative notion. As will be seen, the experiments compare situations that require more or less effort (without necessarily quantifying it).

on processing with extra cognitive effects. In order to appreciate this interaction between effort and effect, compare, for instance, utterances (5) and (6):

5. I don't eat meat.
6. I don't eat dead flesh.

Decoding the two-word phrase "dead flesh" involves more processing than decoding the synonymous single word "meat". If both utterances were given exactly the same literal interpretation, then the extra effort required by (6) would make it marginally less relevant. Given that the hearer has certain expectations concerning relevance, however, the hearer will compensate the extra effort spent upon processing (6) by searching for extra cognitive effects (for instance, by understanding that the speaker considers people who eat meat to be morally inferior).

How can these factors be tested in an experimental setting? One way to manipulate the "effort" factor is to decrease or increase the cognitive resources available to the listener, rendering less or more likely an enriched interpretation. In the case where cognitive resources are decreased, the hearer will have less of an opportunity to process the utterance deeply (as in Bott & Noveck, 2004, Exp. 4). On the contrary, one can imagine a situation where the listener is encouraged to process an utterance more deeply because she is encouraged to spend more time processing the utterance or because his attention is focused in one way or another. In the former case, a literal interpretation of an utterance could fall short of fulfilling the listener's expectations of relevance. The listener should then prefer an enriched, and more informative, interpretation (provided the resources needed for inference-making are available). Consequently, if one allows, or encourages, a listener to apply more effort to process a disjunctive utterance, the listener should look for more effects and access more exclusive (enriched) interpretations.

Experimental paradigm

In the current experiments, a participant is shown a five-letter string and has to decide whether or not a given description ("There is an A or a B") is true or false. If the resources available for the interpretation of the description are in some way reduced, one would expect the participant to be less likely to enrich the description, making the literal interpretation of "or" adequate for responding. On the other hand, if the paradigm gives the participant the opportunity to apply resources to the task, one would expect participants to be more likely to enrich the interpretation and to provide a response indicative of an exclusive "or". Experiment 1 presents three conditions in order to manipulate depth of processing. In the control condition, the letter string remained visible until the participant responded to the descriptive sentence. In one experimental condition, participants' processing time was cut short (by removing the five-letter string before the descriptive sentence appeared), and, in another, participants were given the opportunity to carry out deeper processing (by extending the minimal amount of time necessary for responding).

Another way to encourage participants to apply more cognitive resources to the interpretation of the descriptive sentence is to attract their attention to a critical term (in our case, the word "or"). Indeed, attracting someone's attention amounts to encouraging the interlocutor to apply more cognitive resources in order to process what is under focus (Raz & Buhle, 2006). Therefore, increasing the saliency of a word is likely to change the way it is processed by increasing the attention it gets. One means to attract attention to a lexical item is to use contrastive stress (Pierrehumbert & Hirschberg, 1990). To return to Example 1, if Mary wanted to emphasize her intention that Peter buy flowers or champagne, *but not both*, she could say it with emphasis on the disjunction:

7. I'd like flowers OR champagne.

In Experiment 2, we investigate how prosody influences the interpretation of "or" by accenting it. We hypothesized that the presence of a contrastive stress on "or" would give rise to deeper processing, which would trigger the search for more effects and hence an enriched interpretation. In Experiment 3, we test the same hypothesis in a

within-participant experiment where the same participants hear "or" accented as well as nonaccented.

To be more specific, we presented French speakers five-letter-long strings followed by either disjunctive or conjunctive sentences concerning the letters in the string. Participants were then required to answer *true* or *false*. For example, after reading the word TABLE, the participant could receive a sentence such as "There is an A or a C", which is true, or a sentence such as "There is an A and a C", which is false. There were eight conditions altogether based on the two connectives ("and" or "or") and the four lines of the truth table (first letter *true*–second letter *true*, *true–false*, *false–true*, and *false–false*). Thus, the *true–true* case for the disjunctive sentences (e.g., "There is an A or a B"), which is the most relevant one for us, was just one of eight possible conditions. The remaining seven conditions served as control conditions.

To summarize, in order to determine whether deeper processing can change the interpretation of "or" we designed three experiments. In Experiment 1, we manipulated the time available for the participant to provide a response. Our prediction was that, when given the opportunity to spend more time processing the stimulus, participants would be more likely to derive a pragmatic inference. In Experiment 2, we had participants apply more effort either by capitalizing and underlining the "or" (in the *written* condition) or by adding a contrastive stress (in the *oral* condition). Our prediction was that such cues would favour the production of enriched interpretations. Experiment 3 followed up on Experiment 2 by testing whether the added emphasis on "or" could produce higher rates of exclusive interpretations in a within-participant experiment that presented the sentences only orally.

EXPERIMENT 1

According to relevance theory, all other things being equal, an effect ought to vary as a function of the cognitive effort applied (i.e., depth of processing). In the present case, the more participants apply cognitive resources to process a sentence, the more likely they are to arrive at an exclusive interpretation. To test this prediction, we manipulated the amount of cognitive resources that participants could use to process a sentence by asking them to judge its validity under three different time constraints. A rushed presentation should therefore discourage deeper processing with respect to a control condition that did not rush the presentation (the string remained visible as the descriptive sentence appeared and as participants responded). Conversely, asking people to wait to give their answer (the "*extra-time*" presentation) should encourage deeper processing, hence more enriched interpretations, than in the two other conditions.

Method

Participants

A total of 59 participants were recruited from the Université Catholique de Lyon and the Université de Lyon 2. All were native French speakers and received a gift worth approximately 5 euros for participation. Participants were randomly assigned to one of the three experimental conditions.

Stimuli and design

Three categories of stimuli were used: words, pseudowords, and nonwords[6] (see Table 2 for some examples). The words and the pseudowords were chosen from the BRULEX database (Content & Radeau, 1988) according to the mean frequency of the digrams or trigrams that composed them. We used that factor so as to control both words

[6] At first, we thought that the relative difficulty associated with those three types of items (words are more easily processed than pseudowords, and these, in turn, are more easily processed than nonwords) could possibly play an interesting role in our task. Insofar as nonwords require more processing effort to be remembered than pseudowords and words, we expected that fewer cognitive resources would be available for a potential enrichment of the disjunctive true–true (TT) utterances. When piloting revealed that word type had no impact on later truth evaluation, we decided to keep these three types of words because their presence obscured the goal of the task, but we did not include this factor in the analysis.

Table 2. *An illustration of each of the eight conditions presented in the experiments*

Connective	*Truth condition*	*Example of word*	*Word type*[a]	*Sentence*	*Logically correct response*
Or	TT	TABLE	Word	There is an A or a B.	T[b]
	TF	JAMIS	Pseudoword	There is an A or a B.	T
	FT	RSOUB	Nonword	There is an A or a B.	T
	FF	POJET	Pseudoword	There is an A or a B.	F
And	TT	BYAMS	Pseudoword	There is an A and a B.	T
	TF	DEVAT	Pseudoword	There is an A and a B.	F
	FT	LIBRE	Word	There is an A and a B.	F
	FF	XEGHI	Nonword	There is an A and a B.	F

Note: TT = true–true. TF = true–false. FT = false–true. FF = false–false.

[a]Note that each condition included all three word types.

[b]A true response is justified because the sentence contains a true disjunct, and a false response is justified because "or" can be taken to pragmatically imply "not both".

and pseudowords with the same orthographic criterion. We also made sure that no word had repeating letters. The stimuli were presented in white letters at the centre of a black computer screen and appeared in a random order.

After having seen the five-letter string, participants were asked to evaluate a test sentence of the following form: "There is a *letter 1* and (or) a *letter 2*." The sentences were presented at the centre of the screen, three lines below the five-letter combination. For example, a participant would see a word such as TABLE followed by the sort of sentences below:

"There is an A or a B" (*true–true* condition, TT).

"There is an A or an X" (*true–false* condition, TF).

"There is an X or a B" (*false–true* condition, FT).

"There is an X or a Y" (*false–false* condition, FF).

These four possibilities correspond to the four lines of the truth table (see Table 1). For the remaining four types of sentence, "or" was replaced by "and". In other words, each participant ultimately saw 96 sentences equally distributed among the two connectives ("and" and "or") and the four lines of the truth table (TT, TF, FT, FF; see Table 2).

For each participant, the sentences were generated randomly so that there were 12 trials for each of the eight types of sentence, the 12 five-letter combinations being evenly divided between words, pseudowords, and nonwords (i.e., each condition included 4 words, 4 pseudowords, and 4 nonwords). Furthermore, we were careful that the two letters in the test sentences were not polysyllabic in French (such as "Y" and "W"). Note that a single word could be associated with any of the eight sentence types depending on what had been generated randomly. For example the word "TABLE", which is presented only once per subject, could be the source for judging any one of the eight sentences (e.g., "There is a T and a B" or "There is a K or a B", etc.).

Procedure

Participants were presented with instructions at the beginning of the experiment. These explained that they would see a cross in the centre of the screen (which was designed to focus their attention on the beginning of the word to appear). They were told that they would see a chain of letters (e.g., "ZRETY") appear there and that a sentence to evaluate as true or false would follow (e.g., "There is a Z and an H"). The assignment of the right and left hands for *true* responses was randomly assigned and was counterbalanced across the experiment. After giving their answer, the participants saw a new word and a new sentence to evaluate. In the *fast-word*

condition, the letter string was shown for one second and was then replaced by the target sentence. In the *normal* condition, the letter string was shown for one second alone, and then the test sentence appeared below the string. Both the string and the test sentence remained until the participant provided a response. In the *extra-time* condition, the target appeared as it did in the *normal* condition, but participants were instructed to wait three seconds before answering. An error message ("Too fast") appeared if participants gave their answer before the imposed three-second waiting period.

Before the experimental session, participants saw four practice statements and five dummy sentences to avoid problems associated with starting the experimental phase. All participants saw the same practice and dummy sentences. The procedure used for practice trials was identical to that for the experimental trials. However, participants were encouraged to ask questions during the practice phase and to work independently during the experimental session. Participants were not told of the existence of the dummy sentences. Each experimental session was divided into two blocks in order to give participants a moment to pause. The program to run this experiment was written in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Results

In what follows, we first summarize how we analysed the data. This is followed by a description of the seven control sentences. We then turn to the cases that are central to the paper—the *true–true* condition.

Analyses for all statements are based on the proportion of correct answers on a literal interpretation of "and" and "or". It is important to note that in the *true–true or* condition, we take the participants who give an answer that reflects a nonliteral interpretation to be enriching the semantics of "or". Those participants are thus not treated as making a mistake but rather as providing an answer that is also correct on an enriched interpretation of "or".[7] Arcsine transformations were carried out before analysis to improve the conformity of the data to the standard assumptions of analysis of variance (ANOVA; e.g., Howell, 1997). Table 3 displays the mean percentage of correct responses (on a literal interpretation) for all eight conditions (note that, for the *true–true or* statements, the figures reflect the proportion of *true* responses, i.e., the rate of inclusive, literal interpretations).

Responses to the *true–false* and *false–true* conditions were combined because they are logically identical and only differ in the order in which the true letter appears in the question. Therefore, we made statistical comparisons between three cases: *true–false*; *false–false*; and *true–true*. Below, we analyse performance on the control items before turning to our main concern, the *true–true* condition with the connective "or".

Responses to statements containing the conjunction ("and") are represented on the left side of Table 3. For those responses made under the *normal* and *extra-time* presentation conditions, participants were at ceiling for all statements. The majority of participants responded accurately nearly 100% of the time in all "and" cases, indicating that they had no difficulty performing the basic task. For those responses made under the *fast-word* presentation conditions, participants responded less accurately (although still at an average of 91% correct) than did those under the *extra-time* and *normal* conditions. There were differences between the three types of question (*true–true*, *true–false*, and *false–false*) for "and" under these fast conditions, such that the *true–true* and the *true–false* questions were more difficult than the *false–false* question ($M = 0.86$ and $M = 0.90$ vs. $M = 0.98$, respectively). Statistical tests were inappropriate for comparisons between the three conditions because variance was negligible in many cells. In summary, participants responded accurately under all conditions using "and" but they found the *fast* condition slightly more difficult.

[7] Presumably some of the responses that we are taking as enriched interpretations are, in fact, errors. However, a dramatic change in error rates uniquely in this one condition is unlikely. If errors are made in the "true–true or" condition, one can safely assume that they are at a level comparable to that for the controls and that they ought to affect both sorts of interpretation proportionally.

Table 3. *Percentage of correct responses to each of the conditions in Experiment 1*

	A and B				*A or B*			
	TT TABLE	*TF JAMIS*	*FT RSOUB*	*FF POJET*	*TT TABLE*	*TF JAMIS*	*FT RSOUB*	*FF POJET*
Presentation	*True*	*False*	*False*	*False*	***True***	*True*	*True*	*False*
Fast-word	86	90	90	99	***80***	82	82	92
Normal	96	95	99	98	***75***	92	94	98
Extra-time	98	97	98	99	***52***	95	92	99

Note: TT = true–true. TF = true–false. FT = false–true. FF = false–false. The examples all assume that the test sentence concerns the letters A and B (which is why these letters are presented in bold in the example strings). The figures in bold refer to "true" responses to the main case of interest where, for example, a participant sees the word TABLE and says "true" to the statement "There is an A or a B". The only other available response is "false", and thus false responses to TT (and other true) cases can be calculated by subtracting the figure from 100.

When the connective "or" was used, response patterns for the control statements were similar. For both *true–false* and *false–false* questions, responses were highly accurate under all three presentation conditions ($M = 0.93$ overall). However, as with the "and" connective, responses were less accurate overall under the *fast-word* presentation condition than under the two other presentation conditions ($M = 0.86$ vs. $M = 0.95$ and $M = 0.96$, respectively).

We now turn to responses to the case of interest—"or" when the truth conditions are *true–true*. We anticipated that should participants be encouraged to expend extra effort in responding, they would make more exclusive interpretations of "or". This appears to be the case, as illustrated by the high rates of inclusive responding in the *short-* and *normal*-presentation conditions and the relatively low rates of inclusive responding in the *slow*-presentation condition ($M = 0.80$ and $M = 0.75$ vs. $M = 0.52$). This difference is reliable using an ANOVA with presentation condition as a between-participant variable, $F_1(2, 56) = 3.35$, $MSE = 0.325$, $p < .05$, $F_2(2, 144) = 15.279$, $MSE = 0.24$, $p < .0005$.[8]

To verify that the *extra-time* condition was responsible for the effect, we ran *t* tests between the three presentation conditions, using the response to the *true–true or* statement as a dependent measure. The *extra-time* condition showed reliably fewer inclusive interpretations than the *fast-word* presentation, $t_1(39) = 2.34$, $p < .05$, $t_2(189) = 5.73$, $p < .0005$, and it also showed fewer inclusive interpretations than the *normal* presentation (although the effect was marginal in the subjects analysis), $t_1(37) = 1.89$, $p = .07$, $t_2(190) = 3.57$, $p < .0005$. Further evidence that the *normal* vs. *extra-time* conditions were reliably different is shown by a significant interaction between presentation condition (*normal vs. extra-time*) and truth condition (*true–true*, *true–false*, and *false–false*), $F_1(2, 74) = 3.83$, $MSE = 0.13$, $p = .026$, and, $F_2(2, 122) = 12.82$, $MSE = 0.12$, $p < .0005$ (*true–false* and *false–false* conditions did not differ across presentation conditions, ts < 1). There was no significant difference when comparing the *normal* presentation to the *fast-word* presentation ($p_1 > .7$ and $p_2 > .07$).

This drop in inclusive responding cannot be due to a general trend of finding questions more difficult under the *extra-time* presentation condition because responses in this condition were generally more accurate overall in the control conditions. This is confirmed by a significant interaction between truth condition and presentation condition, as reported above.

[8] By convention we refer to *F* values obtained with participants as a random factor as F_1 (or t_1) and with items as a random factor as F_2 (or t_2).

Discussion

This experiment manipulated the time that participants had available to make truth value judgements. We found that when participants were prompted to take a prolonged period of time before answering, they were more likely to respond *false* to a disjunctive sentence whose two disjuncts were true (in comparison with the other two experimental conditions, i.e. the *normal* and the *fast-word* presentation conditions). This implies that they were more likely to derive the enriched interpretation when they were encouraged to spend more time processing the sentence than when they were not encouraged to do so (as in the *normal* and the *fast-word* presentation conditions).

The control sentences provide a context in which to appreciate the differences found among the *true–true or* statements. They showed that performance in the *extra-time* condition only improved with longer processing. In fact, 88% responded correctly overall to the control problems in the *fast-word* condition, 96% did so in the *normal* presentation, and 97% did in the *extra-time* condition. Yet, *true* responses to the *true–true or* condition go down consistently across these three conditions (80%, 75%, and 52%, respectively).

These results confirm two very specific predictions. First, the interpretation of "or" that first arises corresponds to the literal meaning of "or", which confirms previous findings highlighting a preference for inclusive interpretations. When not required to wait three seconds to answer, the rate of answers reflecting an inclusive interpretation was indeed quite high (80% in the *fast-word* condition and 75% in the *normal* condition). This is consistent with prior literature and is not surprising (e.g., see adults in Paris's study, 1973). Even though the *fast-word* and *normal* conditions differ markedly (as shown by rates of correct responses to controls), participants retain a strong preference for inclusive interpretations. Second, an exclusive interpretation arises as part of an effort to enrich the semantics of "or". These results are in line with the prediction according to which deeper processing yields the enriching effects. In this case, participants were evidently encouraged to apply deeper processing in analysing the test sentences when they were forced to focus on them for at least three seconds.

An alternative interpretation of the data would be to consider that the *extra-time* condition allows the participants to represent the two possible interpretations of the disjunctive utterance. Under an ambiguity account, this would amount to arguing that the two lexical entries of the word are simultaneously represented, leading the participants to uncertainty. There are theoretical arguments against this position (see Introduction), and a more precise analysis of the data also contradicts such a claim. Indeed, when presented with the *true–true or* utterances in the *extra-time* condition, 17 out of 21 participants responded consistently in a way that defies chance prediction (i.e., a participant is consistent in at least 10 out of 12 cases), and 2 others were consistent in three quarters of the cases (9 out of 12). This consistency shows that the 52% rate of inclusive responses does not reflect uncertainty or indecision within individual participants.

It is also important to point out that adopting exclusive interpretations is a matter of degree across participants. While we highlight the decline in inclusive interpretations across the three conditions, the rate does not drop below 50%. Likewise, the percentage of *true* responses in the *true–true or* condition, even in the *fast* and *normal* conditions, was never as high as the lowest rate of *correct* responses across the control conditions. A percentage of people appear to adopt the enriched interpretation even when the paradigm does not encourage them to do so (i.e., in the *fast* and *normal* conditions). By comparing the percentage of correct responses in the control conditions to the percentage of literal responses in the *true–true or* condition, it appears that some responses (anywhere from 2–24%) can be attributed to enrichments even in the *fast* and *normal* conditions (combined). What is critical is that the percentage of enrichments increases significantly (to anywhere from 40–47% of responses) once participants are assigned to the *extra-time* condition.

This indicates that the inference at stake, though easy to access, is not automatically generated. If it was automatically generated, one would expect much lower rates of inclusive answers in the *fast* condition and (especially) in the *normal* condition.

That said, we expected a difference between the *normal* and *fast–word* conditions with respect to rates of inclusive interpretations, and we did not find one (at least not a significant difference). This is arguably due to the high rate of inclusive interpretations in the *normal* condition. The rate of inclusive interpretations would have had to have been very high indeed (at ceiling) in order to have found a significant difference between the *normal* and *fast–word* conditions.

These results are readily interpretable in the relevance theoretic framework, which posits that the inference will be made in certain contexts and not others, depending on the hearer's expectations of relevance. In a context that neither mandates nor precludes an exclusive interpretation, an interpreter whose expectations of relevance are low will not make the effort of constructing an enriched interpretation. An interpreter with higher expectations of relevance might make such an effort. In an experimental setting such as the one we used, where there is no clear indication of the degree of relevance to expect, generally low expectations together with individual variations (some participants are more eager or more attentive than others) are to be expected. What is critical is that the preferred and spontaneous response is overwhelmingly inclusive in the *fast* and *normal* conditions.

Overall, these findings are consistent with the literature in two respects. One, participants are inclined to treat "or" inclusively. Two, this experiment showed that extending processing time increases the likelihood that participants will adopt enriched interpretations. Taking a complementary approach, Bott and Noveck (2004) showed that limiting available processing time results in a lower likelihood of enriched interpretations. As we wrote in the Introduction, there are many ways one can encourage the application of cognitive resources to disjunctive sentences. In the next two experiments, we explore how contrastive stress can also actively increase depth of processing.

EXPERIMENT 2

In this experiment, we investigated how an added emphasis on "or" affects its interpretation. For instance, the sentence "There is an A OR a B" ought to encourage participants to pay close attention to the disjunction. The added emphasis is indeed supposed to attract the hearer's attention and thereby to underline some implicit feature that is critical to the meaning of the sentence. If one assumes that the initial interpretation of "or" is inclusive and that added stress is an indication that the listener should apply deeper processing, it follows that there ought to be more exclusive interpretations when "or" is stressed. In other words, added emphasis on "or" ought to prompt the "not both" inference and thus an exclusive interpretation.

Four different conditions were used based on a two (written versus oral presentation) by two (emphasis or no-emphasis) design, where emphasis refers to a stress placed on "or" in the *true–true* condition. In the written condition, the emphasis was conveyed by both underlining and capitalizing "or" (OR) and in the oral condition, prosodic stress was placed on "or". Otherwise, the seven control conditions remained identical to those in Experiment 1 (without emphasis anywhere in the sentence). In the previous experiment longer reading offered the participant the opportunity to use more cognitive resources to interpret the utterance and allowed her to increase the depth with which she processed it. In this experiment, the participant is also led to use more cognitive resources insofar as the contrastive stress placed on the word "or" attracts one's attention. By definition, paying attention means processing at greater depth. We thus investigate whether the consequences for interpretation of this deeper processing is that "or" will be understood exclusively (i.e., following the enriched interpretation).

Method

Participants

A total of 75 participants were recruited from the Université Lumière Lyon 2. All were native

French speakers. Participants were randomly assigned to one of the four experimental conditions.

Stimuli and design

There were four different groups based on four sorts of presentation of the evaluation sentences. One variable was based on the modality (written vs. oral), and the other one was based on the presence versus absence of emphasis. The second factor was manipulated in the *true–true or* condition only. The remaining seven sorts of control sentences (i.e., the four sentence types in the "and" condition and the three control sentence types in the "or" condition: *false–false* and the two *true–false* cases in the "or" condition) did not carry any sort of emphasis. We chose to put an accent uniquely where the prosodic cue can affect the truth value by turning a potentially true statement into a false one (the TT condition). As far as the other conditions are concerned, a prosodic cue cannot affect the participant's applied truth value in a detectable way. In the FT, TF, and FF conditions, a contrastive stress on "or" could lead the participant to enrich its semantics towards an exclusive interpretation but there would be no way to verify this since the enrichment would make no difference to the response pattern. For instance, "There is an A OR a B" after the word "CHILD" would lead the participant to answer "False" regardless of whether she had an inclusive or an exclusive reading of the word. The structure of Experiment 2 is identical to the one in Experiment 1 except that there was only one presentation condition (the one referred to as the *normal* presentation in Experiment 1), and it was designed to determine whether added stress on the word "or" in the *true–true* condition prompts more enriched interpretations.

In order to present the task in a verbal mode, the experiment was conducted with *E-prime* (Schneider, Eschman, & Zuccolotto, 2002), which makes it possible to present sound files easily. Unlike in Experiment 1 where the software provided each participant with a unique set of stimuli, there was one set of 96 stimuli randomly distributed among eight possible conditions (two connectives: "and" and "or", and four truth conditions: *true–true*, *true–false, false–true, false–false*). Thus, each participant saw 12 experimental items for each of the eight conditions as well as 6 training items and 5 dummies.

In the *written–unstressed* condition, the presentation of the stimuli was identical to that of Experiment 1. The *written–stressed* condition differed only in the *true–true* condition in which the "or" was underlined and capitalized. The oral stimuli were recorded by the first author using Wavestudio with a standard microphone (Sony ECM 719) in the PCM format and at a frequency of 22 kHz. The stimuli were then presented to two independent judges unaware of the goal of the study. For each item, they were asked to judge whether the connective carried a stress or not. The interrater reliability was 100%. All the audio files were then normalized in intensity and length with Adobe Audition. Each sentence was followed by a 100-ms silence and preceded by a silence that was adjusted so that the total length of the file was 2,300 ms. Participants first saw a word on the screen for 2 s and then heard the target sentence through headphones (Sennheiser HD 457). Both the word and the sentence remained on the screen until the participant provided an answer. In the *oral–unstressed* condition, all sentences were pronounced using a plain tone of voice; in the *oral–stressed* condition, the "or" carried a focalization accent in the *true–true or* condition.

Results

As in Experiment 1, analyses were performed on the proportion of responses reflecting literal interpretations. A total of 7 participants were eliminated from the analysis because they gave less than 50% of correct answers in at least one of the control conditions.[9] They were replaced so that we had equal numbers of participants in all four groups. Table 4 displays the mean percentage

[9] A total of 3 participants were eliminated in the *written–unstressed* condition and 4 in the *written–stressed* condition.

Table 4. *Percentage of correct responses to each of the conditions in Experiment 2*

		A and B				*A or B*			
Presentation		*TT TABLE*	*TF JAMIS*	*FT RSOUB*	*FF POJET*	*TT TABLE*	*TF JAMIS*	*FT RSOUB*	*FF POJET*
Written	Unstressed	93	96	98	100	***81***	92	89	97
Written	Stress on OR TT	93	95	96	99	***58***	88	87	98
Oral	Unstressed	95	97	97	99	*77*	90	88	98
Oral	Stress on OR TT	95	98	98	100	*27*	89	97	99

Note: TT = true–true. TF = true–false. FT = false–true. FF = false–false. The examples all assume that the test sentence concerns the letters A and B (which is why these letters are presented in bold in the example strings). The figures in bold refer to "true" responses to the main case of interest where, for example, a participant sees the word TABLE and says "true" to the statement "There is an A or a B". The only other available response is "false", and thus false responses to TT (and other true) cases can be calculated by subtracting the figure from 100.

of logically correct responses for all conditions. As can be seen, the rate of correct responses to the control sentences was high, and there were no differences among them when comparing the four presentation conditions (*written–unstressed*, *written–stressed*, *oral–unstressed*, and *oral–stressed*).

As in the previous experiment, there were differences between the sentence types. For the "and" connective, participants found the *false–false* questions particularly simple, and most participants were at ceiling. The *true–false* and *true–true* questions were apparently more difficult but rates of correct performance were still very high overall. For the "or" connective, the *false–false* questions were again answered very accurately but there was more of a drop in performance for the *true–false* questions than there was for the equivalent cell in the "and" condition. In summary, participants had no obvious problems dealing with the oral presentation of the question, and performance was similar to that of Experiment 1.

The aim of the experiment was to examine the effect of stressing the connective on the rate of inclusive responding to "or". The fifth column in Table 4 refers to the rate of inclusive responding in the *stressed* and *unstressed* conditions. Rates of inclusive responding are similar in the *unstressed–written* and *unstressed–oral* conditions ($M = 0.81$ vs. $M = 0.77$) and do not differ from the *normal* condition of Experiment 1 ($M = 0.81$), indicating that this rate of inclusive responding is robust across different samples of participants and in different presentation formats.

As can be seen in Table 4, stressing the disjunction in these cases leads to fewer inclusive responses, as indicated by a drop from $M = 0.79$ to $M = 0.43$ overall, $F_1(1, 64) = 16.31$, $MSE = 0.315$, $p < .001$; $F_2(1, 11) = 474.88$, $MSE = 0.0039$, $p < .001$. This drop is greater when the "or" is stressed orally (50% decrease) than when it is stressed graphically (23%). The interaction is reliable in the items analysis only, $F_1(1, 64) = 1.68$, $MSE = 0.315$, $p = .20$; $F_2(1, 11) = 55.55$, $MSE = 0.0049$, $p < .001$.

We also classified each participant as being inclusive or exclusive according to the type of answer given. Thus, if more than 50% of their responses were inclusive (or exclusive), participants were classified as inclusive (or exclusive). This resulted in a pattern that mirrored that of the continuous analysis, with 14, 10, 13, and 4 participants classified as inclusive for the *written–unstressed*, *written–stressed*, *oral–unstressed*, and *oral–stressed* conditions, respectively. There were reliably fewer inclusive participants in the *stressed* than in the *unstressed* conditions, $\chi^2(1) = 10.38$, $p < .01$, and the *oral–stressed* condition differed from the others, $\chi^2(3) = 14.93$, $p < .01$. Thus, a categorical analysis shows that an oral stress on "or" is most efficient in prompting enriched interpretations.

Discussion

Experiment 2 aimed to encourage the production of enriched interpretations by prompting

participants to apply more cognitive resources to the processing of "or" through graphic or prosodic cues. To test this idea, we used two different modalities (oral and written) to determine how robust and generalizable our data were. It is interesting to note that a plain spoken voice did not produce exclusive interpretations any more than did a written statement. There is thus no main effect on the control condition based on modality. In the critical *true–true or* condition, our data show a strong effect based on the presence of both graphic and prosodic cues. Overall, these data indicate that both graphic and prosodic cues can influence the production of inferences. The effect tended to be stronger when focus was conveyed through a verbal prosodic cue rather than a graphic one but the interaction was not statistically significant. Note that the 27% rate of inclusive interpretations (resulting from the high rate of enriched interpretation production) in the *oral–accented* condition is among the lowest we have come across in laboratory tasks dealing with scalar terms. Thus, Experiment 2 revealed that prosody had a great impact on the processing of "or": A great majority of people had an inclusive reading when such a cue was absent whereas a great majority of participants had an exclusive reading when it was present.

Although the results from Experiment 2 are impressive, our hypothesis would pass a more severe test if our findings could be replicated while using a within-subject design. Indeed, if the effect we observed in the previous experiment is robust then there should be more exclusive interpretations when the "or" is accented and more inclusive interpretations when there is no accent. That is, the accent on "or" should have a direct effect on inference production among a single group of participants.

EXPERIMENT 3

Here we investigate how an added emphasis on "or" affects its interpretation within a group of participants who hear "or" both with and without an accent in the same critical *true–true* condition. If one assumes that the initial interpretation of "or" is inclusive and that added emphasis is an indication that the listener should apply deeper processing, it follows that added emphasis on "or" ought to prompt the "not both" inference and thus an exclusive interpretation. Ideally, one would find cases where participants respond differently to the two sorts of presentation. The structure of this experiment is identical to that of Experiment 2 except that we use only the oral presentation, and the critical *true–true or* condition contains both stressed and unstressed items.

Method

Participants

A total of 27 participants from the Université Lumière Lyon 2, all of whom were native speakers of French, were tested.

Materials, task, and experimental design were the same as those in Experiment 2's *oral* conditions. There were only two differences. The first difference is that among the critical 12 *true–true or* sentences that followed the five-letter-long strings, half of them (2 following words, 2 following pseudowords, and 2 following nonwords) were presented in an unstressed fashion while the other half were presented with a stress on "or". The other difference is that, unlike in Experiment 2, keys for representing *true* and *false* were varied (i.e., participants were randomly assigned to one of two sorts of key-press—having the left-finger key-press represent *true* or having the right-finger key represent *true*).

Results and discussion

As before, the control items produced high rates of correct responses. Participants responded correctly to the conjunction statements at rates that were 97–98% across the four conditions. Rates of correct responses to the three control conditions, the *true–false*, the *false–true*, and the *false–false* conditions, for the disjunction were $M = 0.88$, $M = 0.92$, and $M = 0.99$, respectively. These high rates indicate overall competence in the task

Table 5. *Percentage of correct responses to each of the conditions in Experiment 3*

	A and B				*A or B*			
	TT TABLE	*TF JAMIS*	*FT RSOUB*	*FF POJET*	*TT TABLE*	*TF JAMIS*	*FT RSOUB*	*FF POJET*
Oral presentation	98	97	98	98		89	92	99
Nonaccented					***68***			
Accented					***48***			

Note: TT = true–true. TF = true–false. FT = false–true. FF = false–false. The examples all assume that the test sentence concerns the letters A and B (which is why these letters are presented in bold in the example strings). The figures in bold refer to "true" responses to the main case of interest where, for example, a participant sees the word TABLE and says "true" to the statement "There is an A or a B". The only other available response is "false", and thus false responses to TT (and other true) cases can be calculated by subtracting the figure from 100.

and resembled those found in the oral conditions of Experiment 2 (see Table 5).

As in Experiment 2, we carried out analyses based on participants and items (as hand assignment had no effect we did not take that factor into account). The only difference with the analyses we carried out in Experiment 2 is that the items here become a within-subject type variable because the sentences were broken down into two separate groups (*stressed* vs. *unstressed*). Rates of inclusive interpretations to sentences presented without an accent on "or" (68%) were significantly higher than rates of inclusive interpretations to sentences presented with an accent on "or" (48%), $t_1(28) = 3.05$, $p < .005$, $t_2(5) = 14.77$, $p < .0001$. Thus, a plain presentation of "or" led to relatively high rates of inclusive interpretation, and a prosodic cue on "or" led to more exclusive interpretations (i.e., to more enriched inferences). This is strong evidence showing that stress on "or" can prompt an individual listener to transform the term from its initial inclusive reading to an exclusive one.

This experiment led to a replication of the results obtained in Experiment 2 while using a within-participant design. Note that the effect was diminished compared to what was obtained in Experiment 2 (Experiment 2: 77% vs. 27%; Experiment 3: 68% vs. 48%). The weakening of the effect with respect to Experiment 2 can be readily explained by the fact that participants tend to adopt an interpretation that is maintained throughout the experiment so as to remain consistent. Indeed, nearly half of the participants (13 of 29) provided exactly the same (high or low) percentage of inclusive interpretations in the *stressed* and *unstressed* conditions, and 9 others provided rates of inclusive interpretations that were very comparable (e.g., 100% inclusive in one condition and 83% in the other). This means that 22 participants adopted an interpretation that they maintained in the *stressed* and *unstressed* conditions. Interestingly, the remaining 7 participants (who did change the way they answered with respect to the way "or" was accented) provided a high percentage of inclusive interpretations in the *unstressed* condition and a relatively low percentage of inclusive interpretations in the *stressed* condition.[10] No participant provided the opposite pattern (i.e., a high rate of inclusive interpretations uniquely in the *stressed* condition).

It makes sense for the participants to choose one interpretation of "or" across the experiment in that it minimizes the amount of effort involved. Still, note that the increase in the proportion of exclusive responses in the *stressed* condition was in one direction, and the results replicated those in our between-subjects experiment, which constitutes strong evidence in favour of our hypotheses. These results show that the accent can encourage

[10] This is based on a 50% differential between the two conditions.

listeners to change their truth value judgement in the *true–true or* condition.

GENERAL DISCUSSION

The present work assessed the influence of depth of processing on the pragmatic enrichment of "or". We started by considering two theories: GCI and relevance theory, which make clear predictions concerning the processing of "or" sentences. We then focused on further testing the latter because it has received much experimental support. More precisely, we were interested in examining the factors that lead a hearer to enrich the literal reading of "or" into a more informative one—namely, an exclusive interpretation. We expected to find support for our hypotheses because our own prior experiments with logical terms, mostly with the French equivalent to the quantifier "some" ("certains"), indicated that (a) participants start off with a weak interpretation, which may be enriched into a stronger interpretation ("Some but not all"), and that (b) a forced reduction in the amount of applied cognitive resources reduces the chances that a participant produces an enriched interpretation.

In the present experiments, we manipulated different sorts of cue with the aim of affecting the amount of processing resources that a participant can apply to the disjunctive sentence (or utterance). In one condition of Experiment 1, participants were required to wait three seconds before they were allowed to answer. This was a way to give them the opportunity to spend more time evaluating the sentences and thus to look for a more informative meaning. Experiments 2 and 3 used a more natural means to have participants apply more effort. Emphasis was put on "or", either graphically (OR) or prosodically (with the focalization accent), in order to encourage participants to apply more effort to interpreting the disjunction in the utterance. Indeed, prosody functions as a natural highlighting device, which attracts the listener's attention to a specific stimulus. Following a relevance theoretic line of argument, we postulated that the conditions encouraging deeper processing of the utterance (i.e., having to wait three seconds to answer or putting emphasis on "or") would give rise to more enrichments and hence to more exclusive interpretations. This prediction was confirmed in the three experiments.

The higher number of inclusive interpretations in the neutral conditions (i.e., the *normal* and *fast-word* conditions in Experiment 1 and the *unstressed* condition in Experiments 2 and 3) favours a model according to which the literal interpretation of the connective "or" is readily available and in which the enriched interpretation results from a nonautomatic process. This view is compatible with relevance theory and a growing literature highlighting the initial preference of literal over enriched interpretations with respect to logical terms (Bott & Noveck, 2004; Breheny et al., 2006; see also Noveck, 2001, and Pouscoulous, Noveck, Politzer, & Bastide, 2007).

Thus, our data suggest that pragmatic inferences are sensitive to a range of cues and are generated when they are warranted and necessary to fulfil the listener's expectations of relevance (rather than being produced by default every time the word appears). Indeed, when engaged in the interpretation of an utterance, listeners are guided by a presumption of relevance according to which any effort should be offset by some effect. These conclusions are best accounted for within the relevance theoretic framework (rather than within the framework of GCI theory) where the interpretation based on the literal interpretation of a word (on its encoded semantics) can often be relevant enough and not require any further enrichment. It also follows that further devoting processing resources to an utterance should lead to pragmatic enrichments (due to an expectation that further processing ought to provide for a more informative interpretation). These results might appear somewhat counterintuitive to readers who feel they have an easy access to exclusive interpretations of disjunctions. To these readers, we point out that the extra work required to draw out the exclusive interpretation is relatively small and is perhaps not even accessible to consciousness.

Prosodic cues specifically have been shown to have a strong influence on pragmatic enrichments here. We employed the focalization accent as a natural means to orient the hearer's interpretation, in our case towards exclusivity, by having the participant make more effort (Wilson & Wharton, 2006). This choice was motivated by an extensive literature highlighting the role of prosody in structuring information. More precisely, it has been repeatedly argued that pitch accents increase the salience of the material they are associated with (Pierrehumbert & Hirschberg, 1990) and that this phenomenon can serve pragmatic functions (Dik, 1997; Sperber & Wilson, 1986/1995). Indeed, given that this type of accent is optional and can be suppressed without affecting the information status of the material in focus, the hearer is likely to interpret it as an attempt, made by the speaker, to increase the salience of part of the utterance (House, 2006). This will compel the listener to structure the context in such a way that this extra salience makes sense. In House's words, "the effect is to constrain access to the context within which cognitive effects will be derived" (House, 2006, p. 1549). This implies that placing an accent on "or" encourages the hearer to look for a different interpretation from the one she would have arrived at had the salience of the word not been increased. From there, we argue that stress does not itself directly encode anything but is rather "a sort of vocal equivalent to pointing, a natural means of drawing attention to one particular constituent in an utterance" (Sperber & Wilson, 1986/1995, p. 203).

One potential criticism of the current experiments is that our results will not generalize to "natural" conversational contexts. After all, the sentences used in the three experiments are arguably not heard routinely in daily conversations. Consequently, it could be argued that nonstandard sentences (designed for experimental purposes) prompt nonstandard mechanisms. We would like to make two points in response to that criticism.

First, our main goal was to test whether or not an increase in depth of processing could prompt the production of inferences; the experiments were not designed to determine in which precise conversational contexts enriching inferences appear. In order to accomplish our goal, we manipulated the amount of cognitive resources available to processing the utterance while keeping all other factors (i.e., semantics, syntax, encyclopaedic knowledge, etc.) unchanged. Our materials are thus experimentally sound precisely because they allow us to make comparisons that would not have been as fine-grained with more real-life material where such factors are hard to control.

Our second point is that claims made with more abstract materials (Bott & Noveck, 2004) were recently replicated within a more natural setting. While investigating terms like "some" and "or", Breheny et al. (2006) used contexts designed to trigger (or not to trigger) pragmatic enrichments. With respect to disjunctive sentences, they demonstrated an asymmetry between two conditions, with longer reading times when the "not both" inference was made than when it was not. Thus, they found that the last segment of the paragraph (i.e., "the class notes or the summary") was read faster in (9) than in (10).

> 9. John heard that the textbook for Geophysics was very advanced. Nobody understood it properly. He heard that if he wanted to pass the course, he should read / the class notes or the summary.
>
> 10. John was taking a university course and working at the same time. For the exams he had to study from short and comprehensive sources. Depending on the course, he decided to read / the class notes or the summary.

These results, together with ours, suggest that the mechanisms involved in understanding abstract material are quite similar to those used in everyday comprehension. Furthermore, as it has already been argued in Bott and Noveck (2004), we insist on the fact that, to the best of our knowledge, no one has ever suggested that the cognitive mechanisms involved in an experimental setting are different from those applied to ordinary verbal exchanges.

CONCLUSION

The goal of this paper was to investigate how processing effort relates to the enrichment of "or". More precisely, we tested the claim according to which increased processing effort triggers the search for a maximally informative interpretation of "or"—namely, an exclusive interpretation. That claim was supported. Compelling participants to wait three seconds before providing their answer increased their rate of exclusive disjunctive interpretations. Likewise, this effect was also found when participants apply more effort through emphasis on the word "or". This work adds to the recent, growing literature that aims to account for the conditions under which the two disjunctive interpretations arise (Chierchia, Guasti, Gualmini, Meroni, & Crain, 2004; Noveck, Chierchia, Chevaux, Guelminger, & Sylvestre, 2002), and it helps clarify the early experimental literature on disjunctions, which often reported high rates of inclusive interpretations in surprising terms (Braine & Rumain, 1981; Evans & Newstead, 1980; Paris, 1973). That exclusive interpretations were not more common in these early studies often confounded linguists and psychologists because many have the intuition that disjunctions are practically synonymous with exclusiveness. These contradictory impressions can now be squared. Indeed, a disjunctive statement is generally true when just one of the mentioned disjuncts is, making inclusivity acceptable even if there is more than one true disjunct. However, as this work shows, an exclusive interpretation of a disjunction can be made readily available as long as the listener provides a little extra (and even perhaps imperceptible) effort.

Original manuscript received 5 December 2006
Accepted revision received 25 May 2007
First published online 17 December 2007

REFERENCES

Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, *34*, 433–456.

Bezuidenhout, A., & Morris, R. (2004). Implicature, relevance and default inferences. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics.* Basingstoke, UK: Palgrave Macmillan.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Braine, M., & Rumain, B. (1981). Children's comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*, 46–70.

Breheny, R., Katsos, N., & Williams, J. (2006). Are scalar implicatures generated on-line by default? *Cognition*, *100*, 434–463.

Chierchia, G., Guasti, T., Gualmini, A., Meroni, L., & Crain, S. (2004). Adults' and children's semantic and pragmatic competence in interaction. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics.* Basingstoke, UK: Palgrave Macmillan.

Content, A., & Radeau, M. (1988). Données statistiques sur la structure orthographique du français [Special issue: Statistical data regarding the orthographic structure of French]. *Cahiers de Psychologie Cognitive*.

Dik, S. (1997). *The theory of functional grammar*. Berlin: De Gruyter.

Evans, J. St. B. T., & Newstead, S. E. (1980). A study of disjunctive reasoning. *Psychological Research*, *41*, 373–388.

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

Horn, L. R. (1984). Toward a new taxonomy for scalar inference. In D. Schiffrin (Ed.), *GURT*. Washington, DC: Georgetown University Press.

Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, *61*, 121–174.

House, J. (2006). Constructing a context with intonation. *Journal of Pragmatics*, *38*, 1542–1558.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Noveck, I. A. (2001). When children are more logical than adults: Investigations of scalar implicature. *Cognition*, *78*, 165–188.

Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning*, *8*, 297–326.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*, 203–210.

Noveck, I. A., & Sperber, D. (in press). The why and how of experimental pragmatics: The case of "scalar inferences". In N. Burton-Roberts (Ed.), *Advances in pragmatics*. Basingstoke, UK: Palgrave.

Paris, S. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, *16*, 278–291.

Pelletier, F. J. (1977). "Or". *Theoretical Linguistics*, *4*(1/2), 61–74.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). Processing costs and their impact on the development of scalar implicature. *Language Acquisition*, *14*, 347–375.

Raz, A., & Buhle, J. (2006). Typologies of attentional networks. *Nature Reviews Neuroscience*, *7*, 367–379.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford, UK: Blackwell. (Original work published 1986)

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*, 1559–1579.