Pragmatic inferences modulate N400 during sentence comprehension: evidence from picture-sentence verification

Lamar Hunt III[a,b], Stephen Politzer-Ahles[a], Linzi Gibson[c], Utako Minai[b], Robert Fiorentino[a]

[a]Neurolinguistics & Language Processing Laboratory, Department of Linguistics, Dole Human Development Center, Room 4037, 1000 Sunnyside Ave., University of Kansas, Lawrence, KS 66045

[b]Developmental Psycholinguistics Laboratory, Department of Linguistics, Dole Human Development Center, Rooms 4062 & 4066, 1000 Sunnyside Ave., University of Kansas, Lawrence KS 66045

[c]Department of Psychology, Fraser Hall, Room 426, 1415 Jayhawk Blvd., University of Kansas, Lawrence KS 66045

Address correspondence to

Lamar Hunt III

Department of Linguistics

University of Kansas

1541 Lilac Lane

Blake Hall, Room 427

Lawrence, KS 66044-3177

Tel: +1-785-864-1757

Fax: +1-785-864-5724

E-mail: lhunt3@ku.edu

Abstract

The present study examines the online realization of pragmatic meaning using event-related potentials (ERPs). Participants read sentences including the English quantifier *some*, which has both a semantic meaning (*at least one*) and a pragmatic meaning (*not all*). Unlike previous ERP studies of this phenomenon, sentences in the current study were evaluated not in terms of their truth with respect to the real world, but in terms of their consistency with a picture presented before the sentence. Sentences (such as "The boy cut some of the <u>steaks</u> in this story") were constructed such that either 1) both the semantic and pragmatic interpretations were true with respect to the preceding picture (when the boy in fact cut some but not all of the steaks); 2) neither interpretation was true (when the boy in fact cut none of the steaks); or 3) the semantic interpretation was true but the pragmatic interpretation false (when the boy in fact cut all of the steaks). ERPs at the object word, which determined whether the sentence was consistent with the story, showed the largest N400 effect for objects that made the sentence false, whereas they showed an intermediate effect for objects that made the sentence false under the pragmatic interpretation but true under the semantic interpretation. The results suggest that this pragmatic aspect of meaning is computed online and integrated into the sentence model rapidly enough to influence comprehension of later words.

Keywords: scalar implicature, pragmatics, ERP, N400, picture-sentence verification

1. Introduction

Event-related potentials (ERPs) can shed light on incremental language comprehension by revealing expectations hearers or listeners make as a sentence is unfolding. Words that are unexpected, as in the final word in the sentence "She spread the bread with socks", are known to elicit larger N400s (a negative-going centro-parietal ERP component) than expected words, as in "She spread the bread with butter" [13]. While the critical words in this example ("socks"/"butter") are expected or unexpected because of their plausibility with respect to the real world (as people do not spread bread with socks), subsequent research has shown that the N400 is also sensitive to modulations of expectancy driven by various linguistic factors (for reviews, see [12,20]; see also [16,15,14,3]). These modulations of expectancy may derive from semantic constraints (language-inherent rules about how lexical items may be composed into a sentence), discourse constraints (expectations about how an utterance relates to the surrounding context), and pragmatic constraints (expectations about the intentions of the speaker making the utterance). This study examines the role pragmatic information based on *scalar implicature* plays during incremental language processing.

Scalar implicature is a pragmatic phenomenon wherein a weak term such as ‘some’ is taken to mean that a stronger term is not true (e.g., ‘some’ is interpreted as *not all*) [9,11,18]. ‘Some’ thus has two interpretations: a purely semantic interpretation (*at least one*) and a pragmatic interpretation (*at least one but not all*). The pragmatic interpretation is cancellable: for instance, while "I had some of the cake" implies that the speaker did not have all of the cake, the speaker can then say "In fact, I had all of the cake" without contradicting her previous utterance [6]. The semantic interpretation, however, is not cancellable (a speaker cannot say "I had some of the cake; in fact, I had none of it" without contradicting herself). This dissociation between pragmatic and semantic meaning opens the possibility for pragmatic violations in otherwise semantically well-formed sentences, making it an excellent test case for examining the sensitivity of the N400 to pragmatic constraints while controlling for other semantic factors known to affect N400 amplitude.

ERP studies on scalar implicature processing have primarily focused on whether *underinformative* sentences elicit a larger N400 than sentences that are both pragmatically and semantically well-formed. An underinformative sentence is acceptable according to the semantic interpretation of the scalar term, but unacceptable according to its pragmatic interpretation. For example, the underinformative sentence "Some turtles have shells" is semantically acceptable because at least one turtle in the world does have a shell, but it is pragmatically unacceptable because *all* turtles have shells. Noveck and Posada [17] found a reduced N400 for objects in underinformative sentences compared to semantically and pragmatically acceptable sentences (e.g., "Some houses have bricks"). However, Nieuwland and colleagues [14] found that in participants with high pragmatic ability there was a larger N400 to underinformative sentences, while the same pattern that was found in Noveck & Posada [17] emerged in participants with lower pragmatic ability.

Both studies were unable to control lexico-semantic relationships across conditions, which is a factor the N400 is known to be sensitive to [12]. For example, although "Some turtles have shells" is underinformative, the word 'turtles' has a close lexico-semantic association with 'shells'. Likewise, while "Some houses have bricks" is pragmatically appropriate, the lexico-semantic association between 'houses' and 'bricks' is weaker, thus increasing the N400 amplitude of critical words in this condition. Nieuwland and colleagues [14] found that this difference was systematic in their stimuli. Hence, the reason the N400 was not sensitive to pragmatic violations in the low-pragmatic ability group might have been that, for those participants, lexico-semantic relationships are a greater determining factor on the size of the N400 than pragmatic appropriateness [14].

In the present study, we control for the effect of lexico-semantic associations by presenting a given sentence along with a visual-world context which renders the sentence True (semantically and pragmatically acceptable), False (semantically unacceptable) or Underinformative, depending on what happens in the visual context. Hence, the sentences themselves are lexically identical across conditions. The presence of an increased N400 in the Underinformative condition would therefore demonstrate that the N400 is sensitive to the scalar implicature interpretation. Moreover, given the nature of the N400 as an index of expectations about upcoming lexical items, such a finding would add to the growing body of evidence supporting the claim that scalar implicature processing influences predictions about upcoming lexical items during incremental sentence processing [14,8].

2. Materials and methods

2.1. Participants

The participants were 24 right-handed native English speakers (15 females, mean age 22.1, range 18-42), all students at the University of Kansas. They had normal or corrected-to-normal vision, and no known neurological impairments. Each participant provided informed consent to participate in the study and received $10 per hour for their participation. All methods were approved by the institutional review board of the University of Kansas.

2.2. Stimuli

We constructed 171 target sentences by varying the subject, verb and object of sentences such as the following: “The student cut some of the steaks in this story.” The object ("steaks" in the example) was the critical word. Object nouns were not repeated. Each target sentence was preceded by a unique visually-presented story that consisted of two slides, a “before” slide and an “after” slide (see Figure 1). The images used were hand-drawn and then digitized and colored in Adobe Illustrator 10.0.3. The “before” slide depicted three sets of four objects each, and an actor in the upper right corner of the screen. One of the sets corresponded to the object that would be mentioned in the following sentence (e.g., the steaks). The middle of the screen displayed a verbal description, e.g. “This is a story about a student who wants to cut steaks, brownies, and tomatoes. Let’s see what happens.” In the "after" slide, some of the objects were shown to have been affected (e.g., cut).

[Figure 1 about here]

ERPs elicited by the same word were compared across conditions. The semantic and pragmatic correctness of the word was manipulated by systematically varying the preceding "after" slides as follows. Out of the three sets of objects in each story, in one set all four of the objects were affected, in one two out of four objects were affected, and in the other none of the objects were affected; the locations of each group were counterbalanced. For instance, in the example shown in Figure 1, the True condition is the one in which two out of the four steaks are cut. The False condition is the one in which none of the steaks were cut. The Underinformative condition is the one in which all four of the steaks were cut; in this condition, the word "steaks" should be among the expected continuations if the reader interpreted *some of* as meaning *at least one*, but should not be among the expected continuations if the reader interpreted *some of* as also meaning *not all*. Note that the status of the other objects in the set (e.g., the brownies and tomatoes) does not affect the semantic truth or falsehood of the critical sentence; however, their presence and the way they are affected ensures that True, False and Underinformative continuations are possible on each trial. Each sentence appeared in all three conditions across participants in a Latin square design, creating three lists.

We measured the cloze probability of the critical words in each condition by asking 27 participants (who did not participate in the ERP study) to perform a sentence completion task using the same sentential and visual stimuli. We left a blank between “some of the” and “in this story”, and participants were allowed to fill in the blank using one or more words. The cloze probabilities, averaged across participants and items, were 89.9% for True continuations, 4.1% for Underinformative continuations, 1.8% for False continuations, 4.2% for continuations including both the True and Underinformative objects (e.g. "steaks and brownies"), 0.02% for continuations including both the True and False objects, and 0.02% for continuations including both the False and Underinformative objects.

There were 171 filler trials which shared the same basic structure as target trials, but the number of object sets in the pictures varied from one to three sets, and the filler sentences included the quantifiers *all of* and *none of* as well as *some of*. This was to reduce the proportion of trials with *some of*, the quantifier of interest. In order to have an equal number of True and non-True (False or Underinformative) sentences overall, 57 filler trials were False and 114 (including all *some of* sentences) were True.

2.3. Procedure, data acquisition, and analysis

Participants were tested in a quiet EEG room, and the task was administered using Paradigm (Perception Research Systems, Inc.). To keep each recording session as brief as possible, each participant was recorded during two separate sessions, each of which included 171 trials and lasted approximately an hour, including four breaks. The time between sessions was between 48 hours and 3 weeks. To record in two sessions, the sentence lists were each split into two sub-lists, the order of which was counterbalanced across participants. Participants were instructed to watch each story, read the sentence that followed, and press a button to decide whether the sentence "matched" what happened in the story. Each recording session began with four practice trials, none of which were Underinformative. During each trial, the "before" slide appeared for 7500 ms, directly followed by the "after" slide for 5500 ms. This was followed by the sentence, presented using rapid serial visual presentation with 300 ms for each word and 200 ms between words, a rate which pilot participants found comfortable. After the sentence ended, a prompt appeared on screen asking participants to judge whether the sentenced matched what happened in the story. This remained until participants responded or 5000 ms passed. The inter-trial interval was 500-750 ms.

The EEG was recorded continuously at a sampling rate of 1 kHz using an elastic electrode cap (Electro-Cap International, Inc.) embedded with 32 Ag/AgCl scalp electrodes arranged in a modified 10-20 layout. Electrode sites were FPZ, FZ, FCZ, CZ, CPZ, PZ, OZ, FP1/2, F3/4, F7/8, FT7/8, FC3/4, T3/4, C3/4, TP7/8, CP3/4, T5/6, P3/4 and O1/2; AFZ was used as the ground. Polygraphic channels were placed above and below each eye and on the left and right outer canthi (to monitor blinks and horizontal eye movements), and on each mastoid (the left mastoid electrode served as the online reference). Impedances for all scalp electrodes were kept below 5 kOhms, and the recordings were amplified with a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Inc.) with a bandpass of 0.1 to 200 Hz.

Trials containing blinks, horizontal eye movements or muscle artifacts within a -200 ms pre-stimulus to 1000 ms post-stimulus epoch were rejected. 17.4% of trials were excluded from further analysis, with no significant difference in the number excluded across conditions ($F(2,46) < 1$). The EEG was re-referenced offline to the average of both mastoids, epoched (from 200 ms before to 1000 ms after the presentation of the critical word), baseline corrected using a -200 to 0 ms pre-stimulus interval, and averaged. Mean amplitudes were calculated for two time windows, 250-450 ms and 450-650 ms, based on visual inspection of the waveforms and on the previous literature. For statistical analysis, six regions of interest were created by averaging the mean amplitudes of electrodes within the respective region: Left-Anterior: F3, FC3, FT7; Midline-Anterior: FZ, FCZ, CZ; Right-Anterior: FC4, F4, FT8; Left-Posterior: CP3, TP7, P3; Midline-Posterior: CPZ, OZ, PZ; and, Right-Posterior: P4, CP4, TP8. ERPs were analyzed statistically using 3 (CONDITION) × 6 (REGION) repeated measures ANOVAs; the Greenhouse-Geisser correction was applied to all effects (all effects had more than one degree of freedom in the numerator). Where significant effects of CONDITION or interactions including CONDITION were found, the conditions were compared via three pairwise *t*-tests (False vs. True, Underinformative vs. True, and False vs. Underinformative) with significance values corrected using the Bonferroni procedure.

3. Results

Acceptance rates. Participants in the ERP experiment accepted 95% of True sentences (range: 74-100%, SD: 6.3) and 9% of False sentences (range: 0-35%, SD: 7.7). Acceptance for

Underinformative sentences varied across participants (mean: 49%, range: 2-100%, SD: 42.5), with 11 participants rejecting Underinformative sentences at a rate of 80% or more, 10 accepting them at a rate of 80% or more, and only 3 showing no strong tendency. This within-participant consistency in responses to Underinformative sentences is consistent with several previous studies [1,17,19; but c.f. 5, inter alia]; we refer to participants who consistently rejected Underinformative sentences as *pragmatic responders*, and those who consistently accepted Underinformative sentences as *logical responders*.

[Figure 2 about here]

N400. Visual inspection of the ERP waveforms (Figure 2) suggests that the word in object position in False sentences elicited a larger N400 than True objects, and Underinformative objects elicited an N400 that was larger than that for True but smaller and less broadly distributed than that for False; furthermore, the effect appears to be largest in the posterior regions. Statistical analysis of the mean amplitudes in the 250-450 ms time window confirmed these impressions. The repeated measures ANOVA revealed a significant effect of CONDITION ($F(2,46) = 11.61$, $p < .001$), indicating that averaged across all regions, the N400 for False was significantly more negative than for True ($p = .002$) and for Underinformative ($p = .01$), but that Underinformative did not significantly differ from True in the whole-head analysis ($p = .113$). The ANOVA also revealed a significant CONDITION × REGION interaction ($F(10,230) = 6.33$, $p < .001$). This interaction indicated that the effect of CONDITION was significant at every region except the left anterior (ps < .004). False objects elicited significantly more negative ERPs than both True and Underinformative objects in all regions but left anterior (ps < .029), and Underinformative objects elicited significantly more negative ERPs than True objects in the midline and right posterior regions (ps < .029). Because the N400 for False objects appeared to be broader in distribution than the N400 for Underinformative objects, we re-analyzed the topographies of the mean differences using a scaling method [10], which suggested that the topographical difference was due to a quantitative change in source strength rather than to a qualitative change in the underlying sources.

Although the division of participants into subgroups composed of consistent responders results in small subgroups (n=10 pragmatic responders, and n=11 logical responders), we conducted an exploratory between-groups analysis to examine whether there are emerging differences in ERP patterns between the two groups. In this analysis, the CONDITION × GROUP interaction was significant ($F(2,38) = 5.04$, $p = .011$). Logical responders did not show any effects ($F(2,18) = 1.09$, $p = .351$). Pragmatic responders, though, showed a significant effect of CONDITION ($F(2,20) = 13.38$, $p = .001$), indicating that False objects elicited significantly more negative N400s than True and Underinformative objects (ps < .011), and Underinformative objects elicited marginally more negative N400s than True ($p = .096$). This pattern of results for the pragmatic responders illustrates that the distinction between False and Underinformative sentences observed in the overall analysis also holds for this subset of participants who do not distinguish these sentences behaviorally.

Late positivity. Although the main focus of this study was the N400, visual inspection of the ERP waveforms also suggested that False and Underinformative objects elicited more positive ERPs than True objects later in the epoch. A repeated measures ANOVA on the 450-650 ms mean amplitudes (baseline-corrected using the mean of the 250-450 ms window, to control for earlier differences in N400 amplitude [7]) revealed a significant main effect of CONDITION ($F(2,46)$, 12.80 $p < .001$), indicating that both Underinformative and False objects elicited more positive late waveforms than True (ps < .007) and did not differ from one another ($p = .102$).

There was also a significant CONDITION × REGION interaction ($F(10,230) = 6.18$, $p < .001$). The effect of CONDITION was significant at all regions (ps < .011), but whereas the positivity for False objects was significant in all regions (ps < .037), the positivity for Underinformative objects was only marginal in left anterior ($p = .097$) and significant in the other regions (ps < .014); furthermore, the positivity for False objects was marginally larger than that for Underinformative objects in the right posterior region ($p = .084$) and significantly larger in the left and midline regions (ps < .045).

The exploratory comparison between responder types for the 450-650 ms time window revealed a significant CONDITION × GROUP interaction ($F(2,38) = 6.85$, $p = .005$). Logical responders did not show a significant effect of CONDITION ($F(2,18) < 1$), whereas pragmatic responders did ($F(2,20) = 16.05$, $p < .001$). Pragmatic responders showed a more positive ERP for False and Underinformative than True objects (ps < .001), and the positivities for False and Underinformative did not differ ($p = .257$). The CONDITION × REGION × GROUP interaction was also significant ($F(10,190) = 2.60$, $p = .032$), indicating that while the above pattern was observed in all regions, the effects were smallest in the left anterior region.

4. Discussion

In order to probe whether the N400 is influenced by scalar implicature processing, the present study tested whether sentences that are underinformative with respect to a context because of a pragmatic inference elicit an N400. Objects in both Underinformative and False sentences elicited a biphasic N400-late positivity pattern, which was larger for False than Underinformative. In the present study, the sentences were lexically identical across conditions, and thus only differed in terms of their felicity and truth relative to the preceding visual context.

Therefore, the N400 in the Underinformative condition is not due to lexical differences, but to the violation of an expectation based on a scalar inference. This suggests that the pragmatic interpretation of ‘some’ influenced the incremental interpretation of the sentence, and that the N400 reflects incremental utilization of scalar implicature information [see also 14].

This conclusion is independent of whether the N400 indexes lexical access or integration of the lexical item with the overall meaning of the sentence [12, for review of these theories]. On the integrative account, because no correct representation of the context can be constructed using the False sentence, critical words in the False condition are the most difficult to integrate into the meaning of the sentence. Thus, a larger N400 for False compared to Underinformative and True is predicted. The larger N400 for Underinformative compared to True can also be explained if constructing a logically true but pragmatically infelicitous representation also causes integrative difficulty. On the lexical access account, because the critical word in the False condition is the least expected, a larger N400 is predicted for False compared to True and Underinformative. Moreover, although the pragmatic violation engendered in the Underinformative sentence rules out critical words in this condition as the most likely candidate, because the pragmatic interpretation is cancellable, a continuation with the critical word in the Underinformative condition is not completely unexpected and thus activation for the Underinformative object noun may be maintained to some degree. Thus, the conclusion of this study—that the N400 is sensitive to the incremental deployment of pragmatic information—does not depend on one particular view of the N400.

The 450-650 ms time-window showed positivities for Underinformative and False objects compared to True objects, consistent with the timing and topography of the P600. Previous research has also reported positivities appearing after N400s elicited by unexpected words [2,4]. It has been hypothesized that this effect is due to an expectancy disconfirmation process associated with high constraint violations (i.e., words with a low cloze probability in a context in which the number of plausible continuations is minimal). Given that the contexts tested in this study involved a small number of plausible continuations, the late-emerging positivities in the current study may be functionally linked to those reported in these previous studies.

In summary, this study demonstrates that when lexical differences are controlled for, N400 amplitudes differentiate between words that make a sentence true with respect to a context, those that make a sentence false, and those that make a sentence semantically true but pragmatically underinformative. This finding suggests that the processes reflected by the N400 are sensitive to information garnered from scalar inference, as well as information garnered from semantics. This methodology, using a visual-world context combined with EEG, presents a way forward in examining scalar implicature in various conversational contexts, toward adjudicating among alternative theories of scalar implicature processing [11].

Acknowledgements

We thank Angela J. Chon for assistance in stimulus development, and the University of Kansas Honors Program for funding support.

References

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

[2] K.A. DeLong, T. Urbach, D. Groppe, M. Kutas, Overlapping dual ERP responses to low cloze probability sentence continuations, Psychophysiology. 48 (2011) 1203-1207.

[3] K. DeLong, T. Urbach, M. Kutas, Probabilistic word pre-activation during language comprehension inferred from electrical brain activity, Nature Neuroscience. 8 (2005) 1117-1121.

[4] K.D. Federmeier, E.W. Wlotko, E. De Ochoa-Dewald, M. Kutas, Multiple effects of sentential constraint on word processing, Brain Research. 1146 (2007) 75-85.

[5] A. Feeney, S. Scafton, A. Duckworth, S.J. Handley, The story of some: Everyday pragmatic inferences by children and adults, Canadian Journal of Experimental Psychology. 54 (2004) 128-133.

[6] H.P. Grice, Logic and conversation. In: Studies in the way of words, Harvard University Press, Cambridge, MA, 1989, pp. 22-40.

[7] P. Hagoort, Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations, Journal of Cognitive Neuroscience. 15 (2003) 883-899.

[8] J.K. Hartshorne, J. Snedeker, The speed of inference: Evidence against rapid use of context in calculation of scalar implicatures, submitted.

[9] L.R. Horn, On the Semantic Properties of Logical Operators in English, Ph.D. dissertation, University of California, Los Angeles, United States – California, 1972.

[10] H. Jing, R.T. Pivik, R.A. Dykman, A new scaling method for topographical comparisons of event-related potentials, Journal of Neuroscience Methods. 151 (2006) 239-249.

[11] N. Katsos, C. Cummins C, Pragmatics: from theory to experiment and back again, Language and Linguistics Compass. 4 (2010) 282-295.

[12] M. Kutas, K.D. Federmeier, Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP), Annual Review of Psychology. 62 (2011) 621-647.

[13] M. Kutas, S. A. Hillyard, Reading senseless sentences: Brain potentials reflect semantic incongruity, Science. 207 (1980) 203-205.

[14] M. Nieuwland, T. Ditman, G. Kuperberg, On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities, Journal of Memory and Language. 63 (2010) 324-346.

[15] M. Nieuwland, G. Kuperberg, When the truth is not too hard to handle: An event related potential study on the pragmatics of negation, Psychological Science. 19 (2008) 1213-1218.

[16] M. Nieuwland, J.J.A. Van Berkum, When peanuts fall in love: N400 evidence for the power of discourse, The Journal of Cognitive Neuroscience. 18 (2006) 1098-1111.

[17] I.A. Noveck, A. Posada, Characterizing the time-course of an implicature: An evoked potentials study, Brain and Language. 85 (2003) 203-210.

[18] I.A. Noveck, D. Sperber, The why and how of experimental pragmatics: The case of 'scalar inferences'. In: N. Burton-Roberts (Ed.), Advances in pragmatics, Palgrave, Basingstoke, 2007, pp. 184-212.

[19] E.M. Tavano, The balance of scalar implicature, Ph.D. dissertation, University of Southern California, Los Angeles, United States – California, 2010.

[20] J.J.A Van Berkum, The neuropragmatics of 'simple' utterance comprehension: An ERP review, in: U. Sauerland, K. Yatsushiro (Eds.), Semantics and Pragmatics: From Experiment to Theory, Palgrave Macmillan, Basingstoke, 2009, pp. 276-316.

Figure captions

Figure 1: A sample set of pictures for the target sentence "The student cut some of the steaks in this story". A participant first saw the "Before" slide depicted in the upper left, along with the text depicted in the top right (presented at the center of the slide during the experiment). A participant next saw one of the three pictures in the bottom row, followed by the sentence. This sentence is True when following the bottom left picture (in which some but not all of the steaks were cut), Underinformative when following the bottom center picture (in which all of the steaks were cut), and False when following the bottom right picture (in which none of the steaks were cut).

Figure 2: Grand average ERPs and acceptance rates for the three sentence types. The leftmost column shows grand average waveforms (with a 30 Hz low-pass filter applied for visualization purposes) at electrode Pz for all participants (Figure 2A), logical responders only (Figure 2B), and pragmatic responders only (Figure 2C). The center column shows topographical maps of the effects for the False and Underinformative conditions, relative to the True condition, averaged over the 250-450 ms window. The rightmost column shows the acceptance rate for sentences in each condition.

Figure 1 (Stimulus)

Click here to download high resolution image

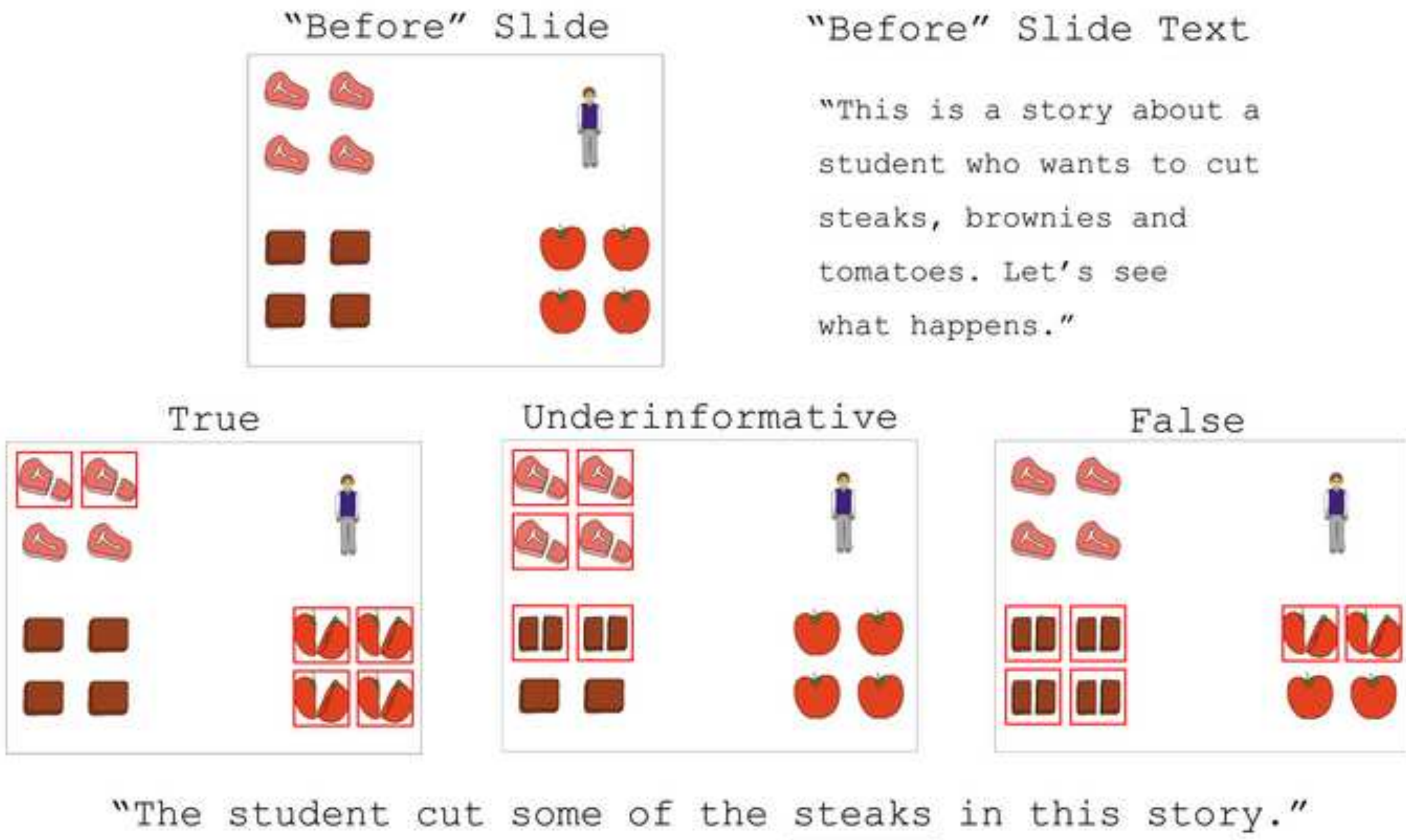

Figure 2 Color (ERP)
Click here to download high resolution image