Nothing but the truth, but not the whole truth: Adults choose to mention agents and patients in proportion to informativity, even if it doesn’t fully disambiguate the message.

Melissa Kline

Laura Schulz

Ted Gibson

**Abstract**

How do we decide what we choose to say to ensure our meanings will be understood? The Rational Speech Act model (RSA, Frank & Goodman, 2012) asserts that speakers plan what to say by comparing the potential informativity of words in a particular context. We present the first example of an RSA model of *sentence level* (who-did-what-to-whom) meanings. In these contexts, the set of possible messages must be abstracted from the set of common-ground entities (people, such as *Jane, Marco;* objects such as *apple, banana*) to possible events (*Jane eats the apple*, *Marco peels the banana*), with each word in a message contributing unique semantic content. How do speakers accomplish the transformation from context to possible compositional messages? In a communication game, participants were required to describe transitive scenes (e.g. *Jane pets the dog*), with only two words, in contexts where those two words either were or were not enough to uniquely identify an event. Adults made utterance choices sensitive to informativity in these contexts, matching the predictions of the RSA even when there was no fully ‘correct’ or successful choice. Thus we show that adults’ communicative behavior can be described by a model that accommodates the active construction of possible messages from a given discourse, beyond the set of possible entities in common ground. By presenting the first case of RSA modeling of basic who-did-what-to-who argument structure, this study suggests that full-blown natural speech may indeed result from speakers who model and adapt to the listener’s needs.

**Introduction**

Communication requires continually making decisions about what information to include and exclude. It is not always necessary or desirable to fully describe an event: if someone asks *What are you doing*?, *I’m eating* might be sufficient, and possibly preferable to longer alternatives like *I’m eating a sandwich* or *I’m eating a grilled cheese sandwich*. For a speaker to successfully communicate with a listener in this way, the two need to implicitly agree on some shared principles of communication. Grice (1975) codified these conversational assumptions as a series of ‘maxims’, including the maxims of Quantity (‘give as much information as is needed, but no more’) and Relevance (‘say something that furthers the goal of the conversation’). Thus a speaker can refer to *a sandwich* alone if the alternative is a salad, but should refer to *a grilled cheese sandwich* if the alternative is peanut butter and jelly.

Adults use statistical information to predict words and structures (Levy, 2008; MacDonald, 2013; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and in turn many features of language production also seem to be shaped to improve the chances of successful communication. Formalizations based on information theory (Shannon, 1949) have been given to explain a variety of effects in natural production including phonological reduction, lexical choice (e.g., between *math* and *mathematics*) and inclusion of optional arguments (Aylett & Turk, 2004; Jaeger, 2010; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Resnik, 1996; van Son & van Santen, 2005). However, people might produce helpful features of speech without intending to be helpful, just because those features are easy to produce (cf. Keysar, Barr, & Horton, 1998). To tell whether production is intentionally helpful, the speaker should flexibly adapt to context when the situation changes. A large body of work has focused on how referring expressions for objects (*that, that big sandwich, that grilled cheese sandwich*) are related to the specific nonlinguistic information available in a scene (e.g. Brennan & Clark, 1996; Brown-Schmidt & Tanenhaus, 2008; Nadig & Sedivy, 2002; Pogue, Kurumada, & Tanenhaus, 2016; Sedivy, 1999). As listeners, adults assume that speakers are making informative choices (e.g. by directing their attention to a small cup with a larger competitor rather than a lone small plate, upon hearing *The small…*) Are they correct? Frank & Goodman (2012) developed a model of communication based on a speaker-listener pair with two key dynamics: listeners succeed by assuming that the speaker will maximize the information transferred, and the speaker assumes that the listener will interpret utterances in this ‘smart’ way. This model captures a wide variety of phenomena in pragmatics, particularly on the side of comprehension, predicting the interpretation of scalar implicatures, hyperbole, and metaphor, among others (Goodman & Frank, In Press; Goodman & Stuhlmüller, 2013; Kao, Levy, & Goodman, 2013; Kao, Wu, Bergen, & Goodman, 2014). A smaller body of experimental work has addressed the extent to which speakers actually are as rational as listeners assume. For instance, in a context with a *blue circle* as a target with a *blue square* and a *green square* as distractors, adults limited to a single word produce CIRCLE to identify the target shape, not BLUE or CIRCLE (both good descriptions of the target in isolation) at random (Frank & Goodman, 2012).

Of course, human language is much more complicated than this task in many respects. The prototypical sentence identifies not an object but a proposition about the world, and deriving this set of possible propositions from a conversational context may be complex. Once a proposition has been chosen, we have many choices about how to encode it in a sentence, including argument structure and verb choice (*she ate it/she put it in her mouth*) in addition to choices about how to refer to the individual arguments. Do we convey sentence level meaning using something like the rational speaker model? We focus on a class of basic propositions about events, transitive sentences like *John feeds the dog*. Verbs vary in the kinds of entities can appear as the subject or object: *I eat* \_\_\_\_\_\_\_ will almost certainly be followed by a food (relatively high predictability) while *I see \_\_\_\_\_\_\_\_* has few selectional restrictions. Resnik (1996) modeled argument inclusion in terms of this kind of conceptual content, showing that the tendencies of verbs to include or omit objects could be explained by the selectional restrictions on their arguments. Thus, distributional patterns of language use will impact the informativity of a particular verb-argument pair.

However, little attention has been paid to the role of *nonlinguistic* information on the informativity of sentences. A key challenge for understanding referential contexts for who-did-what-to-whom event information is a parallel to the problem in verb acquisition: events are transitory, and unlike object-noun labeling, references to events often take place when the event itself is in the past or future (Gleitman, 1990). As with referring expressions for objects, participants must determine what is necessary to say to pick out a referent event from a set of possibilities. Even assuming the speaker is referring to an event that might occur in the immediate context, deriving the set of possible messages requires several steps. A speaker trying to design an informative event utterance must consider not only the set of possible verbs, but also what could correspond to each argument position (agent, patient); Thus the size of the context set of events is the product of the possible verbs, subjects and objects, e.g.:

(4) {John, Sue, George, Maria, Jenny} x {feeds, chases, pets} x {the dog, the cat} = 30 events

We use this logic to create a ‘toy’ world in which there are always exactly seven referents (people and objects), and the messages to be communicated are interactions between these entities (e.g. *John feeds the dog*). Critically, we manipulate the amount of information that can be conveyed by subjects, verbs, and objects by altering the makeup of these seven referents. In object reference studies, a noun phrase like *my sandwich* or *my grilled-cheese sandwich* is assumed to be informative if and only if it uniquely identifies one out of several referents in the context, under-informative if it could apply to more than one object, and over-informative if it includes additional modifiers (*my grilled cheese sandwich* in a context with only one sandwich). RSA models assume a richer sense of ‘informativity’ in which words are informative to the extent that they reduce the number of possible interpretations by any amount.

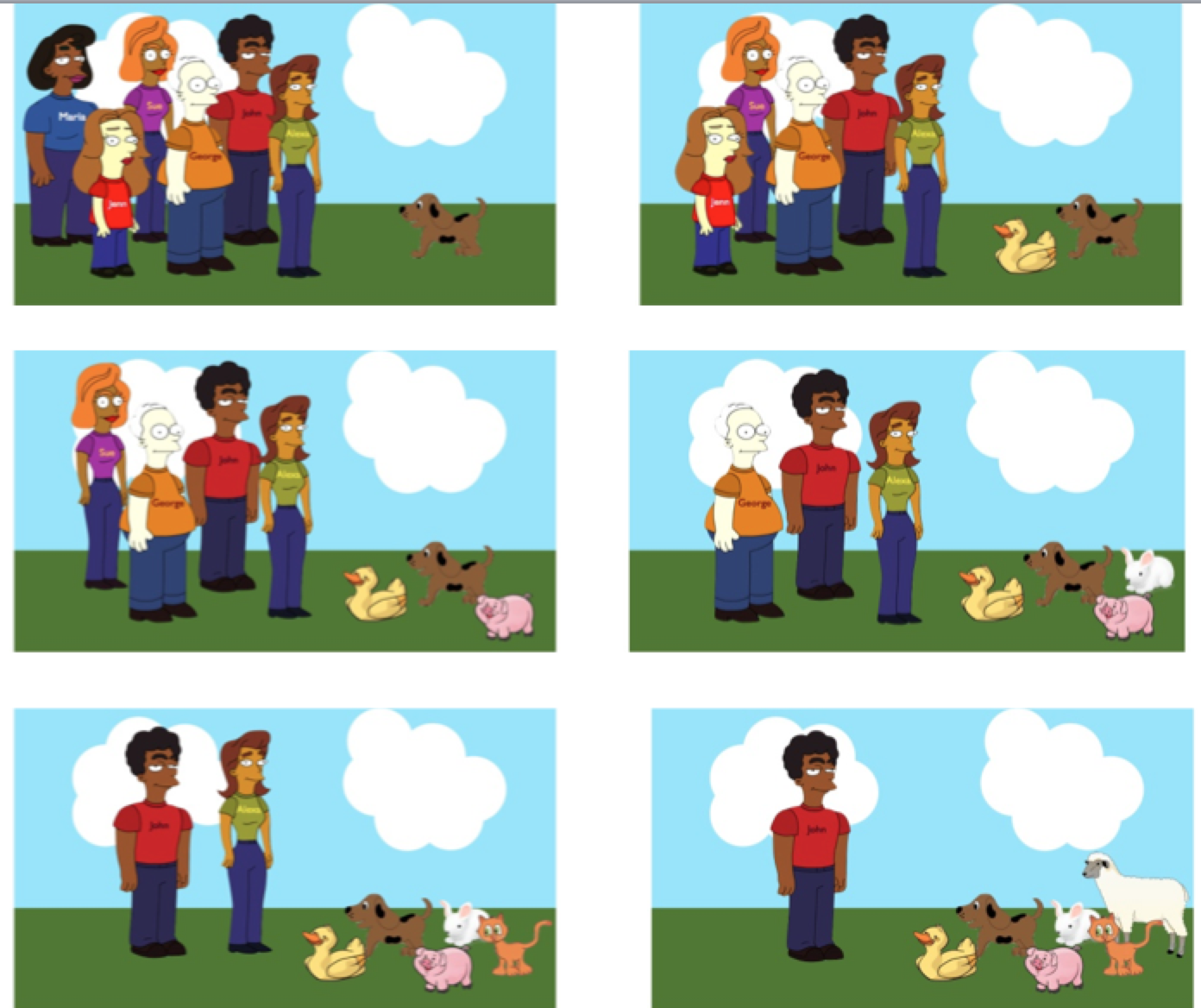




Figure 1: Context and event images for JOHN FEEDS THE DOG.

By manipulating the makeup of the people/object set, we can ask whether choices about what to say are informative even when a listener would be unable to identify a single correct choice of meaning. To do this, we use a production task that restricts the producer to just two words, focusing on the broadest reference selection choice of whether to mention an element at all. Figure 1 shows the full set of conditions. For the top left array, FEED DOG fails to resolve the ambiguity; on the other hand, JOHN FEED eliminates the ambiguity by specifying the agent and relying on an intelligent listener to identify the unique patient in context. However, in the bottom right array the reverse is true: FEED DOG becomes the sentence that can resolve ambiguity. In the remaining arrays, there are multiple potential agents and potential patients, but different words reduce ambiguity to different degrees: in the top right array, mentioning John narrows down the possible events to just two alternatives (he feeds the dog or duck) rather than five (somebody feeds the dog). If the RSA model extends to descriptions of argument structure relations, adults should be able to select informative arguments: when there are more agents than patients, participants should be more likely to mention subjects, even if ambiguity between multiple messages remains. However, if participants use a simpler strategy of determining just whether or not a given utterance successfully conveys the intended event, then they should still choose informative arguments in the deterministic cases, but perform at chance when both arguments remain ambiguous.

**Methods**

Participants91 English-speaking adults participated on Amazon’s Mechanical Turk (AMT). Participants were screened to be located in the United States and self-reporting English as their first language (an additional 21 participants were excluded who did not meet these criteria). No other demographic information about participants was collected. The task took approximately 13 minutes to complete and participants were paid $1.00. All participants gave informed consent in accordance with requirements of the MIT institutional review board.

Stimuli We created cartoon stimulus sets for each of twelve verbs (*eat, feed, hold, drink, kick, drop, wash, pour, throw, touch, read*, and *roll*). Each set consisted of an action picture, and six possible context pictures showing possible agents and patients (see Figure 1). The people were generated using a character-creation website (Brooks et al., 2007) with distinct features and names on their shirts. The objects were members of an appropriate category (e.g., various foods). The total number of agents and patients in each context summed to 7: (1 agent, 6 patients, abbreviated 6\_1), (2 agents, 5 patients, 2\_5), … (6 agents, 1 patient, 6\_1). All stimuli, code, and analyses are available at www.github.com/mekline/Subject-Drop/.

ProcedureStimuli were presented using Python and the EconWillow package (Weel, 2008), accessed through AMT. People were told that they were providing descriptions for another participant:

*In this task, you will be describing a series of cartoon scenes for another participant. You can only send two words to your partner at a time, but you must still try to convey what is happening in each scene. For each scene, your partner will be able to see an initial set of possible participants (the presentation scene). However, they won't be able to see the actual action that takes place or the description sentence ("The monkey smells the orange.") Your job is to let them know what happens in the action scene using just two of the words from the description. Again, you can use only two words to respond.*

Each participant saw twelve trials in a random order, with two items at each context level (e.g., 6\_1, 3\_4). On each trial, participants saw the context display for ten seconds, read the sentence corresponding to the action they would see (e.g., “John feeds the dog”), and then saw the corresponding action picture for 10 seconds. Finally, the context picture reappeared and participants were asked to provide a two-word description of the event. Participants were given two text boxes in which to enter their response. If they entered more than one word in each (screened by checking for spaces, e.g., “baby rolls”), they were told to try again. After each trial participants saw a progress bar, the time it had taken them to enter a response on that trial, and their average speed so far. This was intended to encourage participants to answer as quickly as possible. After seeing all twelve trials, participants were told the purpose of the study and given a unique code to receive payment.

**Results**

Responses were first checked for minor variations such as capitalization and verb form (e.g., “Eaten” was coded as “eat”). The majority of responses (76%) consisted of two of the possible three content words in the sentence. Because all of the potential agents that participants saw were humans, and all of the potential patients belonged to the same superordinate category, participants sometimes used general descriptors or pronouns (e.g., *Man, she,* or *fruit*). In addition, some participants gave responses that were intended to convey a more general idea of the scene (e.g., *Man hungry*). To avoid unevenly applying exclusion decisions in different conditions, words were coded as referring to the subject or object only if they uniquely identified the referent out of the whole set of 6 (e.g., “woman” was not coded as referring to the subject of the sentence “*Mary ate the orange*” even if she was the unique female). Likewise, words were coded as verbs only if they were a form of the verb itself (e.g., *eaten*) but not if they referred to a related idea (*hungry*). Since not all answers included exactly two codable words, we analyze data for the presence of subjects, objects, and verbs in each two-word response.

The effect of array type on whether participants mentioned the subject in their response was highly significant by a mixed-effect logistic regression with random slopes and intercepts for both item and participant (X2 = 12.3, df= 8, p<0.001). The same was true for objects (X2 = 17.7, df= 8, p<0.001). These patterns are as predicted – as more agent distractors (and thus fewer patient distractors) were present, participants were more likely to mention the subject and less likely to mention the object. We also found that participants overall were somewhat more likely to mention objects than subjects: on the subset of trials (73%) where participants mentioned only one of the two, there were significantly more objects than subjects (p < 0.001, binomial test).

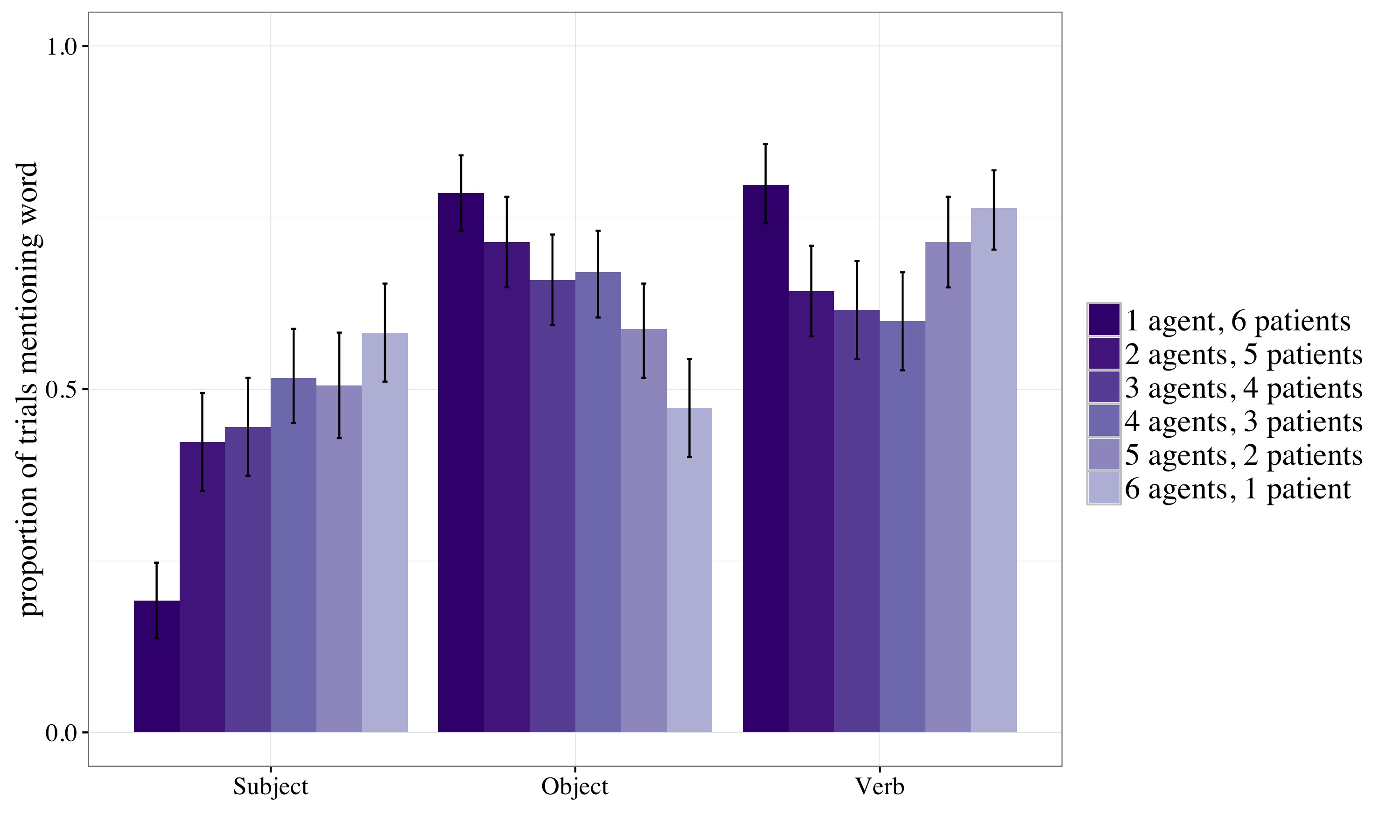


Figure 2: Trials on which participants included subjects, objects, and verbs. Error bars represent 95% bootstrapped confidence intervals.

To test whether participants in fact gave graded responses to the intermediate arrays (e.g., 2\_5), we also examined the effects of array type after removing trials for which a ‘deterministic’ answer could be given (6\_1, 1\_6.). The effects of array type on both subject and object mention were both still significant (Subject mention: X2 = 6.17, df= 8, p<0.05; Object mention: X2 = 4.6, df= 8, p<0.05).

**Model Comparisons**

To evaluate how human performance might reflect pragmatic choices, we compared four possible computational models. First, we considered a non-graded pragmatic account where participants check whether any of them fully identify the target event in context, or not. Second, we adapted the RSA model elaborated in Frank & Goodman (2012) to account for referents and context sets consisting of events rather than objects. Finally, we considered two models that reflect the base rates of subjects, verbs, and objects, one with no pragmatic sensitivity and an RSA model with differing prior preferences for word types. In all models, the shared context is the set of possible events that might occur given the set of agents and patients, and plausible verbs (We assume the prior probability of picking any particular event *e* in *E* is uniform). We assume that each object/person in the scene is classified unambiguously as an agent or patient (wrong in general, but plausible in our experimental context). For the verbs, we assume that participants are considering some set of possible interactions between the agents and patients (e.g., petting, feeding) that both they and the listener have in mind. In principle, the notion of ‘possible verb set’ could be estimated empirically by asking naïve participants to list possible actions between the agent-patient stimuli sets directly. Here, we assume the set is relatively small and does not vary with the number of agents and patients[[1]](#footnote-1) Thus, the shared event context E is

(5)

Next, we consider a set of possible descriptions (*d*) that could be used for some target event *e*. Each of these descriptions might also apply to other events in the possible set; the number of events some description *d* can refer to is notated as |*d*|. We assume that a single unambiguous word refers to each member of the agent-patient-verb triple, and that that both speakers and listeners know that only two-word utterances will be considered. For single-word descriptions, |*d*| can thus be defined easily: the word referring to the agent of *e* can refer to any of the *j \* k* pairings of patients and verbs in E that include the agent. Under alternate generative models, we can thus calculate the probability of a single word description, which we will simplify as p(A). For two-word utterances, words are *not* chosen independently (we assume the speaker won’t say the same word twice); thus for instance the probability of (unordered) **AV** can be calculated as below:

(7)

With these common assumptions, we next describe how single word description probabilites are generated, and then graphically display and evaluate the resulting predictions.

**Pragmatic ‘succeed/fail’ heuristic**: Many common-ground type experiments (e.g. Sedivy 1999) tacitly assume that an utterance is pragmatically helpful if and only if it uniquely identifies the target referent. We assume the verb is always mentioned since it is not depicted in the context, meaning that the possible utterances are (unordered) **AV** or **VP**. The probability of each utterance is thus given by:

(8)

Thus, if there is a single informative choice, the model will select it approximately deterministically,

(9)

but if neither utterance is fully informative (or if both are), the two utterances will be produced at chance,

(10)

**Rational speaker**: We implement Frank & Goodman’s RSA model, which states that a description *d* will be chosen in inverse proportion to how many events that description can apply to:

(11)

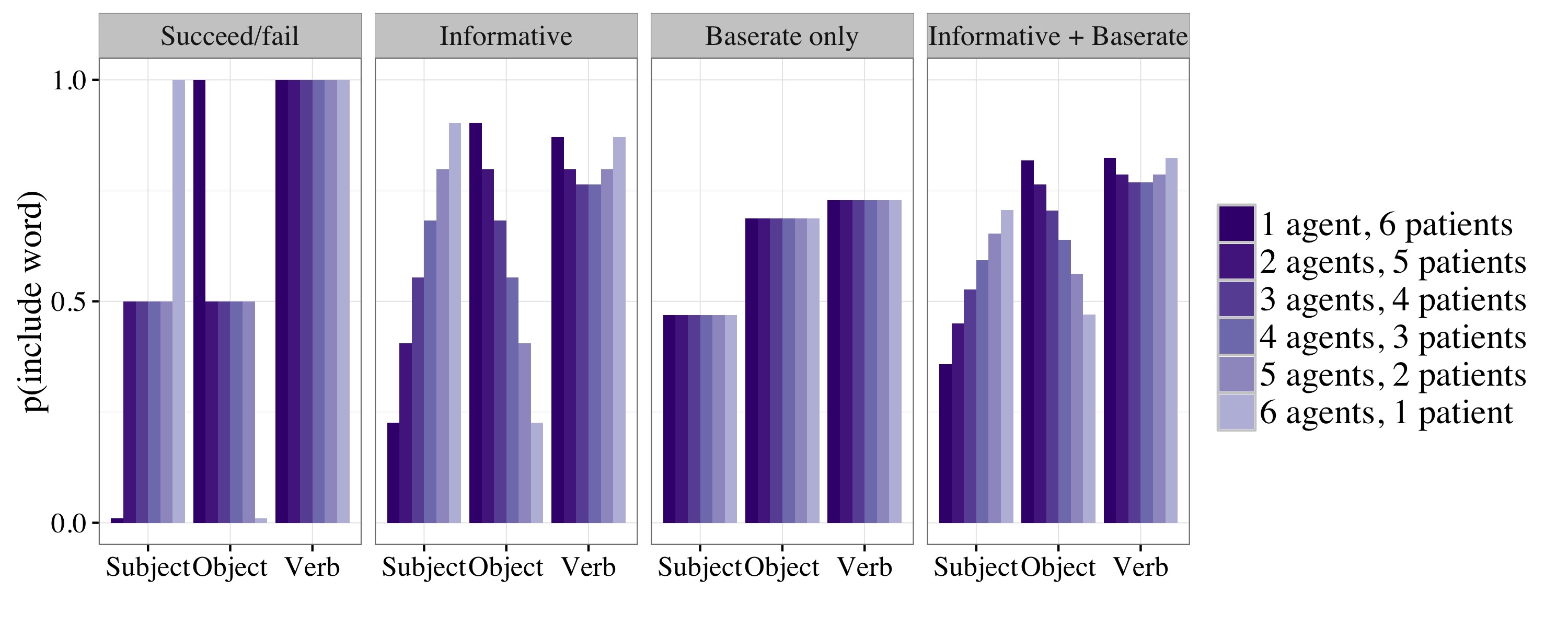
**Baseline model**: In our dataset, we found an overall difference in the frequency with which subjects, objects, and verbs are produced. We thus consider a model that emits this baserate with no pragmatic considerations as a comparison. The probability of each one-word description *d* is simply the base likelihood:

(12)

**Rational speaker, base-rate adjusted:** Because the base-rate model is fitted to this particular data set, and because speakers may plausibly have both baseline and context-specific preferences, we consider a version of the pragmatic model with word-type biases. This is constructed by multiplying the prior probability and the probability under the rational speaker model:

(13)

To parallel the data analysis conducted on the human data, we calculate the probability that each model chooses to include the agent, verb, or patient in a two-word utterance. Because some models involve a parameter estimated from the dataset, we randomly divided the human data into two halves to avoid overfitting and calculated the average likelihood of including subjects, verbs, and objects across context settings. Where there are no data-based parameters we simply compared the model predictions to both halves; when parameters were estimated from the data we used one half of the data to fit the parameters and tested correlations on the other half, repeating this process with the halves reversed. The predictions of each model are shown in Figure 3. The model that considers only base rate performs relatively poorly (r(36) = 0.66) [[2]](#footnote-2), and both the ‘succeed/fail’ model and the rational speaker model provide moderately good fits to the data (r(36) = 0.70 and 0.74, respectively). The closest fit to the human data was found with the baseline-adjusted rational speaker model (r(36) = 0.88).

Figure 3: Predictions of four possible generative models for including Subject, Object and Verb in event descriptions.

**Discussion**

As predicted, when participants described events after seeing arrays of possible agents and patients, their two-word answers reflected the degree to which a given word could convey new information about the event: participants were more likely to mention the subject of the event when the subject was more ambiguous, and more likely to mention the object when the object was more ambiguous. As predicted under an RSA account, this ability was not limited to cases where an event could be uniquely identified: even for the intermediate cases where there were multiple agents *and* patients in the array, participants still tended to choose the two-word sequence that most reduced uncertainty. Quantitative comparison to the RSA reveals a close fit to human data, with a baseline-adjusted version of the RSA performing best, matching the human tendency to produce more objects than subjects.

A paradox of RSA models is that while comprehenders of language assume they are listening to rational speakers, the same people turn around to produce messy, sometime under- or over-informative utterances. Nevertheless, we mainly succeed in getting our meanings across, and it is clear that at least some aspects of adult speech are well designed for robustly transferring information. While there is a rich literature on how speakers do (or don’t) accommodate context when describing individual objects, this study provides the first evidence for audience design in at the level of arguent structure relations. Although the two kinds of shortened sentences (Subject-Verb, e.g., GIRL READ, and Verb-Object, e.g., READ BOOK) are on average equal in length and express the same amount of information, participants recognize that informativity depends on the set of possible alternative events. This holds even when either utterance will leave some ambiguity, suggesting that RSA-type listeners are correct: speakers are choosing what to say and what to omit in a way that maximally reduces uncertainty.

Understanding how listeners and speakers represent contexts and possible messages for verbs and events is a puzzling problem. In noun-referent studies, participants (listeners or speakers) need simply note the number of entities in contexts and the features that discriminate between them (e.g. Stiller, Goodman, & Frank, 2015). For sentence level meanings, the set can be much larger than the number of visible referents: when there are three potential agents and four patients, there are twelve possible events, more if there multiple verbs are consideration: the listener might have to guess at likely relations between *girl* and *apple*, and even if the action is ongoing there may be multiple choices: verbs refer to particular event perspectives rather than pointers to actions (e.g., a girl *swinging a bat* and *hitting a ball towards the outfielder*; cf. Gleitman, 1990; Kline, Snedeker, & Schulz, 2016). These perspectives might differ in argument structure; a listener might need to consider multiple argument sets: an agent and patient, an agent, theme, and recipient, and so on. Furthermore, in the real world many referents, especially humans, can play many roles (e.g., agent and patient of *hugging*), and some possible referent pairs will permit different interactions due to either selectional restrictions or real-world knowledge. We may be able to use the current paradigm to address features of argument structure communication like these: if a speaker learns that *wugging* can be performed by animals but not people, will they take this information into account when designing utterances for a partner who does or doesn’t know this restriction? How far do parallels between messages about object identity and propositions about the world (event descriptions) extend? Which of the complexities of sentence-level predictability do speakers and listeners fold into their models of communicative context? Understanding the dynamics of utterance production in these contexts will further our understanding of how even adults calculate and use informativity to accomplish our communicative goals.

**References**

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Brooks, J., Groening, M., Jean, A., Scully, M., Sakal, R., & 20th Century Fox Home Entertainment. (2007). *The Simpsons movie (Website)*. Retrieved from http://web.archive.org/web/20120317090828/http://www.simpsonsmovie.com

Brown-Schmidt, S., & Tanenhaus, M. (2008). Real-Time Investigation of Referential Domains in Unscripted Conversation: A Targeted Language Game Approach. *Cognitive Science: A Multidisciplinary Journal*, *32*(4), 643–684. https://doi.org/10.1080/03640210802066816

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Goodman, N., & Frank, M. C. (In Press). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*.

Goodman, N., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184. https://doi.org/10.1111/tops.12007

Grice, H. (1975). Logic and Conversation. *Syntax and Semantics*, *3*, 41–58.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Kao, J. T., Levy, R., & Goodman, N. D. (2013). The funny thing about incongruity: A computational model of humor in puns.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007. https://doi.org/10.1073/pnas.1407479111

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The Egocentric Basis of Language Use: Insights From a Processing Approach. *Current Directions in Psychological Science*, *7*(2), 46–50. https://doi.org/10.1111/1467-8721.ep13175613

Kline, M., Snedeker, J., & Schulz, L. (2016). Linking Language and Events: Spatiotemporal Cues Drive Children’s Expectations About the Meanings of Novel Transitive Verbs. *Language Learning and Development*, 1–23. https://doi.org/10.1080/15475441.2016.1171771

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of Perspective-Taking Constraints in Children’s On-Line Reference Resolution. *Psychological Science*, *13*(4), 329–336.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-Specific Generalization of Pragmatic Inferences based on Under- and Over-Informative Prenominal Adjective Use. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02035

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, *61*, 127–159.

Sedivy, J. (1999). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*.

Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, *37*(1), 10–21. https://doi.org/10.1109/JRPROC.1949.232969

Stiller, A. J., Goodman, N., & Frank, M. C. (2015). Ad-hoc Implicature in Preschool Children. *Language Learning and Development*, *11*(2), 176–190. https://doi.org/10.1080/15475441.2014.927328

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*(1–2), 100–123. https://doi.org/10.1016/j.specom.2005.06.005

Weel, J. (2008). *Willow, a Python library for economics*. Retrieved from http://econwillow.sourceforge.net/

1. We tested just two values for the number of verbs (*k*): 5 and 50); we use *k*=5 in all models/graphs. The effect of increasing *k* is to increase the likelihood of including the verb in an utterance. [↑](#footnote-ref-1)
2. All model comparison p values are p < 0.0001 [↑](#footnote-ref-2)