Rational integration of noisy evidence and prior semantic expectations in sentence interpretation

Edward Gibson[a,b,1], Leon Bergen[a], and Steven T. Piantadosi[c]

Departments of [a]Brain and Cognitive Sciences and [b]Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]Department of Brain & Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268

Sentence processing theories typically assume that the input to our language processing mechanisms is an error-free sequence of words. However, this assumption is an oversimplification because noise is present in typical language use (for instance, due to a noisy environment, producer errors, or perceiver errors). A complete theory of human sentence comprehension therefore needs to explain how humans understand language given imperfect input. Indeed, like many cognitive systems, language processing mechanisms may even be "well designed"–in this case for the task of recovering intended meaning from noisy utterances. In particular, comprehension mechanisms may be sensitive to the types of information that an idealized statistical comprehender would be sensitive to. Here, we evaluate four predictions about such a rational (Bayesian) noisy-channel language comprehender in a sentence comprehension task: (i) semantic cues should pull sentence interpretation towards plausible meanings, especially if the wording of the more plausible meaning is close to the observed utterance in terms of the number of edits; (ii) this process should asymmetrically treat insertions and deletions due to the Bayesian "size principle"; such non-literal interpretation of sentences should (iii) increase with the perceived noise rate of the communicative situation and (iv) decrease if semantically anomalous meanings are more likely to be communicated. These predictions are borne out, strongly suggesting that human language relies on rational statistical inference over a noisy channel.

communication | psycholinguistics | rational inference

Traditionally, models of sentence comprehension have assumed that the input to the sentence comprehension mechanism is an error-free sequence of words (e.g., refs. 1–6). However, the noise inherent in normal language use—due to producer or perceiver errors—makes this assumption an oversimplification. For example, language producers may misspeak or mistype because they do not have a full command of the language (due to being very young or a nonnative speaker, or suffering from a language disorder like aphasia), because they have not fully planned an utterance in advance, or because they are trying to communicate under the influence of stress or confusion. Similarly, language comprehenders may mishear or misread things because of their own handicaps (e.g., poor hearing/sight, or not paying sufficient attention), because the environment is noisy, or because the producer is not making sufficient effort in communicating clearly (e.g., whispering or mumbling, or writing in sloppy handwriting). Given the prevalence of these noise sources, it is plausible that language processing mechanisms are well adapted to handling noisy input, and so a complete model of language comprehension must allow for the existence of noise.

Noisy-channel models (7) of speech perception have been prominent in the literature for many years (e.g., 8–11). Furthermore, several researchers have observed the importance of noise in the input for the compositional, syntactic processes of sentence understanding (12–16), leading to the recent proposal of noisy-channel models of sentence understanding (17–19). According to a noisy-channel account, the sentence comprehension mechanism rationally combines information about a priori plausible utterances with a model of the imperfect transmission of the linguistic signal across a noisy channel. Using these information sources, comprehenders infer what meaning a producer most likely intended, given the observed linguistic evidence. Fig. 1 diagrams how communication takes place across a noisy channel, following ref. 7. Here, the producer has an intended meaning m_i and they choose an intended sentence s_i to communicate this meaning. The sentence is conveyed across a noisy channel and is corrupted by producer or comprehender noise, yielding a perceived sentence s_p. The comprehender thus observes s_p and must decode it to its intended meaning m_p. Successful communication occurs when m_i and m_p are the same—when the intended meaning is recoverable from the potentially corrupted input. As a simplification, we here examine situations in which m_i and m_p map directly and unambiguously onto their respective sentences, s_i and s_p. Consequently, we discuss primarily the recoverability of the intended sentence s_i given the perceived sentence s_p, taking the meanings as uniquely determined by the sentence strings.

This approach can be formalized by considering an ideal observer (20–22) model of language comprehension, in which the comprehender engages in optimal Bayesian decoding of the intended meaning:

$$P(s_i \mid s_p) \propto P(s_i) \ P(s_i \rightarrow s_p). \quad [1]$$

In Eq. 1, s_p is the sentence perceived by the comprehender and s_i is the sentence intended by the producer. The left-hand side, $P(s_i \mid s_p)$ gives the probability assigned by the comprehender to any particular hypothesized s_i, given the observed linguistic input s_p. By Bayes' rule, this can be rewritten on the right-hand side of Eq. 1 as the prior probability $P(s_i)$ that a producer would wish to communicate s_i, times the likelihood of s_p given s_i, which is often notated as $P(s_p|s_i)$. We write this likelihood as $P(s_i \rightarrow s_p)$ to make it clear that the likelihood represents the probability of s_i being corrupted to s_p in the process of communication. The prior $P(s_i)$ represents all of the comprehender's relevant linguistic and world knowledge, including for instance the base-rate frequencies of different grammatical constructions and the plausibility of different meanings. This term biases comprehenders toward a priori plausible utterances—things that are likely to be said. The noise likelihood term $P(s_i \rightarrow s_p)$ encodes the comprehender's knowledge of how sentences are likely to be corrupted during language transmission—for instance, the fact that smaller changes to a sentence are more likely than larger ones.

By trading off between the prior $P(s_i)$ and the likelihood $P(s_i \rightarrow s_p)$, comprehenders may arrive at interpretations that differ from the literal meanings of the specific sentences they perceive. That is, if comprehenders perceive an implausible sentence s_p that is "close" to a more plausible sentence under the noise model, they should infer that the producer actually uttered (and intended) the plausible sentence. In this case, the comprehender might

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Fig. 1. Communication across a noisy channel, following Shannon (7).

arrive at a higher overall posterior probability $P(s_i \mid s_p)$ by positing more noise–corresponding to a lower $P(s_i \to s_p)$ than if there were no noise–in combination with a higher plausibility $P(s_i)$. For example, suppose that the comprehender perceives the sentence, "The mother gave the candle the daughter." The prior likelihood for the literal meaning of this sentence is very low—corresponding to the idea that the mother would give her daughter to a candle—with the consequence that the overall posterior probability $P(s_i \mid s_p)$ may be higher for the slightly edited sentence, "The mother gave the candle to the daughter." The prior plausibility for the edited sentence is much higher, and as long as the likelihood $P(s_i \to s_p)$ of deleting a single function word is not too low, we may end up with a higher overall posterior likelihood for the edited sentence. Thus, the critical part of a noisy-channel account is that independent knowledge about likely meanings can lead listeners to interpretations that differ from the literal interpretation of the specific acoustic or visual stream they perceive.

Here, we evaluate this general framework by manipulating the terms in the Bayesian decoding setup, $P(s_i)$ and $P(s_i \to s_p)$ across the five syntactic alternations (23) shown in Table 1, in a sentence comprehension task using visually presented materials. We restrict our attention to alternations in which the content words are identical across the two variants, because (with the exception of confusable words) it seems unlikely that a comprehender would assume that a content word from the intended utterance would be omitted or that a content word from outside the intended utterance would be inserted. For example, in a context in which a boy is not mentioned, people will not interpret "the girl kicked" as possibly meaning that the girl kicked the boy, and they will assume that part of the meaning of the intended utterance includes "girl" and "kicked." Syntactic alternations allow for the same thematic content to be expressed in different ways: by ordering the components of the message in a certain way, we can emphasize one or another part of the message. For example, to convey the idea of the girl kicking the ball, we can choose between the active frame ("The girl kicked the ball") and the passive frame ("The ball was kicked by the girl"), and this choice depends on whether we want to focus the comprehender's attention on the girl and what she did vs. on the ball and what was done to it. Critically, although different syntactic alternations (Table 1) share the fact that the two alternatives are identical in terms of propositional meaning, they can vary in how close the alternatives are, under simple string edits (see also refs. 17–19).

For each of the five alternations that we investigated, we considered semantically plausible and implausible sentences. To construct the implausible versions, we swapped the order of the noun phrases that are involved in each alternation (e.g., "The mother gave the daughter the candle" → "The mother gave the candle the daughter"; "The girl kicked the ball" → "the ball kicked the girl"). When the sentence is plausible, the prior probability is high, and thus comprehenders should interpret the sentence literally. However, the prior probability of implausible sentences is low. Therefore, if comprehenders rationally follow Eq. **1**, their interpretation of implausible sentences should depend on how close the perceived string is to a plausible alternative. For instance, the implausible sentence, "The mother gave the candle the daughter," could have resulted from the plausible sentence, "The mother gave the candle to the daughter," via accidental deletion of the word "to." If the likelihood of deletion is high, comprehenders may infer this deletion and interpret the sentence as the plausible sentence, "The mother gave the candle to the daughter," not the perceived one. Similarly, the implausible sentence, "The ball kicked the girl," could have resulted from the plausible sentence, "The ball was kicked by the girl," via accidental deletion of the words "was" and "by."

We evaluate four specific predictions of this rational noisy-channel comprehension account.

Prediction 1. As a first approximation, we assume that there are two types of string edits: insertions and deletions. We further assume that string edits are independent and that both types of string edits occur with equal probability. This has the consequence that comprehenders should be more willing to forego the literal interpretation when the semantically plausible interpretation involves positing fewer changes to the signal under the noise model, compared with more changes. Under Eq. **1**, comprehenders should prefer sentences s_i such that the likelihood of generating s_p, $P(s_i \to s_p)$, is high. If string edits are independent, then $P(s_i \to s_p)$ increases as the differences between s_i and s_p decrease, so that s_i is more likely to be hypothesized to be the

Table 1. The necessary edits to get from an English construction to its alternation

English constructions	Plausible version	Change	Implausible version
1. Active/passive	a. The girl kicked the ball. (active)	Two insertions	c. The girl <u>was</u> kicked <u>by</u> the ball. (passive)
	b. The ball <u>was</u> kicked <u>by</u> the girl. (passive)	Two deletions	d. The ball kicked the girl. (active)
2. Subject-locative/ object-locative	a. <u>Onto</u> the table jumped a cat. (subject-locative)	One deletion, one insertion	c. The table jumped <u>onto</u> a cat. (object-locative)
	b. The cat jumped <u>onto</u> a table. (object-locative)	One insertion, one deletion	d. <u>Onto</u> the cat jumped a table. (subject-locative)
3. Transitive/intransitive	a. The tax law benefited the businessman. (transitive)	One insertion	c. The tax law benefited <u>from</u> the businessman. (intransitive)
	b. The businessman benefited <u>from</u> the tax law. (intransitive)	One deletion	d. The businessman benefited the tax law. (transitive)
4. DO/PO goal	a. The mother gave the daughter the candle. (DO-goal)	One insertion	c. The mother gave the daughter <u>to</u> the candle. (PO-goal)
	b. The mother gave the candle <u>to</u> the daughter. (PO-goal)	One deletion	d. The mother gave the candle the daughter. (DO-goal)
5. DO/PO benefactive	a. The cook baked Lucy a cake. (DO-benef)	One insertion	c. The cook baked Lucy <u>for</u> a cake. (PO-benef)
	b. The cook baked a cake <u>for</u> Lucy. (PO-benef)	One deletion	d. The cook baked a cake Lucy. (DO-benef)

The five alternations that are investigated in this paper are as follows: 1, active/passive; 2, subject-locative/object-locative; 3, transitive/intransitive; 4, double-object/prepositional phrase object goals; and 5, double-object/prepositional phrase object benefactives. The number of insertions and deletions that are needed to form an implausible alternation from the plausible version is provided for each plausible/implausible pair, as a proposed hypothesis for how the implausible versions might be generated. benef, benefactive; DO, double object; PO, prepositional phrase object.
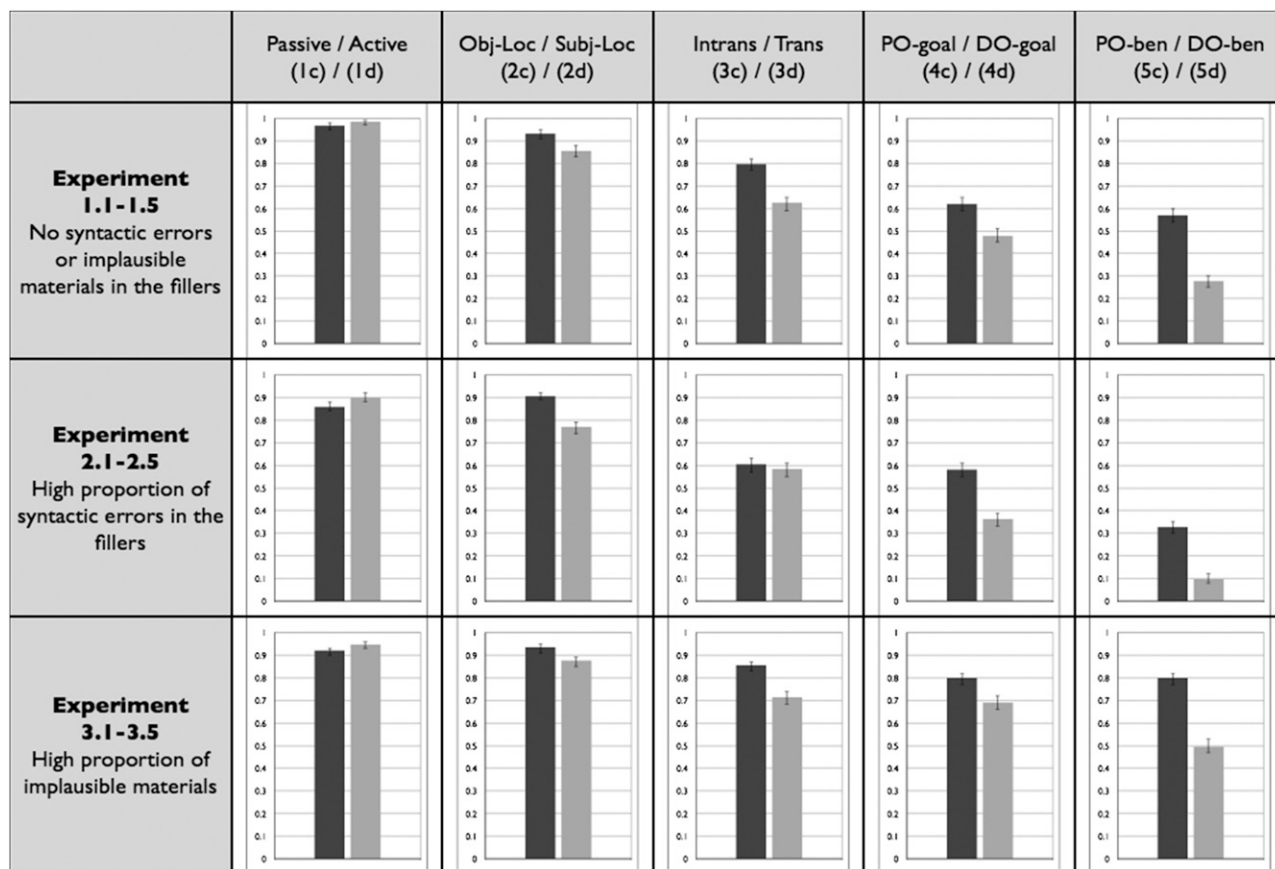
Fig. 2. Percentage of trials in which participants relied on the literal syntax for the interpretation of the implausible syntactic constructions. Error bars are 95% confidence intervals. Examples of each implausible construction and its closest edit to a plausible construction are given in Table 1, e.g., an example of the implausible passive construction is "The girl was kicked by the ball," which can be generated from the plausible active construction, "The girl kicked the ball," by adding in two function words: "was" and "by" in specific locations. Key results are as follows: (*i*) people relied on the literally presented sentence more in the major-change alternations than in the minor-change alternations in each experiments: the percentages in the two left-most constructions are all higher than the percentages in the three right-most constructions; (*ii*) increasing the noise rate lowered interpretation as the literal sentence: the percentages in experiment 2 are lower than those in experiment 1, especially for minor-change alternations; (*iii*) increasing the base rate of implausible events increased interpretation as the literal sentence: the percentages in experiment 3 are higher than those in experiment 1, especially for minor-change alternations; and (*iv*) semantic cues should have a stronger influence on structures whose alternations require a single deletion than those whose alternations require a single insertion: this is visible as a decrease in proportion of reliance on syntax for each of the minor-change alternations comparing the insertion condition on the *Left* to the deletion condition on the *Right*.

meaning if s_p can be created from s_i with fewer string edits. For instance, the deletion of a single word should be more likely under the noise model than the deletion of two words. Thus, alternations 3–5 in Table 1—in which the two alternatives differ from each other by a single insertion or deletion—should be more affected by plausibility than the alternations 1 and 2 in Table 1, in which the two alternatives differ from each other by two edit operations. It should be noted that this prediction follows from the particular string-edit-distance theory assumed here (insertions and deletions). For example, active/passive alternatives may be two insertions or deletions away, but if simple transpositions of content words are allowed, the implausible/plausible variants are just one edit operation apart.

Prediction 2. The noise model $P(s_i \rightarrow s_p)$ should not treat all changes equally. In particular, comprehenders should infer nonliteral meanings more readily when the change involves a deletion, compared with an insertion. This prediction holds generally for a wide range of statistically sensible noise models, and follows from the Bayesian size principle (24, 25): a deletion only requires a particular word to be randomly selected from a sentence, whereas an insertion requires its selection from (a subset of) the producer's vocabulary; the insertion of a specific

word therefore has smaller likelihood $P(s_i \rightarrow s_p)$ than the deletion of a specific word, even under the assumption that insertions and deletions occur equally often. Note that this differs from the symmetric Levenshtein noise model (26) used in refs. 17–19. Thus, semantic cues should have a stronger influence for each of the implausible structures in 3d, 4d, and 5d—in which a word has been deleted from the plausible alternation—than for the implausible structures in 3c, 4c, and 5c—in which a word has been inserted into the plausible alternation.

Prediction 3. Because comprehenders do not know the noise rate–the probability that the noise model will corrupt s_i to a different s_p–in every communicative scenario, they must infer it. Increasing the perceived noise rate should encourage comprehenders to infer a nonliteral but plausible alternative. For example, consider a situation in which you are having trouble hearing the speaker. In such a situation, if you hear an implausible utterance, you may be more likely to attribute it to noise, and infer that the speaker intended something more plausible, than if you encountered the same input in a less noisy environment.

Prediction 4. Increasing the base rate of implausible sentences should discourage comprehenders from inferring anything other

than the literal meaning of the perceived sentence. For example, imagine you are talking to someone who produced many implausible sentences (e.g., a Wernicke's aphasic patient, or an individual suffering from psychosis). In such a situation, you would be more likely to assume that a particular implausible sentence was intended, rather than produced because of an error. In this case, $P(s_i)$ would be more evenly distributed between implausible and plausible sentences, making comprehenders less willing to deviate from the literal meaning of the observed s_p.

Results

We evaluated these predictions in three experiments run over Amazon.com's Mechanical Turk platform in which participants' interpretations were probed with comprehension questions, which were presented simultaneously with the target sentences (*Methods*). The critical dependent measure was the rate at which participants interpreted implausible sentences literally as presented vs. as a close plausible alternative. The same test sentences were used across experiments, with different fillers to evaluate predictions 3–4. In particular, in experiment 2, we increased the perceived noise by having one-half of the filler items contain syntactic errors (such as a deleted or inserted function word), compared with experiment 1, in which the fillers contained no syntactic errors. Additionally, in experiment 3 we increased the base rate of implausible sentences by increasing the rate of implausible filler materials (implausible-to-plausible ratio, 5:16), compared with experiment 1 (1:8). In all experiments, we compare comprehenders' probability of interpreting the sentence as literally presented, using a mixed-effects logistic regression model, with slopes and intercepts by participant and item.

Results are shown in Fig. 2. We do not include the plausible conditions in Fig. 2 because, as expected, these sentences were overwhelmingly interpreted literally: over 95% in each plausible condition in each experiment.

The predictions of the noisy-channel account were borne out across the five alternations in the three experiments. First, the rate of literal interpretation was higher when the plausible alternative involved positing more string edits (major change) compared with fewer (minor change): experiment 1: 93.4% for major change alternations (1 and 2 in Table 1); 56.1% for minor change alternations (3–5 in Table 1; $\beta = 3.37$; $P < 0.0001$); experiment 2: 85.9% for major change alternations; 42.7% for minor change alternations ($\beta = 3.53$; $P < 0.0001$); experiment 3: 92.0% for major change alternations; 72.5% for minor change alternations ($\beta = 2.21$; $P < 0.0001$).

Second, the rate of literal interpretation was higher when the plausible alternative involved an insertion rather than a deletion: experiment 1: 66.1% vs. 46.0% ($\beta = 1.39$; $P < 0.0001$); experiment 2: 50.4% vs. 34.9% ($\beta = 1.45$; $P < 0.0001$); experiment 3: 81.1% vs. 63.3% ($\beta = 1.26$; $P < 0.0001$).

Third, the rate of literal interpretation decreased as the perceived noise rate increased. In particular, the rate of literal interpretation was lower in experiment 2, in which one-half of the filler items contained syntactic errors. This effect was robust across the minor-change alternations (experiment 1: 56.1%; experiment 2: 42.7%; $\beta = -0.92$; $P < 0.0001$), and it showed a trend, sometimes significant, for each individual alternation (Table 1): (3a) intransitive, 71.1%, vs. (3b) transitive, 59.5% ($\beta = -0.67$; $P < 0.05$); (4a) prepositional phrase object (PO)-goal, 54.9%, vs. (4b) double object (DO)-goal, 47.3% ($\beta = -0.63$; $P < 0.2$); (5a) PO-ben, 42.2%, vs. (5b) DO-ben, 21.2% ($\beta = -1.36$; $P < 0.0001$).

Fourth, the rate of literal interpretation increased as the base rate of implausible sentences increased. In particular, the rate of literal interpretation was higher in experiment 3, in which the rate of implausible filler materials was higher than in experiment 1 or 2. This effect was robust across the minor-change alternations (experiment 1: 56.1%; experiment 3: 72.5%; $\beta = 0.58$; $P < 0.01$), and it was significant for each individual alternation (Table 1): (3a) intransitive, 71.1%, vs. (3b) transitive, 78.4% ($\beta = 0.63$; $P < 0.05$); (4a) PO-goal, 54.9%, vs. (4b) DO-

goal, 74.5% ($\beta = 1.09$; $P < 0.005$); (5a) PO-benefactive (benef), 42.2%, vs. (5b) DO-benefactive, 64.5% ($\beta = 1.05$; $P < 0.0005$).

Finally, to ensure that the results are not affected by the online administration of the experiments (via Amazon.com's Mechanial Turk interface), we reran experiment 3 in the laboratory. The results were very similar to those obtained via Mechanical Turk, as shown in Fig. S1.

Discussion

The current results provide strong evidence in support of noisy-channel models of sentence comprehension, in the spirit of refs. 17–19. Four predictions of noisy-channel models were confirmed, each of which showed that factors which should influence a rational Bayesian decoder did influence people's interpretation of sentences. The rate of literal interpretation was affected by how close the literal string was to the plausible alternative, with more differences leading to higher rates of literal interpretation (prediction 1) and insertions leading to higher rates of literal interpretation than deletions (prediction 2). Furthermore, as the perceived noise increased, participants were more willing to posit that sentences were corrupted by noise (prediction 3), and thus the rate of literal interpretation decreased. Finally, as the base rate of implausible sentences increased, participants were less willing to posit changes to sentences to remove the implausibility (prediction 4). Thus, the rate of literal interpretation increased.

The noisy-channel model contrasts with previous sentence comprehension models, which argue that, although meaning may guide initial interpretation in the face of temporary syntactic ambiguity (e.g., refs. 27 and 28), the final interpretation is determined by a sentence's syntax. For example, it has been observed (29) that English speakers primarily use syntactic information when syntactic and plausibility information conflict, in the comprehension of implausible active and passive structures, as in 1c and 1d. From this and related results, it has been argued that English primarily relied on syntactic cues for determining the final meaning of a sentence. A limitation of this proposal, however, is that it does not capture the effects of the edit distance to a sentence string with a plausible meaning on people's reliance on syntax. As we have seen here, when the edit distance is relatively large (for example, between an active structure and a passive structure), people tend to interpret the string literally, even if the content is implausible (29). However, when the edit distance between two alternatives is smaller—as in the double-object/prepositional phrase object alternations or the transitive/intransitive alternations—people have a greater tendency to interpret the string according to the more plausible alternative. This pattern of results is as predicted by the noisy-channel model of sentence interpretation, such that comprehenders appear to combine syntactic cues to meaning with expectations about likely sentences to be uttered and likely mistakes in communication to arrive at the most likely interpretation. This is exactly what we should expect from a system that is designed for communication over a noisy channel.

The cue integration approach proposed in the competition model (29) is similar in spirit to a noisy-channel model in its assumption of noise in the input and its reliance on the integration of information from a variety of sources. However, the specific cue integration approach proposed in the competition model does not appear to be quite consistent with the observed results. In particular, according to the competition model, English speakers rely primarily on syntactic cues for their interpretation of an utterance. However, the results of our experiments show that people's reliance on syntactic cues within a language depends on the particular construction: how far away a plausible alternative is in terms of its edit distance. Thus, although English speakers tend to follow the literal syntactic information for the active–passive alternation (consistent with the cue integration model), they are much less likely to follow the literal syntactic information for the double-object/prepositional phrase object alternation. These across-construction differences are not explained by the cue integration model. Thus, although the cue integration model is important because of its ability to combine information sources

to select interpretations, the particular cue integration model that has been proposed in the literature is potentially limited because it fails to provide a role for reconstruction from noise. A better understanding of when and how this reconstruction is happening may lead to a more complete noisy-channel–based cue integration model.

In the context of the noisy-channel proposal, our experiments address people's expectations about plausible semantic content. Another important component of people's expectations is structural frequency (4–6). Structural frequencies will likely play an important role when the frequencies of the two target structures vary substantially: people will tend to interpret the structure according to its more frequent neighbor. Whereas we have not modeled construction frequency effects, we see a robust effect of structural frequency in the comparison between locative objects (which are common, contingent on the verb being present) vs. locative subjects (which are extremely rare). In particular, people are less likely to interpret the low-frequency locative-subject construction literally compared with the high-frequency locative-object construction (experiment 1: 93.3% vs. 85.6%; β = −2.31; $P < 0.001$; experiment 2: 90.7% vs. 76.9%; β = −5.86; $P < 0.001$; experiment 3: 93.4% vs. 87.6%; β = −5.43; $P < 0.001$). We also see a smaller effect for the comparison between active—the more frequent structure—and passive—the less frequent structure (experiment 1: 98.6% vs. 96.8%; β = −0.26; $P > 0.5$; experiment 2: 90.0% vs. 85.9%; β = −1.65; $P < 0.005$; experiment 3: 94.8% vs. 92.0%; β = −2.23; $P < 0.01$). This replicates the results from ref. 30, which observed that people were more likely to rely on plausibility information for passive structures compared with active structures.

If the noisy-channel hypothesis is on the right track, then we should be able to see evidence in on-line language comprehension for the process of correcting flawed input to a more likely alternative. Indeed, such evidence exists in the event-related potential (ERP) literature. Traditionally, the N400 and the P600 ERP components have been interpreted as indexing semantic/plausibility anomalies (31) and syntactic anomalies (32, 33), respectively. More recently, however, P600 effects have been observed for some materials with semantic incongruities (e.g., refs. 34–36). For example, a P600 effect is observed at the verb "devouring" in the string, "The hearty meal was devouring...," relative to the verb "devoured" in the string, "The hearty meal was devoured...." (35). Furthermore, P600 effects have been reported for orthographic errors, e.g., "fone" for "phone" (37).

The current results suggest an explanation of the P600 effect within the noisy-channel framework (see refs. 38 and 39 for related proposals of the P600 as an error-correction signal). In particular, the P600 component may reflect the correction process that comprehenders engage in when encountering flawed linguistic input. According to this explanation, a P600 is observed for ungrammaticalities (e.g., "Many doctors claims....."; 40) or for semantically unexpected continuations in some cases (e.g., "The hearty meal was devouring") because in both cases it is clear what was plausibly intended. Similarly, in the case of orthographic errors, it is clear what the intended word was. Critically, however, when there is no clear alternative to which the flawed input could be corrected, no P600 component is observed, as in the case of the "classical" semantic anomalies (31), or in Jabberwocky materials in which plausibility cues are absent (e.g., ref. 41). In some cases in which a correction is unlikely, a P600 has nonetheless been reported (e.g., refs. 36 and 42). However, these experiments often use a plausibility/acceptability-judgment task, which is critically not a communicative task. We predict a P600 only in cases in which the task is communicative (e.g., passive reading or reading with comprehension questions) because these are the cases in which the comprehender is likely to model the noise (i.e., evaluate the relative likelihoods of producer's errors, and correct the plausible ones). Conversely, when the task is a plausibility-judgment task, we should see P600s for a wider range of materials, because there should be more correction.

Building on refs. 17–19, our work moves language processing theories toward more realistic types of communicative scenarios, in which sentences are imperfectly observed or incorrectly uttered. The ability to understand noisy input is not surprising from the viewpoint of rational analysis (43, 44), which studies ways in which cognitive systems are "well designed" for the task that they perform. In the case of language processing, comprehension mechanisms are sensitive to the types of information sources that an idealized statistical comprehender would be. However, this capacity is surprising from the traditional view of linguistics and psycholinguistics, which have focused on modeling noise-free input.

In summary, we have demonstrated that comprehenders rationally integrate the likelihood of noise with prior expectations, providing strong evidence for the idea that language understanding is rational statistical inference over a noisy channel. The present work reveals fundamental aspect of human language processing: it is not built only for pristine input. Instead, language processing mechanisms engage in sophisticated on-line integration of prior expectations about likely utterances, with models of how linguistic signals might be corrupted during transmission. As such, evolutionarily or developmentally, language processing mechanisms are shaped to handle many of the complexities of real-world communication.

Methods

Experimental participants were presented with a questionnaire consisting of 60 sentences, like examples 1–5 in Table 1, each followed by a comprehension question, as in 6:

6. a. Active/passive example: The diamond lost the woman.

   Did the diamond lose something/someone? (literal syntax: yes).

   b. Active/passive example: The ball kicked the girl.

   Did the girl kick something/someone? (literal syntax: no).

   c. DO/PO-goal example: The girl tossed the apple the boy.

   Did the apple receive something/someone? (literal syntax: yes).

   d. DO/PO-goal example: The mother gave the candle the daughter.

   Did the daughter receive something/someone? (literal syntax: no).

The target sentences and the questions were presented simultaneously, and participants could read the sentences and questions as many times as they liked before making their choices. Hence there was no memory component to answering the comprehension questions. (Consequently, the methodology does not distinguish on-line and postinterpretive processes as the source of the effects.)

The answer to the question following each target sentence indicates whether the participant used syntactic or semantic cues in interpreting the sentence. For example, in 6a and 6c, a "yes" answer indicates that the reader used syntax to interpret the sentence, whereas a "no" indicates that the reader relied on semantics, whereas the reverse holds for 6b and 6d.

Twenty sets of materials were constructed for each alternation 1–5 in a 2 × 2 design, crossing construction (alternative 1, alternative 2) with the plausibility of the target alternation relative to the other (plausible, implausible). The items were counterbalanced so that one-half had questions like 6a, in which a "yes" answer indicated the use of literal syntax in interpretation, and the other half had questions like 6b, in which a "no" answer indicated the use of literal syntax in interpretation.

Each set of 20 items was divided into four lists according to a Latin square design, and each list was then combined with 60 filler sentences (e.g., "The commissioner wrote a report for the chairman") to form a presentation list. The target materials were the same across the three experiments. In experiments 1-1 through 1-5, the filler items were all plausible and grammatical sentences. In experiments 2-1 through 2-5, the filler items consisted of the filler items from experiments 1-1 through 1-5, but with 30 of these edited to contain syntactic errors: in 10 items, a function word was deleted (e.g., "The commissioner wrote a report for the chairman." → "The commissioner wrote a report the chairman."); in 10 items a function word was inserted (e.g., "The colonel was knighted by the queen because of his loyalty." → "The colonel was knighted for by the queen because of his loyalty."); and in 10 items, a few adjacent words were scrambled (e.g., "A bystander was rescued by the

fireman in the nick of time." → "A bystander was the fireman by rescued in the nick of time."). In experiment 3, unlike in experiments 1–2 (in which each construction was run in a separate experiment), all five sets of target materials were presented together, along with the 60 plausible fillers. As a result, the ratio of implausible materials was much higher in this experiment than in the other experiments: 5 constructions * 10 implausible target syntactic materials together with 5 constructions * 10 plausible target syntactic materials and 60 plausible fillers, resulting in a ratio of 5:16 (compare a ratio of 1:8 in experiments 1-1 through 1-5, and 2-1 through 2-5).

A random order of each experimental list was presented to the participants on Amazon.com's Mechanical Turk, a marketplace interface that can be used for collecting behavioral data over the internet. To constrain the population to American English speakers, we restricted the IP addresses to those in the United States. Furthermore, we asked participants what their native language was, and where they were from originally. Payment was not contingent on answers to these questions. There were 60 initial participants in each experiment, a different set of participants for each experiment (300 participants for experiment 1; 300 participants for experiment 2; 60 participants for experiment 3).

We analyzed only participants who self-identified as native speakers of English from the United States. Furthermore, we only analyzed data from participants who answered at least 75% of the plausible materials correctly.

(The mean across participants and experiments was over 98%.) These restrictions caused the elimination of zero to three participants' data per experiment.

In addition, to test the validity of Mechanical Turk, we reran experiment 3 in a laboratory setting, using participants in the Massachusetts Institute of Technology area. [There now exist several replications of results from in-laboratory studies using Amazon's Mechanical Turk, thus establishing the viability of this method for obtaining experimental linguistic data (45–47).] The presentation of the items was identical to the previous experiments. Sixty native English speakers were recruited for this experiment. The data from 54 of the participants met the inclusion criteria above, and our analyses were restricted to these participants.

1. Frazier L, Fodor JD (1978) The sausage machine: A new two-stage parsing model. *Cognition* 6(4):291–325.
2. Gibson E (1991) A computational theory of human linguistic processing: Memory limitations and processing breakdown. PhD thesis (Carnegie Mellon University, Pittsburgh).
3. Gibson E (1998) Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1):1–76.
4. Jurafsky D (1996) A probabilistic model of lexical and syntactic access and disambiguation. *Cogn Sci* 20:137–194.
5. Hale J (2001) A probabilistic Earley parser as a psycholinguistic model. *Proc NAACL* 2:159–166.
6. Levy R (2008) Expectation-based syntactic comprehension. *Cognition* 106(3):1126–1177.
7. Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:623–656.
8. Jelinek F (1976) Continuous speech recognition by statistical methods. *Proc IEEE* 64:532–556.
9. Aylett M, Turk A (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech* 47(Pt 1):31–56.
10. Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA (2008) Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3):804–809.
11. Dilley LC, Pitt MA (2010) Altering context speech rate can cause words to appear or disappear. *Psychol Sci* 21(11):1664–1670.
12. Bates E, McDonald JL, MacWhinney BM, Appelbaum M (1991) A maximum likelihood procedure for the analysis of group and individual data in aphasia research. *Brain Lang* 40(2):231–265.
13. MacWhinney B, Osmán-Sági J, Slobin DI (1991) Sentence comprehension in aphasia in two clear case-marking languages. *Brain Lang* 41(2):234–249.
14. Blackwell A, Bates E (1995) Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *J Cogn Neurosci* 7(2):228–257.
15. Ferreira F, Bailey KGD, Ferraro V (2002) Good-enough representations in language comprehension. *Curr Dir Psychol Sci* 11(1):11–15.
16. Ferreira F, Christianson K, Hollingworth A (2001) Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *J Psycholinguist Res* 30(1):3–20.
17. Levy R (2008) A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Stroudsburg, PA), pp 234–243.
18. Levy R, Bicknell K, Slattery T, Rayner K (2009) Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proc Natl Acad Sci USA* 106(50):21086–21090.
19. Levy R (2011) Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Portland, OR), pp 1055-1065.
20. Marr D (1982) *Vision* (Freeman, New York).
21. Geisler WS (1989) Sequential ideal-observer analysis of visual discriminations. *Psychol Rev* 96(2):267–314.
22. Geisler WS, Diehl RL (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn Sci* 27:379–402.
23. Levin B (1993) *English Verb Classes and Alternations: A Preliminary Investigation* (Univ of Chicago Press, Chicago).
24. Xu F, Tenenbaum JB (2007) Word learning as Bayesian inference. *Psychol Rev* 114(2):245–272.
25. MacKay DJC (2003) *Information Theory, Inference and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).
26. Levenshtein V (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Sov Phys Dokl* 10:707–710.
27. Trueswell JC, Tanenhaus MK, Garnsey SM (1994) Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *J Mem Lang* 33(3):285–318.
28. Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217):1632–1634.
29. MacWhinney B, Bates E, Kliegl R (1984) Cue validity and sentence interpretation in English, German, and Italian. *J Verbal Learn Verbal Behav* 23(2):127–150.
30. Ferreira F (2003) The misinterpretation of noncanonical sentences. *Cognit Psychol* 47(2):164–203.
31. Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207(4427):203–205.
32. Osterhout L, Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. *J Mem Lang* 31(6):785–806.
33. Hagoort P, Brown C, Groothusen J (1993) The syntactic positive shift as an ERP measure of sentence processing. *Lang Cogn Process* 8(4):439–483.
34. Kuperberg GR, Caplan D, Sitnikova T, Eddy M, Holcomb P (2006) Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Lang Cogn Process* 21(5):489–530.
35. Kim A, Osterhout L (2005) The independence of combinatory semantic processing: Evidence from event-related potentials. *J Mem Lang* 52(2):205–225.
36. Kuperberg GR (2007) Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res* 1146:23–49.
37. Münte TF, Heinze HJ, Matzke M, Wieringa BM, Johannes S (1998) Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia* 36(3):217–226.
38. Coulson S, King JW, Kutas M (1998) Expect the unexpected: Event-related brain response to morphosyntactic violations. *Lang Cogn Process* 13(1):21–58.
39. van Herten M, Chwilla DJ, Kolk HHJ (2006) When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *J Cogn Neurosci* 18(7):1181–1197.
40. Osterhout L, Mobley LA (1995) Event-related brain potentials elicited by failure to agree. *J Mem Lang* 34(6):739–773.
41. Yamada Y, Neville HJ (2007) An ERP study of syntactic processing in English and nonsense sentences. *Brain Res* 1130(1):167–180.
42. Kolk HHJ, Chwilla DJ, van Herten M, Oor PJW (2003) Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang* 85(1):1–36.
43. Anderson JR, Schooler LJ (1991) Reflections of the environment in memory. *Psychol Sci* 2(6):396–408.
44. Chater N, Oaksford M (1999) Ten years of the rational analysis of cognition. *Trends Cogn Sci* 3(2):57–65.
45. Frank MC, Tily H, Arnon I, Goldwater S (2010) Beyond transitional probabilities: Human learners impose a parsimony bias in statistical word segmentation. *Proceedings of the Cognitive Science Society* (Portland, OR), pp 760–765.
46. Gibson E, Fedorenko E (2013) The need for quantitative methods in syntax and semantics research. *Lang Cogn Process* 28(1–2):88–124.
47. Munro R, et al. (2010) Crowdsourcing and language studies: The new generation of linguistic data. *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk* (Association for Computational Linguistics, Stroudsburg, PA), pp 122–130.

Supporting Information

Gibson et al. 10.1073/pnas.1216438110
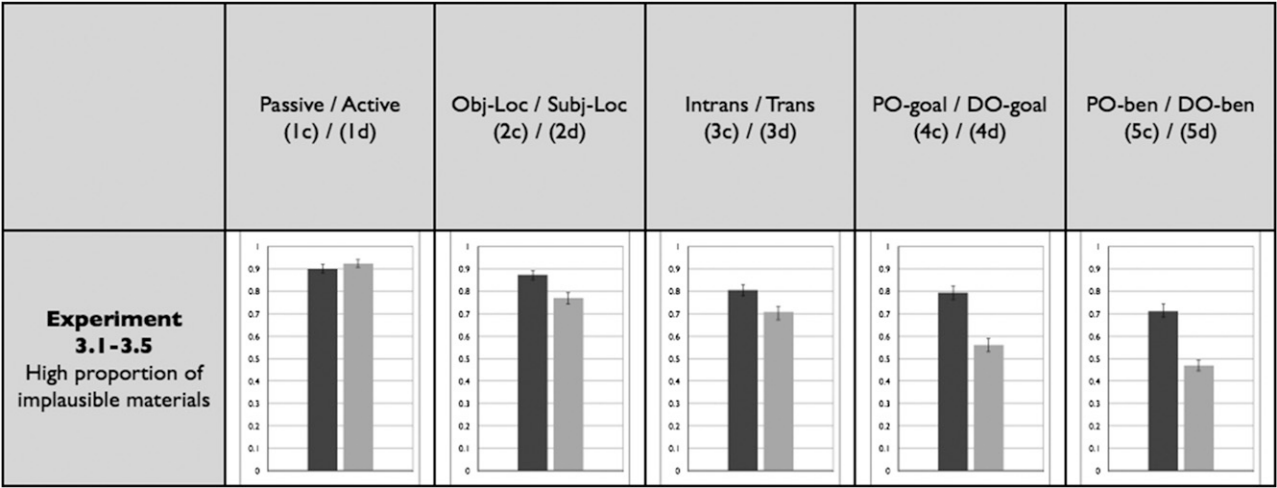
Fig. S1. Replication of experiment 3 using local in-laboratory participants. A comparison with the data gathered from Amazon's Mechanical Turk (Fig. 2, *Bottom*) reveals the same patterns of results. In order to ensure that the results were not affected by the online administration of the experiments (via Amazon.com's Mechanical Turk interface), we reran experiment 3 in the laboratory. The results (shown in Fig. S1) were very similar to those obtained via Mechanical Turk. In particular, the first two predictions of the model were again borne out (the third and fourth predictions could not be evaluated using experiment 3 alone). First, the rate of literal interpretation was higher for the major change alternations than for the minor change alternations (86.6% vs. 67.4%; β = 1.67; *P* < 0.001). Second, the rate of literal interpretation was higher for the insertions than for the deletions (76.9% vs. 57.9%; β = 1.27; *P* < 0.001).