Published in final edited form as:

Cortex. 2017 July ; 92: 19–31. doi:10.1016/j.cortex.2017.02.015.

A rational inference approach to group and individual-level sentence comprehension performance in aphasia

Tessa Warren¹, Michael Walsh Dickey^{2,1}, and Teljer L. Liburd¹

¹University of Pittsburgh

²VA Pittsburgh Healthcare System

Abstract

The rational inference, or noisy channel, account of language comprehension predicts that comprehenders are sensitive to the probabilities of different interpretations for a given sentence and adapt as these probabilities change (Gibson, Bergen & Piantadosi, 2013). This account provides an important new perspective on aphasic sentence comprehension: aphasia may increase the likelihood of sentence distortion, leading people with aphasia (PWA) to rely more on the prior probability of an interpretation and less on the form or structure of the sentence (Gibson, Sandberg, Fedorenko, Bergen & Kiran, 2015). We report the results of a sentence-picture matching experiment that tested the predictions of the rational inference account and other current models of aphasic sentence comprehension across a variety of sentence structures. Consistent with the rational inference account, PWA showed similar sensitivity to the probability of particular kinds of form distortions as age-matched controls, yet overall their interpretations relied more on prior probability and less on sentence form. As predicted by rational inference, but not by other models of sentence comprehension in aphasia, PWA's interpretations were more faithful to the form for active and passive sentences than for direct object and prepositional object sentences. However contra rational inference, there was no evidence that individual PWA's severity of syntactic or semantic impairment predicted their sensitivity to form versus the prior probability of a sentence, as cued by semantics. These findings confirm and extend previous findings that suggest the rational inference account holds promise for explaining aphasic and neurotypical comprehension, but they also raise new challenges for the account.

Keywords

language comprehension; plausibility; sentence processing; Bayesian comprehension; noisy channel

Correspondence to: Tessa Warren, 635 LRDC, 3939 O'Hara St., Pittsburgh, PA 15260, tessa@pitt.edu, Phone: (412) 624-7460.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

People with aphasia (PWA) commonly exhibit impaired performance on sentences with complex or low-frequency structures and few semantic cues to their intended meaning, including reversible passive sentences (e.g., Cho-Reyes & Thompson, 2012), object-extracted relative clauses (e.g., Berndt, Mitchum & Wayland, 1997) and object clefts (Cho-Reyes & Thompson, 2012). However, PWA perform better when world knowledge is an informative cue to sentence meaning (Caramazza & Zurif, 1976; Saffran, Schwartz & Linebarger, 1998). For example, non-reversible passives like *The ball was chased by the dog* typically elicit better performance than reversible passives like *The boy was chased by the dog* (e.g., Schwartz, Saffran & Marin, 1980), presumably because the knowledge that balls cannot chase dogs overrides interference from a potential active interpretation of this passive sentence. Furthermore, the effect of semantic cues appears to be stronger for cases in which the reversed meaning would be impossible (a ball chasing a boy) rather than simply unlikely (a worm swallowing a bird; Caramazza & Zurif, 1976; Schwartz, Linebarger, Saffran & Pate, 1987).

There are a variety of models of aphasic sentence comprehension deficits, all of which are designed to explain why PWA tend to perform better or worse on particular structures. All such models predict that low-frequency non-canonical sentence structures such as passives or object-extracted relative clauses will be more difficult for PWA to comprehend. They also all assume that PWA rely heavily on semantic cues when comprehension is impaired, but they fail to provide an explanatory mechanism that accounts for why. In contrast, the model of sentence comprehension that Gibson et al. (2013) refers to as the noisy channel model, and following Gibson et al. (2015) we will refer to as the rational inference account, integrates syntactic (form-based) and semantic factors in a single framework and thus has the potential to explain patterns of trade-offs between form and meaning across both PWA and neurotypical adults. For this reason, the rational inference account holds considerable interest as a framework for understanding aphasic comprehension. An additional benefit of this model is that it provides a new focus for characterizing the nature of aphasic language deficits, moving towards factors like expectations (e.g. about noise or intention), which are experience-dependent. Experience-based and adaptive effects are just beginning to be explored in aphasia (DeDe, 2013a,b; Knilans & DeDe, 2015; Gahl, 2002; Schuchard & Thompson, 2014; Warren, Dickey & Lei, 2016), but understanding their role may be relevant to both improving models of aphasic language processing and advancing aphasia treatment. Only one study to date, by Gibson, Sandberg, Fedorenko, Bergen and Kiran (2015), has evaluated the rational inference account as applied to aphasic comprehension, and it provides promising support for the theory. The current experiment builds on and expands this work, providing a stronger and more detailed test of the predictions of the rational inference account and evaluating its predictions as compared to those of other models of aphasic comprehension.

A rational inference, or noisy channel, model of sentence comprehension

Gibson et al. (2013) proposed that language comprehension can be modeled as a process of using Bayesian reasoning to compute the probabilities of possible intended sentences given a

perceived sentence. As discussed by Gibson et al. (2015), this rational inference view of language comprehension provides a useful alternative view of sentence comprehension among PWA. Gibson and colleagues (2013) provided the following formula to capture the probability that a given sentence was intended: $P(S_{\text{intended}}|S_{\text{perceived}}) \propto P(S_{\text{intended}}) P(S_{\text{intended}} \rightarrow S_{\text{perceived}})$. This formula states that the probability that a particular sentence was intended given that a particular sentence was perceived ($P(S_{\text{intended}}|S_{\text{perceived}})$) is proportional to the probability of that intended sentence ($P(S_{\text{intended}})$), referred to as the *prior* multiplied by the probability that that intended sentence might have been distorted into the perceived sentence ($P(S_{\text{intended}} \rightarrow S_{\text{perceived}})$), referred to as the *noise term*. Note that under the rational inference model, interpretations that are not faithful to a sentence's input structure, i.e. incorrect interpretations, are understood as being cases in which the listener/reader has assumed that the sentence's structure has been distorted. Gibson et al. used a series of questionnaire experiments to show that neurotypical comprehenders were sensitive to manipulations of the probabilities in this formula during sentence interpretation.

Gibson et al. (2013) manipulated semantic coherence (whether the sentence described a sensible event or not) to influence the prior, $P(S_{\text{intended}})$, i.e. the probability that a sentence was intended. This follows from the assumption that speakers usually intend to produce semantically coherent sentences, so sentences that describe plausible events are more likely to be intended than sentences that describe implausible or impossible events. To manipulate the noise term, $P(S_{\text{intended}} \rightarrow S_{\text{perceived}})$, Gibson et al. tested a set of five syntactic alternations that required different numbers and kinds of edits to change one form of the alternation to the other. On the basis of the Bayesian size principle (e.g. Tenenbaum & Griffiths, 2001), Gibson et al. argued that a given change should be more probable if it requires fewer rather than more edits, and deletions rather than insertions. For example, Gibson et al. tested the double object (DO)/prepositional object (PO) alternation, in which the DO structure, e.g. "The doctor sent the child the nurse," is a single dropped "to" away from the PO structure, e.g. "The doctor sent the child to the nurse," and correspondingly, the PO structure is a single inserted "to" from the DO structure. If deletions are more likely than insertions, then DOs should more often be assumed to be distorted POs than vice versa. Gibson et al. also included the active/passive alternation, in which an active, e.g. "The girl winked at the boy," is two deletions ("was" and "by") away from the passive, e.g. "The girl was winked at by the boy," and the passive is correspondingly two insertions away from an active. Assuming that distortions involving only one edit are more likely than distortions involving two, readers or listeners should more often assume that a DO or PO sentence has been distorted than that a passive or active sentence has been distorted.

Consistent with their predictions, Gibson et al. found that across the five syntactic alternations they tested, participants were more likely to interpret sentences as if they had been distorted from the alternative form when: (1) the alternative form had a more plausible meaning, and (2) the distortion required fewer edits (e.g. DO/PO vs. active/passive), which took the form of deletions rather than insertions (e.g. DOs vs. POs). To investigate whether short-term contextual manipulations of the prior (i.e. the likelihood that a sentence was intended) and noise terms (i.e. the likelihood of distortion) affected comprehension, Gibson et al. ran the same study twice more with adjustments to the filler sentences. In one study, they increased the proportion of semantically incoherent fillers. This should have signaled

that incoherence was more likely to be intended. Consistent with this, in this study, participants more often interpreted semantically incoherent sentences as having been intended and not distorted (i.e. there were more correct responses). In the other study, Gibson et al. included typos and missing words in many of the fillers. This should have increased participants' estimation of the amount of noise in the experiment, and consistent with this, Gibson et al. found an increase in the rate of answers consistent with an assumption of distortion (i.e., incorrect responses).

It is important to note that the factors that Gibson and colleagues manipulated to modulate the likelihood of distortion and the prior probability of a sentence are not the only factors that influence these probabilities. As discussed by Gibson et al. (2015), the frequency of a structure is also an important cue to how likely it is to be intended. This is particularly important for the active/passive alternation, because the differences in frequency and distortion likelihood across these structures favor different structures. Actives are more frequent and thus more likely than passives, but actives are also more likely to be distortions of passives than vice versa, because distortion to an active requires two deletions rather than two insertions and deletions are more likely than insertions. Ultimately, the likelihood of two-edit distortions is very small, so the frequency-driven difference in the likelihood of intention favoring the active structure likely usually drives comprehension. This is consistent with the finding that passive sentences are more likely to receive interpretations that are not faithful to the input (i.e., incorrect interpretations), not only among PWA but also among neurotypical adults and children (Ferreira, 2003; Slobin, 1966; see also Gibson et al., 2015).

A rational inference approach to aphasic sentence comprehension

The rational inference approach predicts at least three distinct patterns of effects in aphasic comprehension. First, PWA should show effects of prior probability and distortion likelihood like the ones that neurotypical comprehenders showed in Gibson et al. (2013)'s studies. Second, following the argument in Gibson et al. (2015), if PWA's sentence comprehension impairments increase the likelihood that their mental representations of the input form (sentence structure) are distorted, then PWA's interpretations should show stronger effects of prior sentence probability than neurotypical controls' do. Third, this should hold not only for PWA in general, but also at the level of individual PWA. PWA whose mental representations of sentence structure are more likely to be distorted should be influenced more heavily by the prior probability of a sentence during interpretation than PWA whose mental representations of sentence structure are less likely to be distorted. Similarly, PWA who are more impaired at judging semantic coherence should be more heavily influenced by input form than PWA who are less impaired at judging semantic coherence. The first two of these predictions were tested in Gibson et al. (2015). The third has not previously been tested.

Gibson et al. (2015) tested a sample of 8 PWA, as well as neurotypical older and younger controls, on semantically coherent and semantically incoherent versions of PO/DO and active/passive structures in an act-out task. The rational inference account predicts that both PWA and controls should be more likely to interpret DO structures than PO structures as if they had been distorted, because the likelihood of a DO being distorted is higher than the likelihood of a PO being distorted (Gibson et al., 2013). Recall that distorted interpretations

are functionally equivalent to incorrect interpretations: they are interpretations that do not correspond to the observed/input structure. Other models of aphasic comprehension make this same prediction for PWA, namely that their performance should be better on PO structures than on DO structures. Mapping theories predict this pattern because PO structures involve a more ‘transparent’ mapping from syntactic order to interpretation (see O’Grady & Lee, 2005). The Lexical Integration Deficit hypothesis (Thompson & Choy, 2009; Meyer, Mack & Thompson, 2012; Mack, Ji & Thompson, 2013) predicts this pattern because PO structures allow PWA to assign a predicted thematic role to the post-verbal NP immediately, whereas this role assignment must be revised in DO structures, requiring additional thematic integration operations. Gibson et al.’s PWA and neurotypical participants performed better on POs than DOs, consistent with these predictions.

The rational inference account also predicts that PWA should be more likely than age-matched controls to interpret semantically incoherent sentences as if they had been distorted (i.e. incorrectly). This prediction is not inconsistent with any other models of aphasic comprehension – this observation is explained by most or all existing models – but it is foundational to the rational inference account. Gibson et al. (2015) found a marginal effect of semantic coherence on PWA’s interpretations of the active and passive items in the predicted direction, but no such effect for the controls. This hints at a stronger effect of prior sentence probability for the PWA than controls.

Finally, the rational inference account predicts that PWA and controls should be more likely to interpret DO or PO sentences as if they had been distorted (i.e., incorrectly) than active or passive sentences. This pattern is unexpected under other current models of aphasic comprehension performance. All existing models of aphasic sentence comprehension predict that PWA will perform more poorly with passives than actives, including structural-impairment models (e.g., the Trace Deletion Hypothesis, Grodzinsky, 2000; the Double Dependency Hypothesis; Mauner et al., 1993; Relativized Minimality models, Garraffa & Grillo, 2008; Varlakosta et al., 2014, the Intervener Hypothesis, Sheppard, et al., 2015; Sullivan et al, 2016; slowed-syntax models, Burkhardt et al., 2003; the Derived Order Problem Hypothesis, Bastiaanse & van Zonneveld, 2006), mapping theories (e.g. Schwartz et al, 1987; O’Grady & Lee, 2005), lexical-impairment models (e.g., the Delayed Lexical Access Model, Love et al., 2008; Ferrill et al., 2012; the Lexical Integration Deficit hypothesis, Meyer et al., 2012; see also the Lexical Bias Hypothesis; Gahl, 2002), and general processing deficit models (Caplan, et al., 2007; Gutman, et al, 2011; Hula & McNeil, 2008; Murray, 2012). As noted above, some current models also predict that PWA will perform more poorly on DO than PO structures. However, no current models of aphasic comprehension performance make strong predictions regarding the relative degree of impairment for passive versus DO/PO structures. The rational inference account is unique in predicting that DO/PO structures will elicit more incorrect responses than passive structures. Gibson et al. (2015) found some support for this prediction: PWA (as well as younger and older neurotypical adults) were more likely to interpret DO/PO sentences as if they had been distorted than active/passive sentences, consistent with sensitivity to a higher likelihood for distortions requiring a single edit than ones requiring two edits.

Gibson et al.'s (2015) results are exciting and suggestive, but not definitive. One concern is that the results provided only weak evidence that compared to controls' interpretations, PWA's interpretations were more strongly influenced by prior sentence probability. The lack of a reliable interaction between group and the influence of semantic coherence in Gibson et al.'s study could have been due to the small sample size, so it will be important to test for this in a larger sample. A second concern is that the act-out task used in Gibson and colleagues' study might have affected the PWA's performance. Act-out or object-manipulation tasks are more challenging for PWA than traditional sentence-picture matching tasks (Caplan et al., 2007; Kiran, Caplan, Sandberg, Levy, Berardino et al., 2012), which may have increased the number of incorrect responses they provided. Converging evidence from sentence-picture matching (a task which is also used clinically to assess syntactic disorders) would strengthen the case that PWA's behaviors match the predictions of the rational inference account. Finally, Gibson et al. did not test the rational inference account's prediction that the degree of an individual PWA's impairment should be related to his or her likelihood of settling on an interpretation that is faithful to the sentence's structure, versus one that is driven by its prior probability.

The current study

The current study tested the same sentence types examined by Gibson et al. (2015), but it expanded on this previous study in four ways. First, it tested a considerably larger sample of PWA ($n=16$), who varied widely in their severity of language and conceptual-semantic impairments. Second, it directly addressed individual differences in sentence-comprehension performance, and tested the prediction that individual PWA's performance would be moderated by their degree of syntactic or semantic impairment. Third, it included a novel manipulation of the degree of semantic incoherence. This had two benefits: (1) the number of active/passive items was twice the number in Gibson et al., increasing power, and (2) this manipulation allowed us to investigate the nature of the function relating degree of semantic incoherence to a sentence's prior probability. As discussed above, there is some evidence that PWA show stronger effects of semantic cues when an alternate meaning describes an impossible, rather than simply unlikely, event (e.g., Saffran et al., 1998). All previous experiments testing the rational inference account compared alternate meanings that were possible or impossible (i.e. Gibson et al., 2013; Gibson et al., 2015). We additionally tested sentences describing implausible, but not impossible, events. If the prior probability of a sentence is influenced by the likelihood of the event it describes, and implausible events are more likely than impossible events, then implausible sentences should be more likely to be intended (have a higher prior probability) than impossible sentences. If this is the case, comprehenders should be less likely to settle on interpretations that are faithful to the perceived input for impossible sentences than implausible sentences. A fourth way the study extended Gibson et al. (2015) is that, instead of using an act-out task, it used a less-demanding sentence-picture matching task. This task is commonly used in clinical assessments and is directly comparable to off-line measures of sentence comprehension accuracy used in other aphasia studies, including sentence-picture verification (e.g., Schwartz, Saffran, Linebarger & Pate, 1987) and plausibility or acceptability judgment tasks (Gahl, Menn, Ramsberger, Jurafsky, Elder et al., 2003; Linebarger, Schwartz & Saffran, 1983). Because distorted interpretations are functionally equivalent to incorrect

interpretations, for the remainder of the paper we will refer to distorted interpretations as incorrect and describe PWAs' performance in terms of accuracy.

Method

Participants

All participants provided informed consent prior to completing any study procedures and were compensated \$10 per hour. They were all native English speakers and all passed a 40dB pure-tone hearing screen (unaided) at 500, 1000, 2000, and 4000 Hz bilaterally.

There were two groups of participants. The first was made up of 16 community-dwelling older adults (15 female) with normal or corrected-to-normal vision and a self-reported lack of speech-language, hearing, or neuropsychological disorders. Their ages ranged from 47 to 69 years (mean: 60.8), and they had between 14 and 18 years of education (mean: 16.7). These older adults completed the Mini-Mental State Exam (MMSE: Folstein, Folstein & McHugh, 1975) and all of their scores (minimum: 29; mean: 29.8) were above lower-quartile cutoff scores for healthy older adults (Bleecker, Bolla-Wilson, Kawas & Agnew, 1988). They also completed Raven's Colored Progressive Matrices (RCPM: Raven, 1965), with scores ranging from 29 to 36 out of 36 (mean: 33.1).

The second group of participants was 16 PWA (6 female) who ranged in age from 50 to 82 (mean: 67.1) and had 12 to 21 years of education (mean: 16.1). They were between 19 and 151 months post onset of aphasia. All PWA were referred from the Western Pennsylvania Participant Registry, a registry of community-dwelling stroke survivors in the greater Pittsburgh area. They were all pre-morbidly right-handed, and had chronic aphasia subsequent to left-hemisphere stroke. See Table 1 for demographic data for all PWA who took part in this study.

The results of standardized language and conceptual-semantic processing testing for the PWA are presented in Table 2. They completed the Comprehensive Aphasia Test (CAT: Swinburn, Porter & Howard, 2004) to characterize their language impairments. The CAT provides a measure of overall aphasia severity (the Mean Modality T-Score, centered at 50 with a standard deviation of 10) as well as sub-scores that measure comprehension and production performance at different levels of language (word, sentence, discourse). A mean modality T-score of 68.2 is the cutoff for the presence of aphasia on the CAT (Swinburn et al., 2004). The participants had mild to moderate overall language impairments, with mean modality T-Scores ranging from 37.4 to 66.1 (M: 54.6). The CAT Spoken Sentence Comprehension T-Score provides a measure of participants' sentence-level listening comprehension impairments. Scores on this portion of the CAT are based on PWAs' ability to comprehend simple and syntactically complex sentences, including passives and sentences with center-embedded relative clauses. Participants spoken sentence comprehension impairments varied, with spoken sentence comprehension T-Scores ranging from 44 to 72 (M: 56.4). Given that spoken sentence comprehension performance contributes to mean modality T-scores, it is not surprising that these measures were highly correlated in this sample of PWA ($r=.85$, $p<.001$). In addition to the CAT, participants completed two measures of conceptual-semantic processing: Pyramids and Palm Trees,

three-picture version (PPT: Howard & Patterson, 1992), and the Kissing and Dancing Test (KDT: Bak & Hodges, 2003). These two picture-based tests measure participants' access to object-related (PPT) and action-related (KDT) conceptual-semantic knowledge. Scores on PPT ranged from .46 to .98 (mean=.88), and scores on KDT ranged from .48 to .96 (M=.82).

To directly index their event knowledge, participants also completed an event task, during which they made judgments about the semantic coherence of photographs of normal and abnormal events (stimuli taken from Proverbio & Riva, 2009). In this task, participants judged as quickly as possible whether photographs showed something that “might normally happen” or not. Scores on the Event task ranged from .48 to .95 (mean=.85). Within the current sample, KDT and Event scores were highly correlated ($r=.74$). Although performance on conceptual-semantic measures and language-impairment measures dissociated in at least one previous sample of PWA (e.g. Dickey & Warren, 2015), in the current sample they were highly correlated (KDT vs. CAT Mean Modality T-Score: $r=.78$, $p<.001$; Event vs. CAT Mean Modality T-Score: $r=.72$, $p<.005$).

Design and Materials

Three sub-experiments and a set of control items were included in the current study. The items for the first two sub-experiments had previously been used in Gibson et al. (2015). The 20 items in the first sub-experiment crossed the double object/prepositional object (DO/PO) structural alternation with semantic coherence implemented as plausible versus impossible (1a-d below).

- (1a) DO/Plausible: The mother gave the girl the candle.
- (1b) DO/Impossible: The mother gave the candle the girl.
- (1c) PO/Plausible: The mother gave the candle to the girl.
- (1d) PO/Impossible: The mother gave the girl to the candle.

These DO/PO items tested the likelihood of assuming distortion in the case of a single edit, because the DO structure is a single deletion (dropped “to”) from the PO structure, and the PO structure is a single insertion (added “to”) away from a DO structure.

The 20 items in the second sub-experiment crossed active/passive structure and semantic coherence implemented as plausible versus impossible (2a-d below). These items tested the likelihood of assuming distortion in the case of two edits, as the active structure is two deletions (dropped “was” and “by”) from the passive structure and the passive is two insertions from the active.

- (2a) Active/Plausible: The girl kicked the ball.
- (2b) Active/Impossible: The ball kicked the girl.
- (2c) Passive/Plausible: The ball was kicked by the girl.
- (2d) Passive/Impossible: The girl was kicked by the ball.

The 20 items in the third sub-experiment had the same active/passive versus semantic coherence design as those in the second experiment, except that in these items the semantic

coherence manipulation contrasted plausible with implausible instead of impossible (3a-d below). A comparison between the results of the second and third sub-experiments will therefore be informative about the relationship between event likelihood and the prior probability of a sentence.

(3a) Active/Plausible: The cat licked the girl.

(3b) Active/Implausible: The girl licked the cat.

(3c) Passive/Plausible: The girl was licked by the cat.

(3d) Passive/Implausible: The cat was licked by the girl.

A final set of 10 control items compared semantically reversible actives and passives (4a&b below). These items served as fillers that would force participants to attend to the syntactic form of stimuli, since their semantic content was uninformative regarding their intended interpretation.

(4a) The man held the woman.

(4b) The man was held by the woman.

Thirteen additional filler items were also included. These had a variety of structures, but their meanings were all plausible.

Two black and white line drawings were created for each item, corresponding to the two events described across conditions. For example, for item 2 above, there were drawings of a girl kicking a ball and a ball kicking a girl (see Figure 1).

These pictures were drawn using a cartoon-like style, to enable participants to visualize even highly implausible or impossible actions (for example, the ball kicking the girl in Figure 1 has a cartoon face and limbs). Note that this should in principle work against the hypothesized effect of semantic (in)coherence, since the style was suggestive of a cartoon-like world where real-world knowledge need not apply, and both coherent and incoherent events were depicted in a similar style.

For each condition within an item, one of the pictures corresponded to the observed sentence's meaning and the other corresponded to the meaning of the potential intended sentence, assuming the observed sentence had been distorted from the alternate form in its alternation. The picture corresponding to the observed sentence's meaning appeared equally often on the left and right sides of the screen.

Four counterbalancing lists were created using a Latin Square design, such that each participant saw only one condition of each item, and an equal number of each condition for each sub-experiment. This meant that there were five observations per condition in each sub-experiment. Participants saw two blocks of items, of roughly equal length, and had the opportunity to take a break in between blocks. Additionally, each counterbalancing list was presented in one of two versions, which differed only in that they switched the order of the two blocks of trials. The order of items within each block was randomized for each participant. Across the entire experiment, participants encountered 53 sentences with meanings that were natural or plausible under a literal reading, 20 sentences with meanings

that were impossible under a literal reading, and 10 sentences with meanings that were implausible under a literal reading.

To verify: (1) the reversibility of the semantically reversible items, and (2) the possibility and plausibility status of the items in Sub-experiments 2 and 3, we ran questionnaire norming studies in which 27 University of Pittsburgh undergraduates first indicated whether the active form of the item described a possible or impossible event (subsequently coded as possible = 1, impossible = 0) and immediately afterwards rated the event's naturalness on a scale of 1 (extremely natural) to 7 (extremely unnatural).¹ Repeated measures ANOVAs (F1 treating participants and F2 treating items as random factors) indicated that the meanings of the reversible control sentences were similarly possible and natural (Possibility for each meaning $M=.98$; Naturalness: for active meaning $M=2.31$, for passive meaning $M=2.35$; all $F_s < .1$). For Sub-experiments 2 and 3, separate ANOVAs across naturalness ratings and possibility ratings crossing the factors of sub-experiment (2 vs. 3) and semantic coherence (i.e. plausible vs. implausible/impossible) confirmed that, as designed, the possibility manipulation in Sub-experiment 2 implemented a stronger contrast of semantic coherence than the plausibility manipulation in Sub-experiment 3. This is evident in reliable interactions between sub-experiment and semantic coherence for both possibility judgments ($F(1,24)=187.46$, $p<.001$; $F(1,19)=137.56$, $p<.001$) and naturalness judgments ($F(1,9)=53.89$, $p<.001$; $F(1,19)=37.15$, $p<.001$). Accompanying main effects indicated that plausible sentences were rated both reliably more possible than semantically incoherent sentences (Plausible Sub-expt 2: $M=.996$, Plausible Sub-expt 3: $M=.97$; Impossible Sub-expt 2: $M=.028$, Impossible Sub-expt 3: $M=.68$; $F(1,24)=592.62$, $p<.001$; $F(1,19)=523.64$, $p<.001$) and reliably more natural than semantically incoherent sentences (Plausible Sub-expt 2: $M=1.54$, Plausible Sub-expt 3: $M=1.95$; Impossible Sub-expt 2: $M=6.6$; Impossible Sub-expt 3: $M=4.67$; $F(1,9)=374.76$, $p<.001$; $F(1,19)=450.01$, $p<.001$).

Procedure

The current experiment was run as part of a multi-experiment testing battery. Older healthy adults were tested in one three-hour session in a laboratory at the University of Pittsburgh. PWA were tested in two three-hour sessions, separated by one to two weeks. They were tested either in a laboratory at the University of Pittsburgh or in their home. Stimuli were presented in E-prime (Schneider, Eschman & Zuccolotto, 2002) on a Dell laptop computer. The experiment began with three practice trials. On each trial, participants saw a fixation cross, centered on the screen. This cross remained on the screen until they pressed the spacebar to reveal two pictures, one depicting the experimental item's syntactically faithful (correct) interpretation and one depicting its syntactic alternation's interpretation. After 1000 milliseconds, participants heard audio of the sentence, recorded by a female native speaker of American English. The two pictures remained on the screen until participants responded by pressing the 'C' or 'N' button on the keyboard to indicate which picture best matched the sentence they had heard. The 'C' button (on the left-hand side of the keyboard) indicated

¹Full compliance with the instructions was rare: 17 participants failed to indicate naturalness ratings for items that they judged impossible and 2 failed to ever rate possibility. This left 25 participants in the possibility analyses. For reversible items, naturalness ratings were unaffected by data loss, but analyses over naturalness ratings for Sub-experiments 2 and 3 were conducted over only the 10 participants who had rated naturalness for all items.

that the left-hand picture best matched the sentence, while the 'N' button (on the right-hand side of the keyboard) indicated that the right-hand picture best matched the sentence. Participants had unlimited time to respond.

Data Analysis

Analyses of the experimental data were done in R (R Development Core Team, 2013; ver 3.0.1) with the lme4 package (Bates, Maechler, & Bolker, 2013; ver. .999999-2) using linear mixed effect logit models with participants and items as crossed random factors (Baayen, 2008). Fixed factors were deviation coded, with values of .5 and -.5. When maximal models failed to converge, the random slopes that captured the least variance were dropped until the model converged. Picture choices consistent with a correct response were coded with a 1; trials on which participants chose the alternative picture were coded with a 0. Given this coding, higher means indicate more accurate responses.

To evaluate whether PWA more heavily weighted sources of evidence that were more reliable for them, we ran analyses including measures of individual PWA's syntactic and semantic processing abilities. To index syntactic ability, we used Spoken Sentence Comprehension T-scores from the Listening Comprehension portion of the CAT (scores were divided by 100 to make their scale more similar to those of the other factors included in our models and then centered). To index semantic ability, we used a composite measure (we will refer to it as KDTE) generated by averaging each participant's KDT and Event task scores. We focused on KDT and Event because they are designed to evaluate knowledge regarding actions and events, which are critical to evaluating the semantic coherence of sentences in the current study. PPT, on the other hand, evaluates knowledge regarding objects, which is less directly relevant to judging semantic coherence in the current study. The fact that KDT and Event scores were highly correlated across our sample ($r=.74$) suggests that they both tap similar knowledge, but their less than perfect correlation suggests that each also measures at least some unique knowledge. Given that many aspects of action and event knowledge are likely to contribute to the evaluation of semantic coherence, we deemed it best to use a composite measure that reflected the full extent of the identified action and event knowledge impairment in our sample, regardless of which test identified it. These KDTE scores were centered before analysis. Within the current sample of participants, KDTE and Spoken Sentence Comprehension T-scores were correlated ($r=.58$, $p<.02$). We also ran individual difference analyses over the combined data from Sub-experiments 1–3, involving DO/PO structures (Sub-experiment 1) and active/passive structures (Sub-experiments 2–3). The structure factor was deviation coded such that active and passive sentences were assigned .5, whereas DO and PO sentences were assigned -.5. This coding was chosen because the rational inference account predicts a higher likelihood of distortion in the DO and PO structures (they require only one insertion or deletion) than for actives and passives (which require two insertions or deletions).

Results

Means for Sub-experiment 1 (involving PO/DO structures) appear in Figure 2. A model testing the three-way interaction among participant group, structure, and semantic coherence

(including random intercepts for both participants and items, and a random slope of the interaction between semantic coherence and structure for participants) showed three main effects: (1) a main effect of group, such that the control participants ($M = .87$) had more correct responses than the PWA ($M = .62$; $\beta = 2.43$, $SE = .65$, $z = 3.76$; $p < .001$); (2) a main effect of structure, such that all participants had more correct responses for PO structures ($M = .84$) than DO structures ($M = .65$; $\beta = 2.24$, $SE = .47$, $z = 4.80$; $p < .001$); and (3) a main effect of semantic coherence, such that all participants had more correct responses for plausible sentences ($M = .93$) than impossible sentences ($M = .57$; $\beta = 3.05$, $SE = .51$, $z = 5.98$; $p < .001$). No interactions were reliable.

Means for Sub-experiment 2 (active/passive structure crossed with possibility) appear in Figure 3, and means for Sub-experiment 3 (active/passive structure crossed with plausibility) appear in Figure 4. We entered the data from both these sub-experiments into a single model testing the four-way interaction of sub-experiment, group, structure, and semantic coherence. This model would support only limited random effects structure and still converge, so the final model included a random slope of the interaction of structure and semantic coherence for participants, and random intercepts for both participants and items. This model revealed the following reliable effects: (1) a main effect of group, such that the control participants had more correct responses ($M = .98$) than the PWA ($M = .75$; $\beta = 4.07$, $SE = 1.03$, $z = 3.94$; $p < .001$) and (2) an interaction between group and semantic coherence, such that control participants' responses were less affected by semantic coherence (Sem+ $M = .98$, Sem- $M = .98$) than PWA's interpretations (Sem+ $M = .88$, Sem- $M = .62$; $\beta = -2.74$, $SE = 1.20$, $z = -2.28$; $p = .023$).

Means for the reversible active/passive control items are presented in Figure 5. We tested for effects of structure over only the PWA's data because models including the controls' data would not converge because controls had almost no incorrect responses. Consistent with many previous findings (e.g. Grodzinsky, 1986, 2000; Schwartz et al, 1980; Saffran et al., 1998), a model with maximal random effects structure showed a reliable effect of structure ($\beta = 1.32$, $SE = .44$, $z = 3.02$; $p < .005$), such that PWA had more correct responses for actives than passives.

To test whether PWA were more likely to provide incorrect responses for DO and PO structures than for actives and passives, and whether higher sentence comprehension T-scores were associated with smaller effects of semantic coherence and larger effects of structure, we ran a model testing for main effects of sentence comprehension T-scores, structure (active/passive versus DO/PO), and semantic coherence, as well as interactions between T-scores and structure and T-scores and coherence, and including random intercepts of participants, items, and sub-experiment, random slopes of structure and semantic coherence for participants, and random slopes of both interactions and the main effect of T-score for items. This model showed main effects of structure ($\beta = 1.02$, $SE = .22$, $z = 4.58$, $p < .001$) and semantic coherence ($\beta = 2.13$, $SE = .31$, $z = 6.85$, $p < .001$), such that active/passive structures and plausible conditions (with higher semantic coherence) elicited more correct responses. There was no hint of an interaction between sentence-comprehension T-score and semantic coherence ($p > .8$). However, there was a positive interaction between T-score and

structure ($\beta = 9.93$, $SE = 3.06$, $z = 3.25$, $p < .005$), such that higher T-scores were predictive of more correct responses for active and passive structures relative to DO and PO structures.

To test whether higher KDTE scores were associated with larger effects of semantic coherence and smaller effects of structure, we ran a model like the one described in the paragraph above, but replaced semantic comprehension T-scores with KDTE scores. The final model included random intercepts of participants, items, and sub-experiment, random slopes of semantic coherence for participants, and random slopes of both interactions for items. This model showed reliable main effects of semantic coherence ($\beta = 2.12$, $SE = .31$, $z = 6.85$, $p < .001$) and structure ($\beta = .90$, $SE = .21$, $z = 4.22$, $p < .001$), such that plausible conditions and active/passive structures elicited more correct responses. Neither interaction was reliable.

Discussion

The current paper tested three predictions of the rational inference, or noisy channel, model as applied to aphasic comprehension: (1) that PWA, like neurotypicals, should be sensitive to sentence likelihood and distortion likelihood, (2) that higher rates of sentence distortion associated with language impairment should lead PWA to rely more heavily on the prior probability of a sentence for interpretation than neurotypicals do (e.g. Gibson et al., 2015), and (3) that this increased reliance on prior probability should also be evident at the level of the individual, and reflect the degree of an individual's language impairment. Additionally, it examined the way that manipulating prior probability via a modulation of degree of semantic coherence (implausible vs. impossible events) affected PWAs' and neurotypicals' interpretation (e.g., Saffran et al., 1998; Warren & McConnell, 2007).

The results of the current experiments confirmed the first two predictions. The current study, which tested PWA and controls with a sentence-picture-matching task, replicated and extended findings from Gibson et al. (2013) for young healthy participants doing a questionnaire task, and findings from Gibson et al. (2015) for PWA and controls doing an act-out task. Both PWA and age-matched controls showed effects of the likelihood of distortion. Both groups provided more correct responses (i.e., chose pictures consistent with the stimulus sentence's structure) when a single deletion rather than a single insertion was involved. This was evident in the main effect of structure in Sub-experiment 1: all participants had more correct interpretations for PO structures than DO structures. Additionally, in an analysis across Sub-experiments 1–3, PWA and controls provided fewer correct responses for the DOs and POs, which only require a change of one word, than for the active/passive alternation, which involves a change of two words. The finding of more incorrect responses for DOs and POs than for actives and passives is consistent with critical predictions of the rational inference account, as well as its assumption that distortion likelihood strongly affects comprehenders' likelihood of choosing a correct or an incorrect interpretation.

However, this finding for PWA is unexpected under most models of aphasic comprehension performance. As noted in the Introduction, all existing accounts of aphasic sentence comprehension predict that PWA should perform worse on passives than actives. The current

study found this effect for reversible actives and passives, but not for the semantically coherent and incoherent actives and passives in Sub-experiments 2 and 3. This was a failure to replicate Gibson et al. (2015), who did find a main effect of structure for active versus passive sentences, even in the presence of strong semantic cues. The reason for the current finding is unclear. Some accounts of aphasic comprehension impairment (such as the Lexical Integration Deficit hypothesis, Meyer et al. 2012, or mapping theories, O'Grady & Lee, 2005) predict that PWA should perform worse on DO than on PO structures, and do not necessarily predict worse performance on passives than DO/POs. However, none of those accounts make a strong prediction that PWA should have more incorrect responses for PO and DO structures than for passives. The fact that performance on actives and passives did not differ in the analysis comparing performance on active/passive and DO/PO structures reinforces the conclusion that performance was better on passives than on DOs and POs because it shows that the effect of active/passive versus DO/PO did not result because high accuracy for the active structures raised the mean accuracy for the active/passive structures. It is possible that resource-related accounts (Caplan et al., 2013; Gutman et al., 2011; Hula & McNeil, 2008; Murray, 2012) might be consistent with the current findings if the resource demands imposed by three-argument verbs are measurably greater than those imposed by passive structures, but this is speculative.

It is important to consider potential disadvantages of using a sentence-picture matching task in the current experiment. This task is binary and only provides a single possible response that is not consistent with the input structure. If participants were likely to settle on other interpretations, this would be a serious concern. However, this concern is alleviated by the fact that when the PWA in Gibson et al. (2015)'s act-out task provided responses that were not consistent with the input sentence's structure, they almost always generated a meaning consistent with its alternate structure (99.4% for active/passive alternation, 87% for PO/DO alternation). Another possible concern is that the forced-choice nature of this task and the relatively small number of items within each sub-experiment may have increased the influence of noise in the PWA data. There are at least two reasons that noise is unlikely to have driven the current effects. First, given that there were no differences between the two active/passive sub-experiments, all analyses grouped them together. This doubled the number of items contributing to active/passive analyses, making those data patterns more robust to guessing. Second, three of four condition means for the PWA in the DOPO sub-experiment were quite extreme, near either 20% or 80% literal interpretations. This indicates that on average PWA interpreted one of five or four of five sentences in a given way. Given the extremity of these means, as well as the fact that the pattern of data was similar to that of the neurotypical controls in the current study and the PWA in Gibson et al. (2015), it seems unlikely that the data patterns of the PWA resulted from noise.

Consistent with the predictions of rational inference account, other theories of aphasic comprehension, and previous findings that semantics can influence sentence interpretations (e.g. Ferreira, 2003; Schwartz et al., 1980), both PWA and healthy controls showed effects of semantic coherence. Both groups were more likely to choose an incorrect interpretation when it had a higher prior probability than the correct interpretation, as evidenced by the reliable main effect of semantic coherence for the DO and PO structures of Sub-experiment 1.

The rational inference account critically predicts that PWA should be both less likely to choose interpretations that are consistent with a sentence's structure (i.e., they should be incorrect more often) and more strongly influenced by prior probability than controls. The findings of the current study were consistent with both of these predictions: PWA chose incorrect interpretations more often than neurotypical controls, as evidenced by the main effects of group across all sub-experiments, and they were more strongly influenced by semantic coherence, as evidenced by the reliable interaction of group and semantic coherence across the active/passive Sub-experiments 2 and 3. The main effect of group replicated Gibson et al. (2015), but the reliable interaction provided stronger evidence for the predicted trade-off between form and sentence likelihood than Gibson et al. (2015) found in their smaller sample. The similarity in the findings across the current study and Gibson et al. (2015), which used an act-out task, suggests both that these patterns of behavior were robust to variations in the cognitive demand imposed by the task, and that limiting the potential interpretations to just two in the current sentence-picture-matching task did not substantially influence the pattern of results.

Additional support for the strong effect of meaning likelihood on PWA's comprehension came from a comparison across the reversible and non-reversible actives and passives in this study. According to the rational inference account, the elimination of semantic cues in the reversible sentences forces the estimation of the prior probability of an interpretation to be based only on its structure, e.g. the fact that actives are more frequent and thus more likely than passives. The data from reversible actives and passives thus provide a baseline against which to evaluate the effects of adding semantic cues regarding the prior probability of an interpretation. It is possible that the importance or salience of the semantic cues was artificially inflated in the current study because the two interpretations provided for each non-reversible sentence were always perfectly distinguished by semantic coherence. However, this was the case for both controls and PWA, so any artificiality was consistent across groups. The controls' performance was unaffected by the addition of semantic cues regarding the prior probability of an interpretation: their performance on reversible and non-reversible actives and passives was almost identical. This is consistent with the rational inference account's prediction that for structures with a relatively low likelihood of distortion (like actives and passives, which require two edits), interpretation will rely less on prior probability. On the other hand, PWA, whose impairments should increase the likelihood of distortion, should have relied more heavily on prior probability. If semantic cues are stronger or more reliable indicators of prior probability than the purely structural cues available in the reversible sentences, then their addition should have strongly influenced PWA's interpretations. It did. From a baseline accuracy of 68% on reversible passives, PWA's performance increased to 89% for semantically coherent passives and decreased to 60% on semantically incoherent passives. From a baseline accuracy of 86% on reversible actives, PWA's performance increased to 89% for semantically coherent actives and decreased to 64% on semantically incoherent actives. These changes in performance are consistent with the predictions of the rational inference account.

The strong effect of semantic coherence on PWA's performance in the current study is consistent with other theories of aphasic sentence-comprehension impairments. However, none of those other theories provide a specific explanation of *why* semantic information has

such a strong effect. Some models claim that guessing or non-linguistic heuristics are responsible for the alternative interpretations that PWA choose (see Grodzinsky, 2000; Meyer et al., 2012), but most make no claims about the mechanisms involved. In contrast, the rational inference account provides a direct explanation of why semantic information has the effect it does: semantic information affects the prior probability of an interpretation, which directly influences the likelihood that that interpretation will be chosen. Thus, the rational inference account provides a more direct account of the semantic-coherence effects than the alternative accounts of aphasic sentence-comprehension deficits described above. Critically, this does not mean that the rational inference account and these alternative accounts are mutually exclusive. The rational inference account provides a framework for understanding trade-offs between the probabilities of distortion and intention. From this perspective, the goal of most theories of aphasic comprehension is to characterize both the mechanisms that lead to high levels of distortion for PWA and the particular set of structures that are affected by these mechanisms. Assuming that there exist internal cues that are informative as to whether a perceived sentence is likely to have been distorted, then the rational inference account and models of aphasic comprehension can be complementary and even potentially integrated. Under an integrated account, patterns of impairment predicted by models of aphasic comprehension should directly contribute to the rational inference account's noise term. Such an account would have the potential to explain both aphasic impairment and its effects on trade-offs in interpretation in a way that no models currently do.

It is important to consider the possibility that in the current study, PWA may have simply relied on semantic cues to interpretation when those cues were present instead of processing sentence structure. Such a strategy would have resulted in strong influences of semantic cues on interpretation, like the ones observed. However, the fact that the current results tracked the likelihood of distortion across structures (with PWA having lower accuracy on DO/PO sentences, which involve fewer edits, than on active/passive sentences, which involve more edits) even in the presence of strong semantic cues is not consistent with this kind of strategy. Also inconsistent with this view is the fact that the controls (whose syntactic processing abilities are unimpaired, and who were definitely influenced by syntactic structure in this experiment) showed similar effects of structure and semantic coherence on the items in the DO/PO sub-experiment.

Although many of the data patterns in the current experiment supported the predictions of the rational inference account, not all of them did. As discussed above, the rational inference account predicts that PWA's increased experience of structural distortion should make semantically-cued sentence intention a more reliable source of evidence for them (Gibson et al., 2015). The interaction between group and semantic coherence across active/passive Sub-experiments 2 and 3 indicated that, at the group level, PWA's interpretations were more likely to be sensitive to the likelihood that a sentence was intended than age-matched controls' interpretations. However, the current experiment found no evidence that optimization towards the more reliable source of evidence held at the level of individual PWA. When individual difference measures of PWA's syntactic ability (CAT Listening Sentence Comprehension T-scores) and semantic ability (indexed by a combination of KDT and Event task scores) were included as predictors in our statistical models, neither measure

ever interacted with semantic coherence. It is possible that this could have been due to the fact that the current experiment's sample size, though relatively large for an aphasia study, was relatively small for an individual differences study. Or alternatively, low power due to the relatively small number of observations per condition per individual may have impeded our ability to detect interactions. However, it is relevant that our models did have enough power to detect an interaction between CAT scores and structure. In an analysis of the PWA's data across all three critical sub-experiments, higher CAT scores predicted a greater fidelity to the form of the input in active and passive structures relative to the DO and PO structures. This interaction suggests that PWA with higher T-scores showed a pattern that is more similar to neurotypical controls. Specifically, PWA with higher T-scores were more likely to have interpretations that deviated from the input form when the likelihood of distortion was high, just like neurotypical controls. This is consistent with all theories that predict more normative performance with less severe syntactic impairment.

There are multiple ways to interpret the lack of finding of heavier weighting of a more reliable source of evidence at the individual level. A strong interpretation would be that this lack of finding indicates that the rational inference account cannot account for trade-offs in aphasic comprehension. Indeed, this is important to consider and may be correct. But it is also possible that the reason differences in weighting were evident in group analyses but not in individual analyses is that it might be easier to observe them for stronger contrasts. For example, changes in the weighting of sources of evidence might be more likely to occur to the relatively large difference in syntactic ability between PWA and neurotypicals, but be less sensitive to smaller, but real, variations in impairment across individual PWA. This possibility converges with the finding that participants' interpretations were unaffected by the impossibility versus implausibility manipulation implemented between Sub-experiments 2 and 3, and shown to be effective in norming. Whereas both controls and PWA's interpretations were sensitive to the gross difference between plausible sentences, on the one hand, and implausible or impossible sentences, on the other, they were not sensitive to the finer-grained difference between implausible and impossible sentences. More research into the conditions under which differences in the weighting of different sources of evidence are and are not observed will be critical for determining how to interpret the current failure to detect them at the individual level.

In summary, the current results replicated and extended the results of Gibson et al. (2015) in showing support for some of the critical and distinctive predictions of the rational inference account. However, the results did not support all of the rational inference account's predictions. This suggests that the rational inference account may be a promising explanation of features of aphasic comprehension, but more investigation is needed.

Acknowledgments

This research was supported by NIH grant R01DC011520 to the first and second authors and by grant number UL1TR000005 to the Clinical and Translational Science Institute of the University of Pittsburgh. It is the result of work supported with resources and the use of facilities at the VA Pittsburgh Healthcare System. We are grateful to Ted Gibson for letting us use the DO/PO and Impossible active/passive sentence stimuli that were used in Gibson, Sandberg, Fedorenko, Bergen & Kiran (2015), to Jane Xu for drawing the picture stimuli, to Molly Warmbein for recording the sentence stimuli, and to Rebecca Hayes, Abel Lei, Evelyn Milburn, and the LABlab (Language and Brain Lab) for help running participants. We also acknowledge helpful comments from the University of Pittsburgh

Reading and Language Group, Scott Fraundorf, and the audience at the 2014 Annual meeting of the Academy of Aphasia in Miami.

Works Cited

- Baayen, RH. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge, England: Cambridge University Press; 2008.
- Bak TH, Hodges JR. Kissing and dancing—a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. *Journal of Neurolinguistics*. 2003; 16(2):169–181.
- Bastiaanse R, van Zonneveld R. Comprehension of passives in Broca's aphasia. *Brain and Language*. 2006; 96(2):135–142. [PubMed: 16129480]
- Bates, DM., Maechler, M., Bolker, B. lme4: Linear mixed effects models using s4 classes. 2013. R package version 0.999999.2 <http://CRAN.R-project.org/package=lme4>
- Berndt RS, Mitchum CC, Wayland S. Patterns of sentence comprehension in aphasia: A consideration of three hypotheses. *Brain and Language*. 1997; 60(2):197–221. [PubMed: 9344477]
- Bleecker ML, Bolla-Wilson K, Kavas C, Agnew J. Age-specific norms for the Mini-Mental State Exam. *Neurology*. 1988; 38(10):1565–1565. [PubMed: 3419600]
- Burkhardt P, Piñango MM, Wong K. The role of the anterior left hemisphere in real-time sentence comprehension: Evidence from split intransitivity. *Brain and language*. 2003; 86(1):9–22. [PubMed: 12821412]
- Caplan D, Waters G, DeDe G, Michaud J, Reddy A. A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*. 2007; 101(2):103–150. [PubMed: 16999989]
- Caplan D, Michaud J, Hufford R. Dissociations and associations of performance in syntactic comprehension in aphasia and their implications for the nature of aphasic deficits. *Brain and Language*. 2013; 127(1):21–33. [PubMed: 24061104]
- Caramazza A, Zurif EB. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*. 1976; 3(4):572–582. [PubMed: 974731]
- Cho-Reyes S, Thompson CK. Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*. 2012; 26(10):1250–1277. [PubMed: 26379358]
- DeDe G. Verb transitivity bias affects on-line sentence reading in people with aphasia. *Aphasiology*. 2013a; 27(3):326–343. [PubMed: 23554543]
- DeDe G. Effects of verb bias and syntactic ambiguity on reading in people with aphasia. *Aphasiology*. 2013b; 27(12):1408–1425. [PubMed: 24683286]
- Dickey MW, Warren T. The influence of event-related knowledge on verb-argument processing in aphasia. *Neuropsychologia*. 2015; 67:63–81. [PubMed: 25484306]
- Ferreira F. The misinterpretation of noncanonical sentences. *Cognitive Psychology*. 2003; 47:164–203. [PubMed: 12948517]
- Ferrill M, Love T, Walenski M, Shapiro LP. The Time-Course of Lexical Activation During Sentence Comprehension in People With Aphasia. *American Journal of Speech-Language Pathology*. 2012; 21(2):S179. [PubMed: 22355007]
- Folstein MF, Folstein SE, McHugh PR. Mini-mental state A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–198. [PubMed: 1202204]
- Gahl S. Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology*. 2002; 16(12):1173–1198.
- Gahl S, Menn L. Usage-based approaches to aphasia. *Aphasiology*. 2016:1–17.
- Garraffa M, Grillo N. Canonicity effects as grammatical phenomena. *Journal of Neurolinguistics*. 2008; 21(2):177–197.

- Gibson E, Bergen L, Piantadosi ST. The rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*. 2013; 11:8051–8056.
- Gibson E, Sandberg C, Fedorenko E, Bergen L, Kiran S. A rational inference approach to aphasic language comprehension. *Aphasiology*. 2015; doi: 10.1080/02687038.2015.1111994
- Grodzinsky Y. Language deficits and the theory of syntax. *Brain and Language*. 1986; 27(1):135–159. [PubMed: 3947938]
- Grodzinsky Y. The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*. 2000; 23(01):1–21. [PubMed: 11303337]
- Gutman R, DeDe G, Caplan D, Liu JS. Rasch Model and Its Extensions for Analysis of Aphasic Deficits in Syntactic Comprehension. *Journal of the American Statistical Association*. 2011; 106(496):1304–1316.
- Howard, D., Patterson, KE. *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company; 1992.
- Hula WD, McNeil MR. Models of attention and dual-task performance as explanatory constructs in aphasia. *Seminars in speech and language*. 2008 Aug; 29(3):169–87. [PubMed: 18720315]
- Kiran S, Caplan D, Sandberg C, Levy J, Berardino A, Ascenso E, Tripodis Y. Development of a theoretically based treatment for sentence comprehension deficits in individuals with aphasia. *American Journal of Speech-Language Pathology*. 2012; 21(2):S88–S102. [PubMed: 22411773]
- Knillans J, DeDe G. Online Sentence Reading In People With Aphasia: Evidence From Eye Tracking. *American Journal of Speech-Language Pathology*. 2015; 24(4):S961–S973. [PubMed: 26383779]
- Love T, Swinney D, Walenski M, Zurif E. How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. *Brain and Language*. 2008; 107(3):203–219. [PubMed: 18158179]
- Mack JE, Ji W, Thompson CK. Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*. 2013; 26(6):619–636. <http://doi.org/10.1016/j.jneuroling.2013.04.002>. [PubMed: 24092952]
- Maurer G, Fromkin VA, Cornell TL. Comprehension and acceptability judgments in agrammatism: Disruptions in the syntax of referential dependency. *Brain and Language*. 1993; 45(3):340–370. [PubMed: 8269330]
- Meyer AM, Mack JE, Thompson CK. Tracking passive sentence comprehension in agrammatic aphasia. *Journal of Neurolinguistics*. 2012; 25(1):31–43. [PubMed: 22043134]
- Murray LL. Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech-Language Pathology*. 2012; 21(2):S51–S64. [PubMed: 22230179]
- O'Grady W, Lee M. A mapping theory of agrammatic comprehension deficits. *Brain and Language*. 2005; 92(1):91–100. [PubMed: 15582038]
- Proverbio AM, Riva F. RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*. 2009; 459(3):142–146. [PubMed: 19427368]
- R Development Core Team. *R: A language and Environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2013. URL <http://www.R-project.org>
- Raven, JC. *Guide to using the coloured progressive matrices: Sets A, Ab, B*. William Grieve & Sons; 1965.
- Saffran EM, Schwartz MF, Linebarger MC. Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language*. 1998; 62(2):255–297. [PubMed: 9576824]
- Schneider, W., Eschman, A., Zuccolotto, A. *E-prime computer software and manual*. Pittsburgh, PA: Psychology Software Tools Inc; 2002.
- Schwartz MF, Linebarger MC, Saffran EM, Pate DS. Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes*. 1987; 2(2):85–113.
- Schwartz MF, Saffran EM, Marin OS. The word order problem in agrammatism: I. Comprehension. *Brain and Language*. 1980; 10(2):249–262. [PubMed: 7407546]

- Sheppard SM, Walenski M, Love T, Shapiro LP. The auditory comprehension of Wh-questions in aphasia: support for the intervener hypothesis. *Journal of Speech, Language, and Hearing Research*. 2015; 58(3):781–797.
- Slobin DI. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*. 1966; 5(3):219–227.
- Sullivan N, Walenski M, Love T, Shapiro LP. The comprehension of sentences with unaccusative verbs in aphasia: a test of the intervener hypothesis. *Aphasiology*. 2016:1–15.
- Swinburn, K., Porter, G., Howard, D. CAT: comprehensive aphasia test. Psychology Press; 2004.
- Tenenbaum JB, Griffiths TL. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*. 2001; 24(04):629–640. [PubMed: 12048947]
- Thompson CK, Choy JJ. Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*. 2009; 38(3):255–283. [PubMed: 19370416]
- Varlokosta S, Nerantzini M, Papadopoulou D, Bastiaanse R, Beretta A. Minimality effects in agrammatic comprehension: The role of lexical restriction and feature impoverishment. *Lingua*. 2014; 148(0):80–94.
- Warren T, Dickey MW, Lei CM. Structural prediction in aphasia: Evidence from either. *Journal of Neurolinguistics*. 2016; 39:38–48. [PubMed: 27041821]
- Warren T, McConnell K. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*. 2007; 14(4): 770–775. [PubMed: 17972747]

Highlights

- Findings provide new evidence that PWA may engage in rational inference.
- PWA are sensitive to differences in probability of distortion across structures.
- PWA's impairment profiles did not predict their reliance on form vs. meaning cues.
- Impossibility and implausibility had similar effects on prior probability.

Figure 1.
Example picture stimuli accompanying item 2 above.

Figure 2.
Means for Sub-experiment 1 (error bars are standard error of the mean calculated over trials).

Figure 3. Means for Sub-experiment 2 (error bars are standard error of the mean calculated over trials).

Figure 4. Means for Sub-experiment 3 (error bars are standard error of the mean calculated over trials).

Figure 5. Means for Sub-experiment 4 (error bars are standard error of the mean calculated over trials).

Table 1

Demographic information, participants with aphasia (PWA)

Participant	Sex	Age at Testing	Months post-onset
201	F	71	39
202	F	70	63
203	M	62	68
204	F	68	69
205	F	54	71
206	F	50	94
207	F	64	30
209	M	65	70
211	M	74	19
212	M	78	120
213	M	62	58
214	M	80	136
216	M	67	52
217	M	67	151
218	M	60	26
219	M	82	23
<i>Mean</i>		67.1	68.1

Table 2

Language and conceptual-semantic testing data, participants with aphasia (PWA)

Participant	CAT Mean Modality T-score	CAT Spoken Sentence Comprehension T-score	Pyramids and Palm Trees	Kissing and Dancing Test	Event
201	45.8	49	0.92	0.92	0.74
202	56.9	58	0.96	0.88	0.91
203	56.6	50	0.83	0.96	0.94
204	59.6	61	0.96	0.92	0.93
205	63.6	65	0.87	0.96	0.95
206	50.1	49	0.88	0.73	0.92
207	66.1	72	0.98	0.92	0.94
209	62.0	61	0.85	0.87	0.91
211	45.6	57	0.85	0.6	0.68
212	62.1	58	0.98	0.96	0.82
213	44.4	50	0.79	0.63	0.9
214	62.4	61	0.98	0.83	0.92
216	47.9	52	0.87	0.85	0.86
217	53.1	57	0.92	0.79	0.87
218	37.4	44	0.46	0.48	0.48
219	60.1	58	0.94	0.88	0.84
<i>Mean</i>	54.6	56.4	0.88	0.82	0.85