

When circumstances change, update your pronouns

Joshua K. Hartshorne (Boston College), Mariela V. Jennings (Boston College), Tobias Gerstenberg (Stanford University), & Joshua Tenenbaum (Massachusetts Institute of Technology)

info@l3atbc.org

Language is frequently ambiguous, with the same sentence having several possible interpretations (*The children made delicious snacks*). A central challenge for the listener is to determine which of the possible intended meanings the speaker actually meant to convey. One particularly prevalent example is third-person pronouns. In principle, the pronouns in (1) and (2) could refer to either Al or Bart or any other male entity:

- (1) Al beat Bart at tug-of-war because he is strong.
- (2) Al beat Bart at tug-of-war because he is weak.

However, most people interpret the pronoun in (1) as referring to Al, and the pronoun in (2) as referring to Bart. Pairs of sentences like these pose a problem for theories of pronoun interpretation which have suggested that pronouns preferentially refer to the subject of the previous sentence (Gordon et al., 1993; Grosz et al., 1995), that pronoun interpretation is a function of verb biases (Arnold et al. 2000; Garvey & Caramazza, 1974; Hartshorne & Snedeker, 2013; Kaiser, 2011; Kehler & Rohde, 2013; Koornneef & van Berkum, 2006; Pyykkonen & Jarvikiivi, 2010; Stewart et al., 2000), and that the pronoun refers to the most causally-connected entity (Sagi & Rips, 2014). Handling such sentences seems to require recourse to world knowledge (cf. Winograd, 1972; Pickering & Majid, 2007; Hartshorne, Gerstenberg, & Tenenbaum, 2014), which leaves open the question of what the format of world knowledge is.

Hartshorne, Gerstenberg, & Tenenbaum (2014) model language interpretation as an inference over a generative model of the speaker (see Figure 1). Crucially, informative speakers tend to say things that are true, allowing the listener to use her expectations about the world (“world knowledge”) to fill in any gaps. Thus, listeners interpret *he* as referring to Al in (1) and *Bart* in (2) because those interpretations result in the most probable worlds. The authors formalized this world knowledge as inference over a generative model of the world (cf. Battaglia, Hamrick & Tenenbaum, 2013). Their model predicts pronoun interpretation in sentences such as (1) and (2) very well, but they do not provide evidence that other theories of world knowledge would not explain human judgements similarly well.

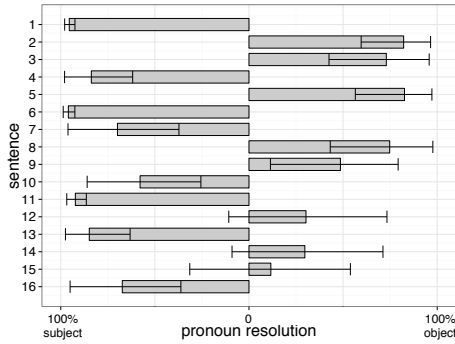
A key advantage of the generative intuitive theory approach is that the definition of world knowledge is highly flexible rather than a list of facts and heuristics. In particular, if we change how the world works we can easily update the model. For example, suppose that a tug-of-war tournament takes place in Ancient Troy and the Trojan god Adynamo is tired of seeing the strongest player always win, so he intervenes to ensure that the weakest player always wins a tug-of-war match. In this scenario, we interpret the pronoun in (1) as referring to Bart, and the pronoun in (2) as referring to Al. We tested this with participants on Amazon Mechanical Turk by creating 16 sentences like (1-2), manipulating the adjective (*strong, weak*), the verb (*beat, didn't beat, lost to, didn't lose to*), and the connective (*because, although*). Participants answered questions about the pronoun's most likely referent (see Figure 2). We find that participants indeed flexibly update their interpretation of the pronoun consistent with what the model predicts.

Alternatively, suppose that a tug-of-war tournament is held between two universities (Willowbrook College and Pinehurst University). Sentences like *Al beat Bart at tug-of-war, because he is from Willowbrook College* are difficult to interpret *unless* one happens to know that students from Willowbrook tend to be stronger than those from Pinehurst. In an AMT experiment (see Figure 3), we find the expected results, with subjects' interpretations about such sentences being strongly modulated in the expected direction once they have been told about Willowbrook's and Pinehurst's populations (“post-diagnostic”) but not before (“pre-diagnostic”). This new information does not affect interpretation of unrelated control sentences.

These studies suggest that pronoun interpretation is heavily influenced by dynamically updated beliefs about the world. We will discuss the relationship between this process and other well-established phenomena like verb biases.

Troy/Non-troy Sentences	
1	Christopher beat Nicholas at tug-of-war, because he is strong.
2	Brian didn't beat Gabriel at tug-of-war, because he is strong.
3	Andrew beat Joseph at tug-of-war, because he is weak.
4	Adam didn't beat Jack at tug-of-war, because he is weak.
5	Christian lost to Austin at tug-of-war, because he is strong.
6	Daniel didn't lose to Seth at tug-of-war, because he is strong.
7	Dylan lost to Ethan at tug-of-war, because he is weak.
8	Richard didn't lose to Carlos at tug-of-war, because he is weak.
9	Daniel beat William at tug-of-war, although he is strong.
10	Isaiah didn't beat Luis at tug-of-war, although he is strong.
11	Brandon beat Ryan at tug-of-war, although he is weak.
12	Marcus didn't beat Charles at tug-of-war, although he is weak.
13	Benjamin lost to Noah at tug-of-war, although he is strong.
14	Patrick didn't lose to Bryan at tug-of-war, although he is strong.
15	Samuel lost to Robert at tug-of-war, although he is weak.
16	Cole didn't lose to Jackson at tug-of-war, although he is weak.

Table 1: Troy stimuli



(a) Non-troy

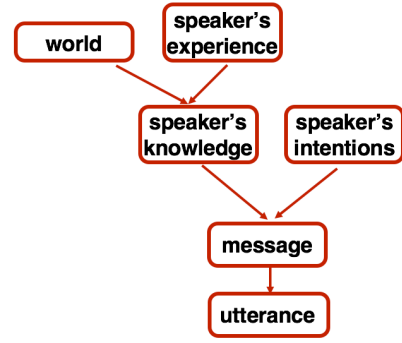
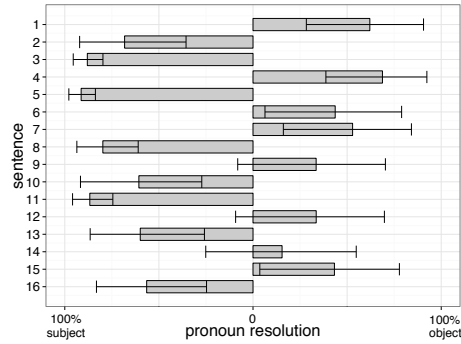


Figure 1

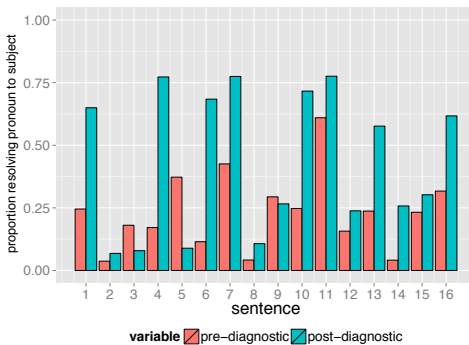


(b) Troy

Figure 2: Pronoun resolution (Troy)

School Sentences	
1	Andrew beat Joseph at tug-of-war, because he is from Willowbrook College.
2	Adam didn't beat Jack at tug-of-war, because he is from Willowbrook College.
3	Christopher beat Nicholas at tug-of-war, because he is from Pinehurst University.
4	Brian didn't beat Gabriel at tug-of-war, because he is from Pinehurst University.
5	Dylan lost to Ethan at tug-of-war, because he is from Willowbrook College.
6	Richard didn't lose to Carlos at tug-of-war, because he is from Willowbrook College.
7	Christian lost to Austin at tug-of-war, because he is from Pinehurst University.
8	Daniel didn't lose to Seth at tug-of-war, because he is from Pinehurst University.
9	Brandon beat Ryan at tug-of-war, although he is from Willowbrook College.
10	Marcus didn't beat Charles at tug-of-war, although he is from Willowbrook College.
11	Daniel beat William at tug-of-war, although he is from Pinehurst University.
12	Isaiah didn't beat Luis at tug-of-war, although he is from Pinehurst University.
13	Samuel lost to Robert at tug-of-war, although he is from Willowbrook College.
14	Cole didn't lose to Jackson at tug-of-war, although he is from Willowbrook College.
15	Benjamin lost to Noah at tug-of-war, although he is from Pinehurst University.
16	Patrick didn't lose to Bryan at tug-of-war, although he is from Pinehurst University.

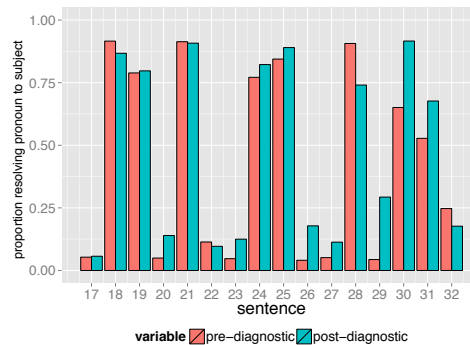
Table 2: School stimuli



School Questions

Control Sentences	
17	Christopher beat Nicholas at tug-of-war, because he is strong.
18	Brian didn't beat Gabriel at tug-of-war, because he is strong.
19	Andrew beat Joseph at tug-of-war, because he is weak.
20	Adam didn't beat Jack at tug-of-war, because he is weak.
21	Christian lost to Austin at tug-of-war, because he is strong.
22	Daniel didn't lose to Seth at tug-of-war, because he is strong.
23	Dylan lost to Ethan at tug-of-war, because he is weak.
24	Richard didn't lose to Carlos at tug-of-war, because he is weak.
25	Daniel beat William at tug-of-war, although he is strong.
26	Isaiah didn't beat Luis at tug-of-war, although he is strong.
27	Brandon beat Ryan at tug-of-war, although he is weak.
28	Marcus didn't beat Charles at tug-of-war, although he is weak.
29	Benjamin lost to Noah at tug-of-war, although he is strong.
30	Patrick didn't lose to Bryan at tug-of-war, although he is strong.
31	Samuel lost to Robert at tug-of-war, although he is weak.
32	Cole didn't lose to Jackson at tug-of-war, although he is weak.

Table 3: Control stimuli



Standard Questions

Figure 3: Pronoun resolution (college)