

**Cost and implicature in word use: Testing predictions of a game-theoretic model of alignment**  
**Hannah Rohde (University of Edinburgh), Scott Seyfarth (UC San Diego), Brady Clark (Northwestern University), Gerhard Jaeger (University of Tübingen), & Stefan Kaufmann (Northwestern University)**  
**hannah.rohde@ed.ac.uk**  
**Implicature; Game theory; Referential alignment; Interactive dialog; English**

Previous work shows that joint communication tasks yield alignment of referring expressions, highlighting the role of interlocutors' experience of shared common ground in establishing convention (Brennan & Clark, 1996; Horton, 2008; Pickering & Garrod, 2004). Less well-established, however, are predictions regarding which form~meaning mappings interlocutors will converge on. To address this, we evaluate alignment in contexts where interlocutors' common ground includes the costs of producing particular forms. Our predictions stem from game theory, a formalism for modeling players' reasoning about communication based on knowledge of the costs/rewards of particular moves and players' understanding that such knowledge is shared (Jaeger, 2008; Lewis 1969).

A game-theoretic model predicts that the use of an otherwise ambiguous form can convey meaning if an unambiguous form is costly and alternative meanings can be conveyed at low cost. In other words, a listener who knows the relative costs of unambiguously referring to X (high-cost) or Y (low-cost) may reason that a speaker using a low-cost ambiguous word X-or-Y intends to convey X, or else she would have used the low-cost word Y. For example, the word 'some' can be used literally to refer to some and possibly all entities ("some fish swim" is true even if all fish swim) or to some-but-not-all entities ("some children are girls"). The literal meaning of 'some' is therefore weaker and conveys less information than 'all'. Its meaning is strengthened from some-possibly-all to some-but-not-all through implicature—i.e., a speaker obeying the maxim of Quantity and intending to convey the more informative meaning 'all' would have used the stronger form, but since the speaker didn't say 'all', the some-but-not-all meaning is favored. A game-theoretic account of 'some' reasons that the implicated some-but-not-all meaning is conveyed given the availability of the low-cost word 'all' for the alternative stronger meaning. To test whether cost-based pragmatic inferencing applies beyond a fixed lexical host like "some", we measure alignment in a communication game with superimposed costs/rewards for production/comprehension.

Participants took turns as Sender and Receiver, naming and identifying 6 objects with 8 words: 3 tree objects and 3 flower objects; 6 unambiguous names and the ambiguous generics "tree" and "flower" (see below). On each turn, the game highlighted one object for the Sender, who then incurred a point cost for communicating a word to the Receiver. If the Receiver correctly identified the intended object, both players earned points and the roles were reversed. Ambiguous words were low-cost. The absolute value of the point costs is less important than the relative ranking of the different signals' costs. Each category contained either one high-cost unambiguous word (Study 1) or two relatively costly unambiguous words (Study 2). Games continued for 20 minutes (~60 trials) unless the pair converged on a form~meaning mapping that permitted low-cost and effective communication.

As predicted, the successful use of ambiguous words reflected the costs of the unambiguous words: Pairs that converged did so with mappings involving low-cost generic words (e.g., "tree") that referred to items with costly unambiguous names (e.g., pine tree). Across trials, high-cost items yielded more ambiguous words produced (mixed-effects model:  $p < 0.001$ ) and successfully understood ( $p < 0.001$ ). Even with more similar costs (Study 2), most pairs converged, though some settled on a convention whereby the generic referred to the second-most-costly item. To rule out a trial-and-error strategy for finding an efficient alignment without recourse to pragmatic inference, we confirmed that Receivers inferred, more often than chance, that the high-cost object was intended when Senders *first* communicated an ambiguous word ( $\chi^2 = 7.26$ ,  $p < 0.007$ ). These results contrast with work demonstrating ambiguity avoidance in the presence of referential competitors (Arnold & Griffin, 2007), by instead pinpointing how pragmatic inference about cost licenses the use of otherwise ambiguous words. Our results are in keeping with existing models of communicative efficiency (Genzel & Charniak, 2002; Levy & Jaeger, 2007) which highlight how production choices reflect the growth of speaker~listener common ground, though such models have not been applied to calculable implicatures regarding referring expressions like the ones that arise here.

**Study 1: one high-cost item per category; Study 2: two more-similarly high-cost items per category**

Objects (presented as images): apple tree, palm tree, pine tree, rose, daisy, tulip

Unambiguous words (Study 1/Study 2 point costs in parentheses, separated by '/'): "apple tree"(60/80), "palm tree"(120/135), "pine tree"(250/170), "rose"(60/80), "daisy"(120/140), "tulip"(280/165)

Ambiguous words: "tree" (80/80), "flower" (80/80)