Recent psycholinguistic studies offer a conflicting picture of how scalar implicatures (SI’s) are calculated during comprehension. While many studies find that SI’s are effortful and slow (see e.g., Huang & Snedeker, 2009), others have found that the enriched meanings of scalar terms are sometimes available as soon as the scalar trigger is spoken (Grodner et al., 2010).

To explain why this happens and what it means, I’ll take you on a whirlwind tour of language processing in the post-modular era, emphasizing three points. 1) Language processing is cascaded with interpretation at one level beginning as soon as information begins to accumulate at the level below; 2) Comprehension depends not just on the bottom-up processes triggered by a word but also by the lingering effects of previously processed words and the top-down influences of central representations; 3) Perceptual processing is also cascaded and is tightly integrated with linguistic representations. Consequently, visual input should place probabilistic top-down constraints on linguistic analysis. Evidence for this comes from the phenomenon of implicit naming (the verbal encoding of perceptual input in a nonlinguistic task) which emerges during infancy.

I will argue that these observations allow us to reconcile the conflicting findings on SI. When SI occurs through bottom-up processes, it is slow and delayed. But, when the demands of the task strongly constrain verbal encoding, SI appears to be instantaneous and effortless, simply because the relevant processes have occurred before the scalar trigger is ever spoken. Two recent experiments confirm the predictions of this proposal.