Pomona College

Self-Paced Counting

A New Experimental Technique to Study Verification Procedures for Quantified Statements

Ben Acland, Jorie Koster-Moeller, Jason Varvoutis, and Martin Hackl: Pomona College

Summary

This poster presents a novel experimental technique (“Self-paced Counting”) that allows us to gather fine grained timing information about how subjects gather information incrementally in verification tasks that involve counting. We show that this technique can detect different verification profiles for semantically equivalent quantified statements and that evidence of this sort can help distinguish between competing analyses of quantifiers that are said to be indistinguishable in their TC import and their compositional commitments.

Motivation

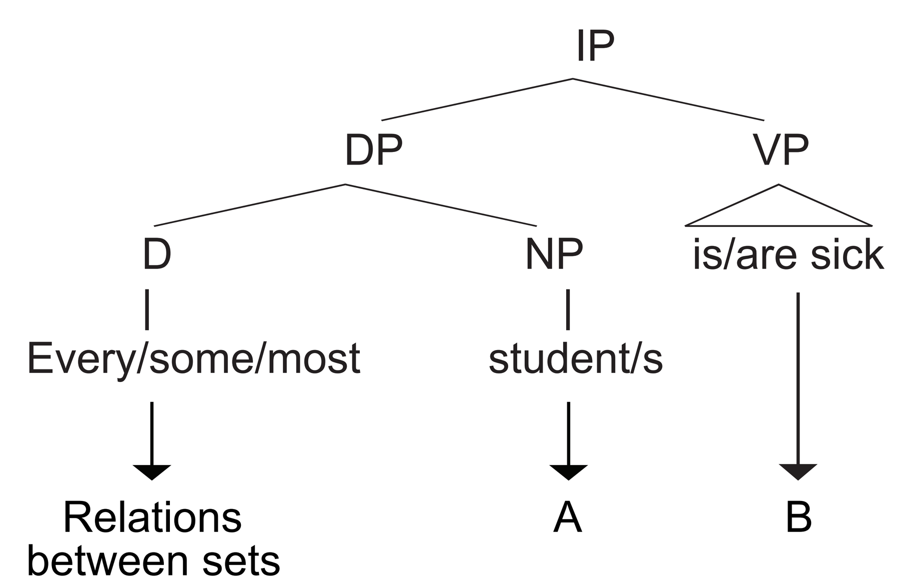
Formal semantic analyses aim to establish a systematic relation between the truth-conditional (TC) import of an expression and its syntactic/combinatorial properties. How TCs are used by other systems of the mind – for instance in verification tasks – is typically not seen as something that formal semantics needs to account for or that could help distinguish between competing semantic analyses.

An area where this lack of interest yields a particularly wide gap that more complete theories eventually will have to bridge is quantification.

Generalized Quantifier Theory

The basic semantic building blocks for quantification in natural language according to Generalized Quantifier Theory are relations between sets of individuals (Barwise&Cooper 1981).

(1)



(2) [[Every]] (A)(B) = 1 iff $A \subseteq B$

(3) [[Some]] (A)(B) = 1 iff $A \cap B \neq \emptyset$

(4) [[Most]] (A)(B) = 1 iff $|A \cap B| > |A - B|$

(5) [[More than half]] (A)(B) = 1 iff $|A \cap B| > ½|A|$

Coarseness of GQT

The internal make-up of Det. does not affect the DP-external semantics.

(6) [[fewer than one]](A)(B) =[[no]](A)(B)=1 iff $A \cap B = \emptyset$

(7) [[most]](A)(B)=[[more than half]](A)(B)=1 iff $|A \cap B|>1/2|A|$

Equivalent statements of the TCs are equally good.

(8) [[no]](A)(B) = 1 iff $A \cap B= \emptyset$, $|A \cap B|<1$, $|A \cap B|=0$, …

(9) [[most]](A)(B) = 1 iff $|A \cap B|>½|A|$, $|A \cap B|>|A-B|$, …

Verification of Quantified Statements

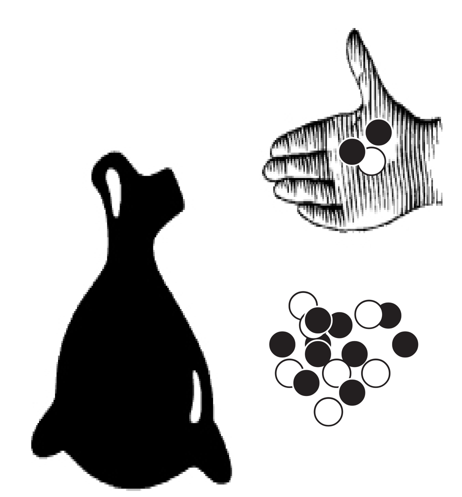
Form:	Most As are Bs.		More than half As are Bs.
TCs:	$\|A \cap B\| > \|A - B\|$	<=>	$\|A \cap B\| > ½\|A\|$
Verfication:	“Vote Counting”		“Counting to a criterion”
	• Compare # of As that are Bs to # of As that are not Bs		• Determine # of As. • Divide by 2. • Compare to # of A that are Bs.

- Can we we find evidence that most and more than half trigger different verification procedures as indicated above – even though they are seen as semantically equivalent?

Self-Paced Counting

The basic idea behind SPC

Find out whether most/more than half of the marbles in the bag are black. You reach into the bag repeatedly until you have enough information. At each handful:



Most (Vote Counting):

- Are there more black than white?
- Does black lead overall?

More than half (Criterion Counting):

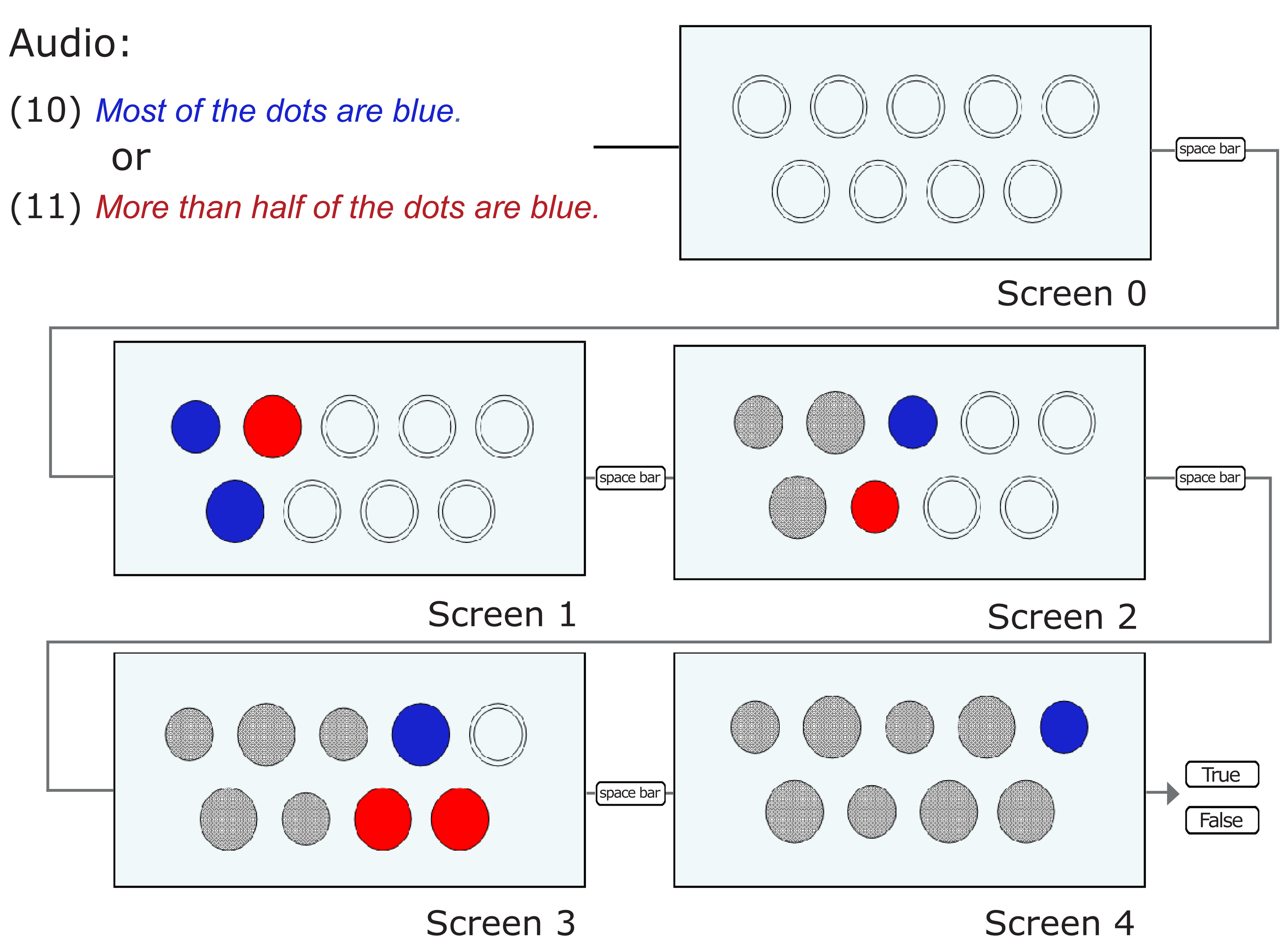
- Add # black marbles to count of black marbles.
- Add # of marbles to count of marbles.
- is # black marbles > ½ # marbles?

Audio:

(10) *Most of the dots are blue.*

or

(11) *More than half of the dots are blue.*



Methods and Materials

Self-paced Counting Methodology:

- Subjects hear a sentence and determine as fast and as reliably as possible its truth/falsity relative to an array of dots.
- The dots in the array are initially empty and ...
- incrementally filled in as subjects press the space bar.
- Previously seen dots are masked.
- Subjects can answer as soon as they have enough information.

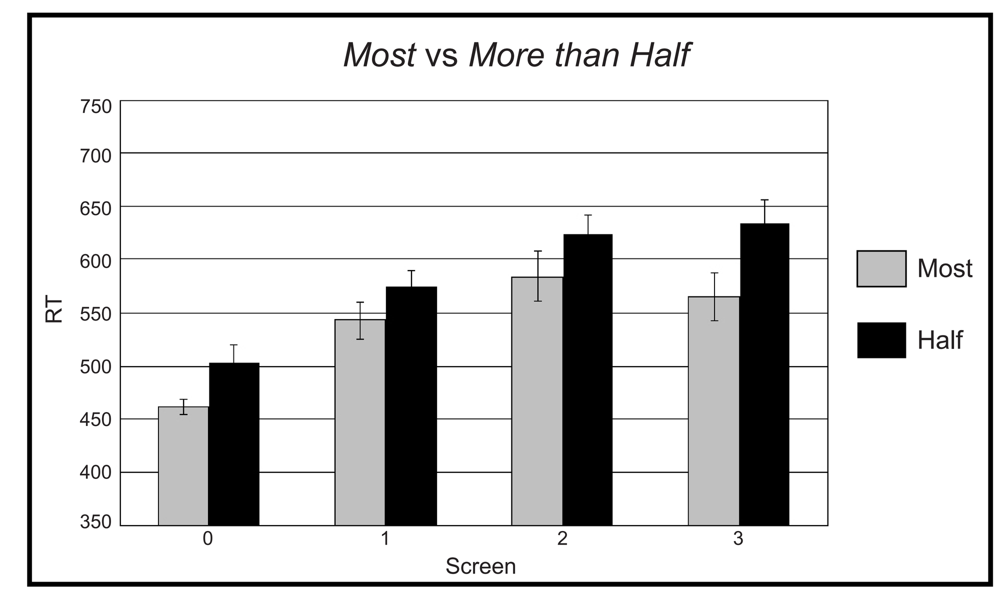
Experiment 1: *Most* / *More than half*

- 24 target items: 12 most, 12 more than half, 6 true/6 false each
- Dot arrays varied in length between 10 and 12
- Within the first 4 frames it is impossible to determine truth/falsity.
- Varied dot-size so that mass is not reliable predictor within frame 1-4.

Analysis:

- Only RTs from correct answers of subjects with ≥ 80% correct answers.
- Repeated measures ANOVA with factors “Determiner (Det)” and “Screen.”

Experiment 1

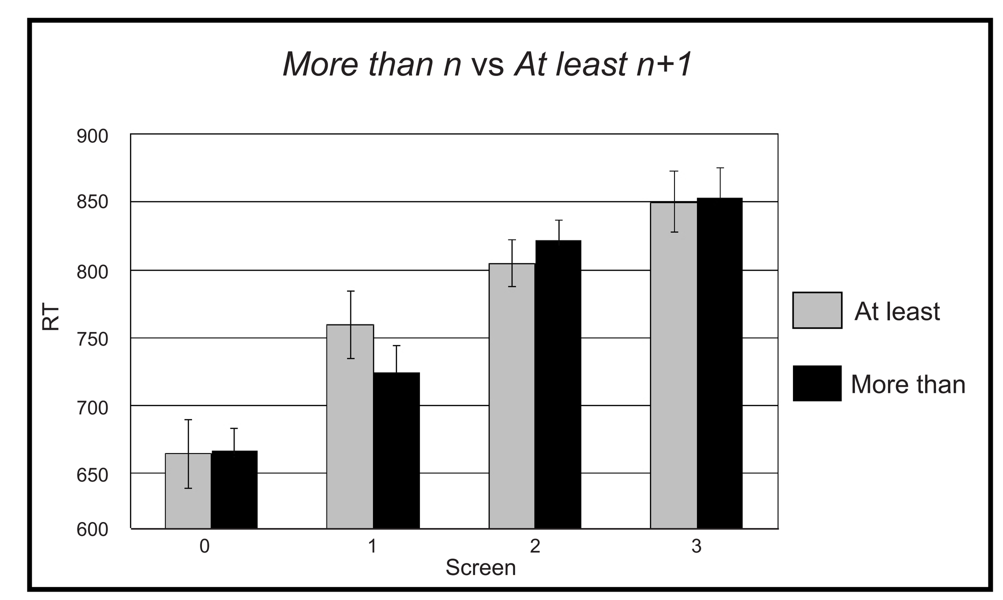


Results (n=20):

- >75% correct: no sig diff
- No sig diff in total RT
- **Main effect of Det** (p=.006)
- **Main effect of Screen** (p=0.14)

Is there a linear dependency between n and R - for counting to n in SPC?

Experiment 2



Target Items:

- 12 at least n+1
- 12 more than n: 4<n<7
- else as in Exp. 1

Results (n=12):

- >90% correct: no sig diff
- No sig diff in total RT
- **Main effect of Screen** (p=0.001)

Conclusions

- *Most*/*More than half* are treated as equivalent determiners.
- Verification strategies differ, suggesting (TC approximations):
  - *Most* triggers vote counting ($|A \cap B| > |A–B|$).
  - *More than half* triggers criterion counting ($|A \cap B| > ½ |A|$).
- GQT is too coarse to explain that. We need a theory that is based on more fine-grained semantic primitives.
- SPC faithfully tracks increases in complexity of counting tasks and offers fine-grained timing information about verification procedures involving counting.