An impossibility theorem of aggregating semantic rankings under non-monotonic quantification

Claire RONG clairerongfr@gmail.com
Department of Cognitive Studies, École Normale Supérieure, Paris

Background. In previous work on scalar items in the scope of universal quantifiers (e.g. *Each box contains books*), graded acceptability judgments were modeled using mixed-effects regression [6]. Those results showed that both categorical semantic distinctions and graded similarity to a 'best' case contributed independently to model fit. We ask whether this generalizes to sentences with scalar items under the non-monotonic quantifier *exactly* n ($n \geq 2$). Here, we treat bare plurals as a scalar item, having a ≥ 1 denotation, pragmatically strengthened to ≥ 2. Consider a sentence like *Exactly two boxes contain books*. Unlike in the monotonic case, situations may satisfy the locally strengthened reading (exactly two boxes contain ≥ 2 books and other boxes might be empty or contain one) without satisfying the literal reading (exactly two boxes contain ≥ 1 book and others are empty). Improving proximity to an intuitive best case may also move a situation closer to falsity by overshooting the exact cardinality requirement. A speaker's felicity judgment may well correspond to a cognitive unifying ranking over situations, integrating both types of rankings. The aggregation problem asks whether a speaker who implicitly entertains several plausible semantic analyses at once can have a principled way to form a single felicity judgment that reconciles the induced rankings while satisfying a small set of natural consistency constraints. We consider two competing but independently motivated theories about how felicity judgments should be organized in the non-monotonic case. We adopt a strategy inspired by Arrow-style impossibility results in social choice theory [1], which have been applied beyond preference aggregation [e.g. 10, 5, 8, 9]. In the analogy with social choice, the two semantic theories play the role of "voters", each inducing a ranking of situations (the "candidates"). Applying choice-theoretic aggregation results to questions in formal semantics is an approach that, to our knowledge, has not previously been explored.

Readings and orderings. We consider sentences of the form *Exactly* n *individuals are* P, where P contains a scalar item. Intuitively optimal situations are those where exactly n individuals satisfy the strengthened meaning of P and all others do not satisfy the literal meaning of P. Each individual may stand in one of three relations to P: **F** (falsifier), **W** (weak verifier, satisfying the literal but not strengthened meaning), or **S** (strong verifier, satisfying the strengthened meaning). A *situation* assigns one of these values to each individual in the domain and is considered up to permutation. For instance, in a domain of four individuals, situations can be represented as strings over {F,W,S} (e.g. FFSS, FWSS). Situations can be organized by minimal changes in verifier status, with distance measured by Hamming distance to an intuitively best case. While alternative metrics are possible and not all situations need be instantiated, the proofs assume a sufficiently rich situation space (see Figure 1 for an example).

We define three bivalent readings: (i) a *literal reading*, satisfied iff exactly n individuals are at least weak verifiers; (ii) an *intermediate reading*, satisfied iff exactly n individuals are at least weak verifiers and not all are weak; (iii) a *strong reading*, satisfied iff exactly n individuals are strong verifiers. The *reading-satisfaction ranking* $\succeq_s$ is defined as follows. For situations s_1, s_2, we have $s_1 \succeq_s s_2$ iff: either s_1 verifies strictly more bivalent readings than s_2; or s_1 and s_2 verify the same number of readings and, in the case where exactly one reading is verified, s_1 verifies the literal reading whenever s_2 verifies only the strong reading. This ordering reflects both the preference for satisfying more readings [3, 7] and an independently motivated salience bias toward the literal reading. The induced strict relation $\succ_s$ is defined by $s_1 \succ_s s_2$ iff $s_1 \succeq_s s_2$ and not $s_2 \succeq_s s_1$, and the induced equivalence relation $\approx_s$ by $s_1 \approx_s s_2$ iff both $s_1 \succeq_s s_2$ and $s_2 \succeq_s s_1$. We define a second ordering based on Presuppositional Exhaustification [2]. Each situation s receives a truth-value in a Strong Kleene framework: **T** if exactly n individuals are S and all others F; **F** if either less than n individuals in total are S or W, or than more than n individuals in total are S or W and among them not exactly n individuals are S; **U** otherwise. The *trivalent ranking* $\succeq_t$ is defined as follows. For situations s_1, s_2, we have $s_1 \succeq_t s_2$ iff $V(s_1) \geq V(s_2)$, where $\mathbf{T} > \mathbf{U} > \mathbf{F}$, and where **U**-situations are partially ordered by their distance to the nearest **T**-situation. The strict and equivalence relations $\succ_t$ and $\approx_t$ are induced in the standard way.

Let S be a fixed set of situations and $\mathcal{O}$ the set of total preorders over S. A *unifying function* is a map $F : \mathcal{O} \times \mathcal{O} \to \mathcal{O}$ where for any pair $(\succeq_s, \succeq_t)$, we write $F(\succeq_s, \succeq_t) := \succeq$ for the resulting preorder, with induced relations $\succ$ and $\approx$. We impose the following constraints on admissible unifying functions F: (i) Unanimity, i.e. if $\varphi \succ_s \chi$ and $\varphi \succ_t \chi$, then $\varphi \succ \chi$; (ii) Independence, i.e. the ranking of any pair $\{\varphi, \chi\}$ depends only on their pairwise rankings under the inputs; (iii) Non-dictatorship, i.e. F does not always follow either $\succeq_s$ or $\succeq_t$. The following theorem generalizes to $n \geq 2$. For $n = 1$, the intermediate and strong readings coincide, yielding a degenerate case for $\succeq_s$. As observed in the literature on non-monotonic quantifiers [4], the main theoretical difficulties arise precisely for $n \geq 2$, independently of the semantic framework. We therefore restrict our attention to $n \geq 2$.

Theorem. There exists no unifying function F from $\succeq_s$ and $\succeq_t$ to a total preorder $\succeq$ over situations that satisfies Unanimity, Independence and Non-dictatorship. **Proof sketch.** Assume F satisfies Unanimity and Independence. By Independence, the behavior of F is fully determined by its action on two-situation profiles. Fix situations φ, χ. Profiles where both orderings strictly agree are settled by Unanimity, and profiles where both rank φ and χ as equivalent impose no constraints on F, since any outcome is compatible with a dictatorship. The only relevant profiles are: $A.\, \langle \varphi \succ_s \chi,\ \chi \succ_t \varphi \rangle$; $B.\, \langle \varphi \succ_s \chi,\ \varphi \approx_t \chi \rangle$; $C.\, \langle \varphi \approx_s \chi,\ \varphi \succ_t \chi \rangle$. Each profile admits three possible outcomes: $\varphi \succ \chi$; $\chi \succ \varphi$; $\varphi \approx \chi$, yielding $3^3 = 27$ candidate mappings. We proceed by exhaustive case elimination, standard in Arrow-style arguments. The unique surviving mapping is one where for all situations φ, χ: if $\varphi \succ_s \chi$ and $\chi \succ_t \varphi$, then $\varphi \approx \chi$; if $\varphi \succ_s \chi$ and $\varphi \approx_t \chi$, then $\varphi \succ \chi$; if $\varphi \approx_s \chi$ and $\varphi \succ_t \chi$, then $\varphi \succ \chi$. To determine whether this unique candidate does define a total preorder, we then test it on full situation spaces. An exhaustive computational check for all sizes of quantificational domains $3 \leq m \leq 10$ and for all $2 \leq n < m$ verifies that the induced relation is non-transitive for the vast majority of (m, n) pairs.

Conclusion. Sentences with scalar items in the scope of non-monotonic quantifiers yield diverging predictions across theories. Empirical evidence suggests that speakers' felicity judgments may integrate aspects of two theories, which we model as two rankings. Using a novel combination of social choice-theoretic aggregation methods and computational verification, we address the meta-theoretical question of whether these two rankings can be cognitively reconciled into a single ranking underlying speakers' judgments. We find that there exists no unifying function that aggregates uniformly, for all *exactly* n with $n \geq 2$, the reading-satisfaction-based ranking and the trivalent distance-based ranking into a total preorder while satisfying Unanimity, Independence and Non-dictatorship. Thus, if speakers entertain several semantic representations simultaneously, the judgments that would result from a compromise ranking face a structural incompatibility, as they cannot satisfy all three Arrow-style constraints at once.

References. **[1]** K. J. Arrow. *Social choice and individual values*. Wiley, 1951. **[2]** I. Bassi, G. Del Pinal, and U. Sauerland. "Presuppositional exhaustification". In: *Semantics and Pragmatics* (2021). **[3]** E. Chemla and B. Spector. "Experimental evidence for embedded scalar implicatures". In: *Journal of semantics* (2011). **[4]** N. Gotzner and A. Benz. "Implicatures in (non-) monotonic environments". In: *Sinn und Bedeutung*. 2022. **[5]** M. Morreau. "It simply does not add up: Trouble with overall similarity". In: *The Journal of Philosophy* (2010). **[6]** C. Rong. "Plurals under quantification: a comparison of English and Mandarin". In: *ESSLLI Student Session* (2025). **[7]** P. Stateva, S. Andreetta, and A. Stepanov. "On the nature of the plurality inference: Ladybugs for Anne". In: *Papers dedicated to A.Reboul.* (2016). **[8]** J. Stegenga. "An impossibility theorem for amalgamating evidence". In: *Synthese* (2013). **[9]** J. Süskind. "An impossibility theorem for biodiversity". In: *Biology & Philosophy* (2026). **[10]** S. D. Zwart and M. Franssen. "An impossibility theorem for verisimilitude". In: *Synthese* (2007).

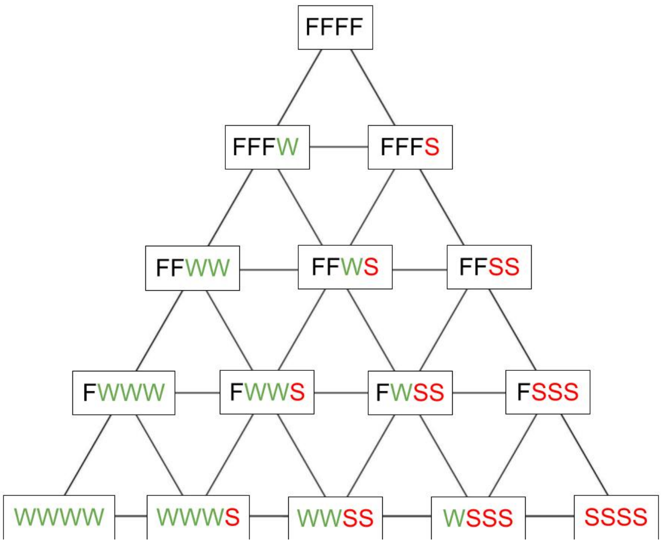

Figure 1: Situation space for four individuals, organized by verifier status and Hamming distance.