

Scaling agents via dimensions

Fabienne Martin (HU Berlin, fabienne.martin@hu-berlin.de), Sehrang Joo (Princeton, sehrang.joo@princeton.edu), Sami Yousif (UPenn, sryousif@sas.upenn.edu)
Frank Keil (Yale, frank.keil@yale.edu), Joshua Knobe (Yale, joshua.knobe@yale.edu)

1. Imagine a driver, Ana, who suddenly has a stroke. She blacks out and has an accident. Is *Ana crashed her car* an appropriate description of this scenario? Do we understand this sentence as assigning the role of agent to Ana, and if yes, does she qualify as one in this case? Previous work showed that such sentences felicitously apply to scenarios involving certain kinds of involuntary agents (van Valin & Wilkins 1996, Fausey et al. 2010), and this even in the absence of adverbials like *accidentally*. This research, however, concentrated on cases involving ‘weak’ agents satisfying more than one agentive dimension (e.g., are effective in that they ‘do’ something, and exert agent control in that they perform with normal abilities, even if not volitional). It remains unclear how in languages such English, such sentences apply to our case, where Ana is not more agentive than an *inanimate* effector. If English allows any entity in subject position as long as its referent participates in the VP-event (see e.g. Guilfoyle 2000, Wolff et al. 2009), no infelicity should arise in that case. But if one assumes that the animate subject of transitive sentences is as a rule interpreted as an agent (DeLancey 1990, see also the agent bias in sentence comprehension studies such as Bickel et al. 2015, Sauppe et al. 2022), and that in the absence of overt marking of reduced agency (e.g., *inadvertently*), the entity associated with the role ‘agent’ is typically expected to satisfy more agentive dimensions than just effectivity, such sentences should be felt problematic in scenarios where the subject’s referent is not more agentive than an inanimate effector.

2. To explore this issue, we ran an experiment where 400 adult participants were shown one of 4 short vignettes about a person, Tom, acting with most or all agentive dimensions (high agency condition) or exerting no more agency than an inanimate effector, and thus crucially not exerting agent control (low agency condition). For example, in one vignette, participants were told that Tom works as a security guard, and as such is not supposed to enter the old storage room. In the high agency condition, Tom deliberately unlocks the storage room and goes in, and thereby causes an adverse outcome (he sets off a sensor inside the room, triggering alarms). In the low agency condition, Tom unexpectedly faints, and as he falls, knocks the door to the storage room open and causes the same outcome. Participants were then asked to evaluate either a causal statement (e.g., *Tom caused the alarms*) or a non-causal statement built with one of four non-causative transitive verbs (*hit, touch, cross* and *enter*, e.g., *Tom entered the room*) on the basis of whether this sentence was a “natural/valid way of describing the event.” Results are displayed in Figure 1. We found no significant interaction between degree of agency and verb type. But there was a significant effect of degree of agency within each verb type ($ps < .001$). This means that whether Tom acted just like an inanimate effector or with more agentive dimension(s) affected participants’ evaluations of both causative and non-causative statements. More specifically, while most participants accepted the test sentences in the high agency condition, they were all over the place in the low agency one, with some accepting them, some rejecting them, and others in-between.

3. Under DeLancey’s 1990 assumption that the animate subject of transitive sentences typically is associated with the agent role, these results show us that there is a lower limit to the applicability of this role. However, the high variability of ratings in the low agency condition also suggests that this limit is of pragmatic rather than semantic nature, since a subset of participants *do* accept the test sentences in this context. This variability falls out naturally if one assumes that ‘agent’ has a weaker meaning (satisfied in the low agency condition), and a stronger one (not satisfied in the same condition). Participants are then expected to vary in their level of acceptance of a pragmatically dispreferred, weaker use of ‘agent’.

4. We propose that in English, ‘agent’ is a multidimensional concept (Kamp 1975, Sassoon 2013), with both an *evaluative* meaning requiring entities to satisfy in the VP-event as many critical agentive dimensions as required by the contextual norm (i.e., typically more than one for animate

agents, or just one – effectivity – for *inanimate* agents), and a *positive* meaning requiring to satisfy one or more agentive dimension. Since the use of the evaluative meaning is stronger, it is expected to be preferred unless there is information to the contrary, via the Strongest Meaning Hypothesis. While these two meanings are encoded by a single (covert) Voice head in English, they may be expressed by different agent-introducing control or out-of-control morphologies in other languages (Davis et al. 2009, Alonso-Ovalle & Hsieh 2021).

We first introduce a predicate *dimension* in order to make reference to a dimension of ‘agent’ (see (1)). We adopt a principle identifying critical dimensions of ‘agent’ (2)). Dimensions listed in (2) can be kept track of by separate morphosyntactic constructions across languages (see e.g. DeLancey 1984, Shibatani 2006, Ganenkov et al. 2009, Fauconnier 2012). We require any agent to be characterized by at least one dimension of ‘agent’ (see (3)).

- (1) $\lambda R.\text{dimension}(R, \lambda x \lambda e.\text{agent}(e, x))$
- (2) $\forall R(\text{dimension}(R, \lambda x \lambda e.\text{agent}(e, x)) \leftrightarrow R = \lambda x \lambda e.\text{intention}(e, x) \vee R = \lambda x \lambda e.\text{control}(e, x) \vee R = \lambda x \lambda e.\text{foreknowledge}(e, x) \vee R = \lambda x \lambda e.\text{effectivity}(e, x))$
- (3) $\forall e \forall x(\text{agent}(e, x) \leftrightarrow \exists R(\text{dimension}(R, \lambda x' \lambda e'.\text{agent}(e', x')) \wedge R(e, x)))$

To specify the number of dimensions of ‘agent’ present in a given instance, we introduce a function *cardinality* $\lambda \mathcal{R}.\text{cardinality}(\mathcal{R})$ for counting the elements of a set. Next, we introduce a function *agential*, a function from events and individuals to degrees: $\lambda x \lambda e.\text{agential}(e, x)$ (of type $\langle e, \langle s, d \rangle \rangle$). The value of this function for e and x is identical to the number of dimensions of ‘agent’ for e and x ; see (4) (cf. the counting-dimension function in Sassoon & Fadlon 2017):

- (4) $\forall e \forall x(\text{agential}(e, x) = d \leftrightarrow \text{cardinality}(\lambda R.\text{dim}.(R, \lambda x' \lambda e'.\text{agent}(e', x')) \wedge R(e, x)) = d)$

We then define a version of *agential*, *agential*⁺, that is restricted to values of at least 1 for d (see (5)), as well as a second predicate *agential*st that restricts *agential*⁺ to degrees that are at least as high as the standard degree in some context c (see (6), where $s_c(\text{agential}^+)$ denotes the standard degree in context c for ‘agential⁺’):

- (5) $\forall e \forall x(\text{agential}^+(e, x) = d \leftrightarrow \text{agential}(e, x) = d \wedge d \geq 1)$
- (6) $\forall e \forall x(\text{agential}^{\text{st}}(e, x) = d \leftrightarrow \text{agential}^+(e, x) = d \wedge d \geq s_c(\text{agential}^+))$

The last step consists in redefining *Voice*_{agent} (Kratzer 1996) introducing agent subjects as encoding a disjunction between *agential*⁺ and *agential*st (cf. (7), where v is the type of events).

- (7) $\text{Voice}_{\text{agent}} \rightsquigarrow \lambda f_{\langle v, t \rangle} \lambda x \lambda e.\text{agential}^{+/st}(e, x) = d \wedge f(e) \quad \langle \langle v, t \rangle, \langle e, \langle v, t \rangle \rangle \rangle$

In a positive context, we expect the function *agential*st to be preferred to that of *agential*⁺ in (7), for the former yields a stronger meaning. Hence, some participants reject the test sentences in the low agency condition (where Tom is not a ‘normal’ agent). But the availability of the weaker meaning explains why other participants accept them (to some variable degree) in the same context. Furthermore, our results suggest that animate agents do not count as normal agents when characterized by effectivity only, while we know from Fausey et al. 2010 that they do when not only being effective but also exerting control. This, in turn, suggests that *control* may be a key critical dimension of agency in language.

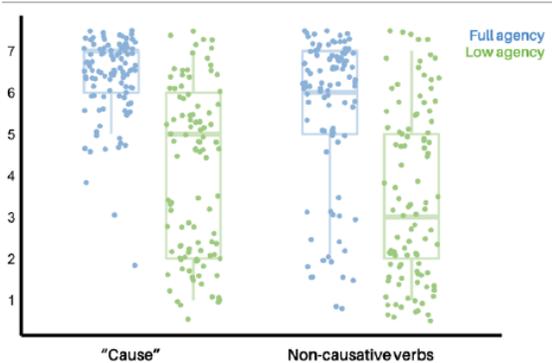


Figure 1: Results of Exp. 1