Using lexical semantics to predict the distributivity potential of Verb Phrases in a large dataset

Lelia Glass[1]

Georgia Institute of Technology

glasslelia@gmail.com

1 Introduction

Applied to a plural subject (*Alice and Bob*), some Verb Phrases (VPs) are understood distributively, meaning that they are inferred to be individually true of each member of the subject (1).

(1) Alice and Bob smiled.

  a. ✓**Distributive:** Alice and Bob *each* smiled.

  b. ✗**Nondistributive:** Alice and Bob smiled jointly without each individually doing so.

Others are understood nondistributively ('collectively'): not inferred to be individually true of each member of the subject, but rather true of the subject as a whole (2).

(2) Alice and Bob met.

  a. ✗**Distributive:** Alice and Bob *each* met.

  b. ✓**Nondistributive:** Alice and Bob met jointly without each individually doing so.

Still others can be understood in both ways (3).

(3) Alice and Bob opened the window.

  a. ✓**Distributive:** Alice and Bob *each* opened the window.

  b. ✓**Nondistributive:** Opened the window jointly without each individually doing so.

There is a long tradition of work on how these sentences should be represented semantically (Champollion, 2010, 2017; Dowty, 1987; Landman, 1989, 2000; Lasersohn, 1995; Roberts, 1987; de Vries, 2015; Winter, 2000, among others). Specifically, how should we capture the two distinct understandings available to (3) — in terms of a semantic ambiguity (of what sort?),

[1]Thanks to the participants at Linguistic Evidence, and two anonymous reviewers, for stimulating comments. This paper is based on my dissertation, Glass (2018b). I am grateful to my committee (Beth Levin, Christopher Potts, Cleo Condoravdi, and Daniel Lassiter) for guidance; to Lucas Champollion for discussion; and to Nanjiang Jiang for helping to collect data. For financial support, I thank the American Council of Learned Societies (ACLS), the Phi Beta Kappa Northern California Association, and Stanford Vice Provost for Graduate Education.

or an underspecified meaning compatible with multiple different situations? (The term 'understanding' is chosen to avoid presupposing an ambiguity, but many authors do posit one.) And if (3) is ambiguous, why are (1) and (2) unambiguous? These questions have been richly discussed (for most recent overviews, see Champollion, 2019; Syrett, to appear).

But another question has largely remained open: which Verb Phrases behave in which ways, and why? Researchers would agree that such inferences are grounded in the nature of the events described by these VPs (Dowty, 1987; Roberts, 1987; de Vries, 2015, 2017): *smiling* involves the face; people have their own faces, so they can only *smile* individually (1). *Meeting* is an inherently social action that an individual person cannot carry out unilaterally (2). Windows can be *opened* by individuals or larger parties (3). But even if the behavior of these exemplars is obvious, it remains an open question how to predict the distributivity potential of a new VP.

To make progress, this paper presents quantitative ratings from online annotators for the distributivity potential of over 2300 VPs (Glass & Jiang, 2017), built from the verbs of Levin (1993) (§2). This dataset provides evidence consistent with a series of theoretically motivated hypotheses which generalize the intuitive analysis of *smile, meet*, and *open the window* (§3):

- **Body-Mind Hypothesis**: Because individuals have their own bodies and minds, VPs describing the actions of an individual body or mind (*smile, jump, meditate, see the photo, swallow a pill*) are understood distributively.
- **Multilateral Hypothesis**: Because individuals cannot carry out inherently multilateral actions alone, VPs describing such actions (*meet*) are understood nondistributively.
- **Causative Hypothesis**: Because the nature of causation allows that multiple individuals' contributions may be jointly but not individually sufficient to cause a result, causative VPs (describing an event where the subject causes the object to change; *open the window*) allow a nondistributive understanding (in addition, perhaps, to a distributive one).

Together, these hypothesized patterns also help to explain an otherwise puzzling, striking asymmetry between transitive and intransitive verbs (§3.2; Link, 1983):

- **Transitive/Intransitive Asymmetry** (§3.2): Many intransitive verbs (*smile*) are only distributive; VPs built from many transitive verbs (*open the window*) can be understood nondistributively as well as distributively.

Because many intransitive verbs are body-mind verbs (distributive), and many transitive verbs are causative (allowing a nondistributive understanding), the Transitive/Intransitive Asymmetry is explained indirectly as a consequence of the types of events that tend to be described by these different types of VPs. More broadly (§4), this paper aims to use rigorous data to explain a much broader swath of data than usually tackled in the distributivity literature. By pinpointing the facets of events that shape the distributivity potential of the VPs describing them, the idea that distributivity 'depends on world knowledge' becomes explanatory.

2 Distributivity Ratings Dataset

To construct the Distributivity Ratings Dataset, the verbs of Levin (1993) (categorized by meaning) were placed into sentences to be rated by online annotators. Each sentence was given as its subject a conjunction of two random human names. Therefore, all verbs that do not make sense applied to humans were excluded; and verbs undergoing the causative/inchoative alternation were tested only in the causative (*broke the vase*) rather than the inchoative (*the vase broke*). Because the stimulus sentences strictly follow a 'subject-verb' or 'subject-verb-object' format (4), verbs were also excluded if they require propositions or prepositional phrases as complements (*decree that; masquerade as; put a book on the table*). To create these sentences, every transitive verb had to be given an appropriate object (4b).

(4) **Stimulus format for VPs built from intransitive and transitive verbs**

a. Name1 and Name2 verbed (*Veronika and Ian giggled*).

b. Name1 and Name2 verbed an object (*Luke and Olivia wrote a book*).

2.1 Choosing objects for transitive verbs

Of course, the object of a transitive verb – both its grammatical properties and its referent – fundamentally shapes the distributivity potential of a full VP.

Looking first at grammatical properties, it matters whether the object is singular or plural. A VP with a plural object can always be understood nondistributively ('cumulatively'; Scha, 1981): if two people *open two windows*, perhaps they do so by each opening one, adding up to two between them – nondistributive (Krifka, 1992). To avoid that possibility, the Distributivity Ratings Dataset uses only singular objects. A VP's potential for distributivity also depends on whether its object is definite or indefinite, which in turn interacts with whether the action described by the verb can be repeated on the same object: Champollion (to appear); Table 1.

Table 1: (In)definiteness and (non)repeatability interact to constrain a VP's potential for a distributive understanding.

	Definite	**Indefinite**
Repeatable on obj.	A&B opened the window.	A&B opened a window.
(open)	✓**Dist:** Each opened it.	✓**Dist:** Each opened one.
		(*same window or different ones*)
	✓**Nondist:** Opened it together.	✓**Nondist:** Jointly opened one.
Not repeatable on obj.	A&B broke the vase.	A&B broke a vase.
(break)	✗**Dist:** Each broke it.	✓**Dist:** Each broke one.
		(*different vases*: 'covariation')
	✓**Nondist:** Jointly broke it.	✓**Nondist:** Jointly broke one.

When the object is definite and the action described by the verb can be repeated on the same object (*open the window, see the photo* – where the same window can be opened, and the same photo seen, more than once), then the VP can be understood distributively (top left cell of Table 1). (As for whether the VP can be understood nondistributively, that depends on other aspects of what we know about the event; people can open windows jointly, but generally cannot see things jointly given that people have their own sensory perception; so *open the window* can be understood nondistributively, while *see the photo* generally cannot.)

When the object is definite and the action described by the verb cannot be repeated on the same object (*break the vase*, where the same vase generally cannot be broken more than once), the VP cannot be understood distributively (the bottom left cell of Table 1). It does not generally make sense for two people to each break the same vase.

When the object is indefinite and the action described by the verb can be repeated on the same object (*open a window, see a photo*), then the VP can be understood distributively (top right of Table 1). The indefinite may or may not be understood to 'covary' (Dotlačil, 2010) with each member of the subject: we might imagine two windows/photos ('distributive with covariation'), or one ('distributive without covariation').

Finally, when the object is indefinite and the action described by the verb cannot be repeated on the same object (*break a vase*), then the VP can be understood distributively (bottom right of Table 1). In that case, the object 'covaries' with each member of the subject: it would be strange for two people to each break the same vase, but they might each break a different one.

For the Distributivity Ratings Dataset, indefinite objects were used (4) to sidestep the issue of whether the action described by the verb can be repeated on the same object. Using an indefinite object allows *break a vase* to have a distributive understanding which would be unavailable with a definite object. Of course, the downside is that when the action can be repeated on the same object, we do not know whether the object is understood to 'covary' or not (if two people *open a window*, is there one window, or two?).

Alongside grammatical properties such as (in)definiteness, the distributivity potential of a VP also depends on the referent of its object. *Open an eye* is distributive, given that people have their own eyes. *Open a vault* and *open a soda* can both be understood in both ways, but presumably differ based on the difficulty of opening vaults versus sodas.

For the Distributivity Ratings Dataset verbs, it seems important to choose objects using a method that systematically controls for these issues. Particularly if the focus of the study is verbs, we do not want the choice of object to confound the data. But it is not obvious what method would control for such confounds. We certainly cannot give every verb the same object (*open a window* vs. *#eat a window*); and a generic object such as *thing* would be unnatural.

In the era of 'big data', it may seem like the answer is to simply choose the most frequent object for each verb from corpus data. But then some verbs would be given body-part objects (which are often strange as singular indefinites; *shake a head, wrinkle a nose*), introducing knowledge about bodies; some verbs would be given container or unit nouns as objects (*cook a minute, mince a tablespoon*); objects from frozen expressions (*keep an eye, abhor a vacuum*); relational nouns that sound strange out of context (*find a way*); or objects that do not make sense in the context of the Levin class within with the verb is classified (*snap a photo* when *snap* is categorized as a change-of-state verb). Corpus data is indispensable for finding naturalistically motivated objects; but it cannot be used indiscriminately.

As a compromise, my strategy was to generate for each verb a set of candidate objects from corpus data (specifically, the 30 most frequent nouns to occur within five words following that verb in the part-of-speech-tagged Spoken section of the Corpus of Contemporary American English; Davies, 2008), and then to hand-select the 'best' object from among these candidates: one which made sense as a singular indefinite count noun with no surrounding context; which fit with the Levin class in which the verb was classified; and which was not a body part (except for the transitive 'Verbs Involving the Body', such as *sprain*, which were given body part objects). I also favored objects which were concrete rather than abstract (*squash a bug* was preferred over *squash a hope*), and those which occurred most frequently. (I also avoided excessively violent sentences or those profiling certain groups of people: *persecute a minority* was preferred over *persecute a Christian/Jew*.) If none of the 30 candidate objects made sense (or if fewer than

30 were generated because the verb is infrequent), the example sentences given in the *Oxford Advanced Learners' Dictionary* were consulted; if no suitable objects were found there either, then the verb was excluded. This blend of bottom-up and top-down methods yields objects that are both naturalistically motivated and controlled for various confounds.

2.2 Study design

In Levin's data, the same verb often appears in multiple classes. *Cackle* is both an 'animal sound' verb and a 'nonverbal expression' verb. *Beat* is a 'sound' verb, a 'hit' verb, and a 'knead' verb. If the verb is intransitive, then it is tested in the same way regardless of its Levin class: it is placed into a sentence of the form *Name1 and Name2 VERBed*. But if the verb is transitive, then it may have a different object in different Levin classes: when *beat* is a 'sound' verb and a 'hit' verb, its object is *a drum*; but when it is a 'knead' verb, its object is *an egg*.

To remove duplicates, any intransitive verb (such as *cackle*) appears only once in the data, with a list of its Levin classes. Any transitive verb appears once for each distinct object with which it was tested. *Beat* is listed once with the object *a drum* (spanning both the 'sound' and 'hit' classes), and once with the object *an egg* (the 'knead' class). These data comprise 2338 unique VPs (1667 transitive verb-object combinations, 671 intransitives).

Next, online participants answered questions of the form (5), with the five answer choices recorded on a Likert scale. ('3=not sure' was the least common response for both questions.)

(5) Naomi and Jeff {smiled, opened a window, ...}.

a. Does it follow that Naomi and Jeff *each* {smiled, opened a window, ...}?
[definitely no] [maybe no] [not sure] [maybe yes] [definitely yes]

b. Could it be that Naomi and Jeff didn't technically *each* {smile, open a window, ...}, because they did so *together*?
[definitely no] [maybe no] [not sure] [maybe yes] [definitely yes]

The two questions (5a) 'each' and (5b) 'together' are worded to probe from two different angles at the same issue: whether the VP is <u>only</u> understood distributively, or whether it can <u>also</u> be understood nondistributively. (As expected for opposite ways of asking the same underlying question, the 'each' and 'together' responses are highly negatively correlated.[2])

In general, most VPs describe events that an individual could plausibly undertake individually (*smile, open a window*), and thus <u>can</u> be understood distributively when applied to a plural.[3] So it is most informative to investigate which VPs have an available nondistributive understanding <u>in addition to</u> a distributive one. Some VPs behave like *smile* in only being distributive; others behave like *open the/a window* in also allowing a nondistributive understanding. The questions (5a–5b) are designed to distinguish the *smile* type from the *open the/a window* type.

The two opposing perspectives in (5a–5b) also force participants to consider both distribu-

[2] In a mixed-effects linear regression in R (Bates et al., 2015) with 'participant' and 'verb' as random effects, every 1-point increase in a VP's (a) 'each' rating is associated with a 0.54 point *decrease* in its (b) 'together' rating ($p < 0.0001$). See §3.1 for discussion of this statistical test.

[3] The exceptions: *meet*-type VPs cannot be understood distributively, given that individuals cannot *meet* alone; and VPs with definite objects describing non-repeatable actions (*break the vase*), which cannot be understood distributively given that the same vase cannot be broken more than once.

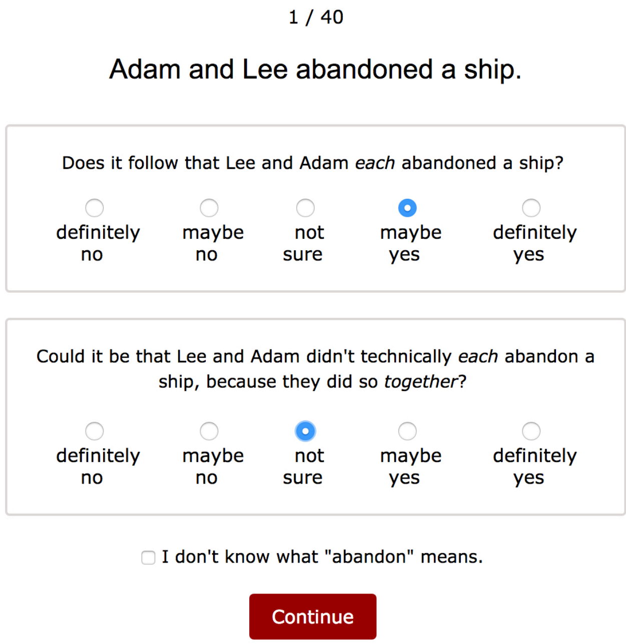

Figure 1: Screen shot from the experiment.

tive and nondistributive understandings – which is why two questions were used as opposed to just one. Otherwise, I feared that participants would be too generous in allowing that two people who merely participated in a particular VP event 'each' fully carried out that event.

Finally, one might worry about the use in (5b) of the notoriously polysemous word *together* (e.g., Moltmann, 2004). If *together* is understood in its sense of socio-spatial proximity and coordination rather than as a marker of nondistributivity, then the question (5b) would be confounded. But I believe that the surrounding context – 'didn't technically *each* VP, because they did so *together*' – helps to disambiguate. Indeed (shown below), participants' responses largely indicate that they understood it to convey nondistributivity, as intended.

2.3 Data collection

Participants using United States I.P. addresses were recruited using Amazon's Mechanical Turk. Data were only analyzed from participates who reported being native English speakers after being advised that they would be paid regardless of their answer.

Each participant encountered 40 questions of the form in (5), a randomized subset of the 2338 unique VPs tested. There were no fillers because there was no controlled manipulation to disguise. An optional checkbox allowed the participant to indicate that they were unfamiliar with the given verb (because some of the verbs, such as *pip, confabulate,* and *carom*, were quite rare); if a participant checked that box, their responses for that VP were ignored.

The goal was to collect three observations for each of the VPs in our corpus. Without a way to keep track of how many observations had been recorded for each VP, we simply hoped that with enough participants encountering 40 randomly chosen VPs each, we would eventually get three observations per VP. This methodology was perhaps not the most efficient, because some VPs were ultimately seen over ten times. We initially ran 270 participants (more than enough for each VP to be seen 3 times). But given our setup, some VPs were seen more times than needed, while others were seen fewer than 3 times. To get at least three observations per VP, we ran

58 additional participants, using only the VPs that had not gotten three ratings initially. We did not throw away any data; if a verb was seen ten times, all ten observations were included in our dataset. After excluding three non-native English speakers, and removing the 484 observations (3.7%) for which the participant indicated that they were not familiar with the verb, there are 325 participants and 12,515 responses for the questions represented in (5); with two subquestions per question, there are 25,030 datapoints in all.

Table 2 illustrates some ratings for the first three participants in these data (all available through the Open Science Framework at https://osf.io/8953e/).

Table 2: Each participant's ratings for both the 'each' question and the 'together' question for each VP they encountered.

ParticID	verb	object	'each' rating	'together' rating	Levin class
Participant1	crack	an egg	2	4	'bang'
Participant1	cackle	n/a	4	2	'snap/cackle'
(…)	(…)	(…)	(…)	(…)	(…)
Participant2	steady	a canoe	4	5	'change-of-state'
Participant2	resent	an intrusion	4	4	'admire'
(…)	(…)	(…)	(…)	(…)	(…)
Participant3	wheeze	n/a	4	2	'hiccup'
Participant3	bend	a wire	2	4	'bend'

3 Motivating and testing hypotheses

These data provide evidence about how features of a VP (and the event it describes) shape its distributivity potential. Concretely, VPs are tagged for various features, discussed in more detail below – those that are transitive vs. intransitive; those describing bodily or mental events; those describing inherently multilateral events; those that are causative. Then we examine how such features of a VP relate to its distributivity potential, as measured by responses to the 'each' and 'together' questions.

3.1 Statistical analysis

The data were analyzed using a statistical test known as a mixed-effects linear regression, using the `lme4` package of R (Bates et al., 2015). A linear regression predicts the value of a continuous dependent variable on the basis of one or more (continuous or categorical) independent variables (see B. Winter, 2013). In this context, the independent variables are the features of the VPs just mentioned. There are two separate dependent variables, each analyzed in its own statistical model. One model predicts, as its dependent variable, the response to the 'each' question; a separate model predicts, as its dependent variable, the response to the 'together' question.

These dependent variables are not strictly continuous, because participants chose among five ordered, categorical responses ('definitely no', 'maybe no', 'not sure', 'maybe yes', and 'definitely yes'). But I treat what is technically an ordered categorical variable as a linear, continuous one: assuming that the difference between 'definitely no' and 'maybe no' is equal to the difference between 'maybe no' and 'not sure', just as the difference between 1 and 2 is equal to that between 2 and 3. This way of handling Likert data is quite common and arguably justified in work on psychology and linguistics (e.g., Carifo & Perla, 2007).

As for the 'mixed effects' structure of these linear regressions: such models are used to factor out influences on the dependent variable unrelated to the hypotheses being tested. An individual participant's rating of a particular VP does not just depend on the independent variables of interest (also known as the 'fixed effects:' here, all the features of the VP noted above), but also on how the specific participant tends to use the ratings scale (some people might give systematically higher ratings than others – a 'random effect'), and also on the specific verb or verb-object combination (different VPs may act differently from one another; another 'random effect'). Mathematically, the intercept in the linear regression is allowed to vary with each participant and each VP. Such a mixed-effects structure makes use of all the available information – that the same subject rated multiple different VPs; that the same VP was rated by multiple different participants – and uses this information to help explain the variance in distributivity ratings. In this way, it is a 'conservative' model, unlikely to find a spurious effect.

All of the results reported below are drawn from two separate mixed-effects linear regressions — one for the 'each' question, one for the 'together' question — including all of the independent variables hypothesized to predict a VP's distributivity potential:

1. whether the verb is transitive or intransitive
2. whether or not the verb describes an action carried out by an individual body or mind
3. whether or not the verb describes an inherently multilateral action
4. whether or not the verb is causative
5. …and (only retained if it improved the model according to the Akaike Information Criterion) an interaction between 'transitive / intransitive' and 'body-mind'

One statistical model predicts a VP's 'each' rating as a function of all these independent variables, allowing intercepts to vary for both participants and VPs.[4] Another model predicts a VP's 'together' rating as a function of the same independent variables, again allowing intercepts to vary for participants and VPs.

By including all of these fixed effects (1–4) at once, these models allow us to isolate the effect of each independent variable, which is important because they overlap (Table 3 – although a VIF test shows that they are not too collinear). For example, most body-mind verbs are intransitive (in fact, 76% of them are: 364 of 476). A model which just used one of these independent variables or the other would conflate the effects of each one: if intransitives are found to differ from transitives, we wouldn't know if this effect is driven only by the body-mind intransitives. In contrast, a model which includes both independent variables isolates the effect of each; if each one is significant, it is predictive independent of the other. Similarly, all causatives as defined here are transitive. A model which just used one independent variable or the other

[4]There is a debate in the literature about how to use random effects: whether the model should always use the maximal number of parameters justified by the study design (Barr et al., 2013), or whether one should decide on a case-by-case basis which random effects actually contribute to the model (Bates et al., 2015). In the spirit of Barr et al. (2013), I tried to run models using the 'maximal' random effects structure (allowing random slopes for each participant depending on each fixed effect, meaning that the model would allow each participant to not just use the ratings scale differently, but also to respond differently to each fixed effect). But these models fail to converge, meaning that there is not enough data to estimate all of these different parameters. Some models converge when subsets of the maximal possible random effects are used: for example, when each participant's slope is allowed to vary depending on whether the verb is transitive or intransitive (but not depending on whether it is a body-mind verb, multilateral, or causative); in those cases, all the results reported below remain significant. Because models with the full random effects structure do not converge, I let only the intercept, not the slope, vary for each 'participant' and 'VP', using more parsimonious models in the spirit of Bates et al. (2015).

(transitive or causative) would blend these effects together: if transitives differ from intransitives, we wouldn't know if this effect is driven only by causative transitives (which in fact are 57% of all transitives; see Table 3); if causatives differ from non-causatives, we wouldn't know if this effect is driven only by the fact that all causatives are transitive. But a model including both independent variables reveals the effect of being causative above and beyond being transitive and vice versa. Furthermore, all causatives are transitive and most (76%) body-mind verbs are intransitive (Table 3), so only a model using all three of these (transitive vs. intransitive, causative vs. non-causative, and body-mind vs. non-body-mind) can disentangle these effects.

Table 3: Number of VPs in each category, and overlap between the categories. For example, 945 of the 1667 transitive verbs are causative, which amounts to 57%.

	trans	**intrans**	**body-mind**	**multi**	**causative**	**total**
trans	1667 (100%)	0	112 (7%)	0	945 (57%)	1667
intrans	0	671 (100%)	364 (54%)	91 (14%)	0	671
body-mind	112 (24%)	364 (76%)	476 (100%)	0	0	476
multi	0	91 (100%)	0	91 (100%)	0	91
causative	945 (100%)	0	0	0	945 (100%)	945

In what follows, I show that each of the independent variables in (1)–(4) significantly predicts the distributivity potential of a VP — both its 'each' rating and 'together' rating. Since these findings are drawn from a model that combines all of the predictors, we can be sure that each effect persists independently of the others.

Finally, I ran these models both with and without an interaction term. As noted (see Table 3), most body-mind verbs are intransitive, but some are transitive; so I allowed the model to make different predictions for verbs that were <u>both</u> transitive <u>and</u> body-mind verbs (*swallow a pill; see a photo*). (No other interactions were justified because no other categories cross-cut each other the way these do.[5]) An Akaike Information Criterion (AIC Akaike, 1974) comparison shows that the 'best' (lowest-AIC; most predictive and parsimonious) model for the 'each' question includes (1–4) and the interaction between transitivity and body-mind verbs (5); while the 'best' (lowest-AIC) model for the 'together' question includes only (1–4) and no interaction. The statistics reported below are taken from these 'best' models.

In the tables given below, I report the predicted 'each' or 'together' ratings, β coefficients (effect size), standard errors, degrees of freedom, t statistics, and significance levels (p) for each of the models used in the analysis. (The Appendix reviews the meaning of these terms, and provides the exact R commands used.) Table 4 characterizes for the model predicting a VP's 'each' rating as a function of (1–4) and the interaction term (5). Table 5 characterizes the model predicting a VP's 'together' rating as a function of 1–4 (but no interaction, because the AIC shows that it does not improve the model).

In sum, all the statistics reported below are drawn from the models in Table 4 and Table 5, isolating the effect of each independent variable. In what follows, I motivate each of these independent variables and its effect on distributivity.

[5] In particular, causatives do not overlap with body-mind verbs because causatives do not specify what the causer did to bring about the result (e.g., Neeleman & van de Koot, 2012), so cannot require a bodily / mental action.

Table 4: Mixed-effects linear regression for the 'each' question, with random intercepts for both participants and VPs, with an interaction between 'body-mind' and 'transitive'. Statistics below come from this model.

	predicted 'each' rating	β	**SE**	**df**	t	p
intercept (regular intrans)	4.08	4.08	0.05	1218	77.4	***
transitive	3.46 (=4.08-0.62)	-0.62	0.05	2245	-13.4	***
body-mind	4.39 (=4.08+0.31)	+0.31	0.05	2167	6.30	***
multilateral (all intrans)	3.68 (=4.08-0.40)	-0.40	0.07	2451	-5.5	***
causative (all trans)	3.28 (=4.08-0.62-0.18)	-0.18	0.03	2182	-5.9	***
trans * body-mind	4.01 (=4.08-0.62+0.31+0.24)	+0.24	0.08	2169	3.1	**

Table 5: Mixed-effects linear regression for the 'together' question, with random intercepts for both participants and VPs, but no interactions. Statistics below come from this model.

	predicted 'together' rating	β	**SE**	**df**	t	p
intercept (regular intrans)	3.23	3.23	0.05	1277	65.4	***
transitive	3.67 (=3.23+0.44)	+0.44	0.04	2234	10.6	***
body-mind	2.62 (=3.23-0.61)	-0.61	0.04	2201	-14.3	***
multilateral (all intrans)	3.53 (=3.23+0.30)	+0.30	0.08	2460	3.8	***
causative (all trans)	3.81 (=3.23+0.44+0.14)	+0.14	0.03	2206	4.4	***
body-mind * trans	(not included)					

3.2 Transitive/Intransitive Asymmetry

Decades ago, Link (1983) hinted at a relation between a VP's distributivity potential and its argument structure:

(6) **Transitive/Intransitive Asymmetry:** Most intransitives (*smile*) are only distributive, while VPs built from many transitives (*open the window*) can be understood nondistributively.

After observing that *carry the piano* (built from a transitive verb) can be understood both distributively and nondistributively, Link observes that many intransitive verbs are distributive only: 'Common nouns and intransitive verbs like *die*, however, seem to admit only atoms in their extension. I call such predicates *distributive*' (Link, 1983: 132). He reiterates (Link, 1983: 141): 'Most of the basic count nouns like *child* are taken as distributive, similarly IV [intransitive verb] phrases like *die* or *see*'.

Of course, we have already seen exceptions: *see the photo* is built from a transitive verb and is distributive; *meet* is intransitive and nondistributive; *lie* is intransitive and can be understood in both ways. Unlike the other hypotheses proposed below, the Transitive/Intransitive Asymmetry is just a hunch. If it is manifested, then we face a deeper question of why it would be so.

Indeed, according to the models described above (§3.1; Figure 2), an intransitive verb is predicted to have an 'each' rating of 4.08, while a VP built from a transitive verb is predicted to have a rating of 3.46 – a large difference, and a highly significant one ($p < 0.0001$). Turning to the model predicting the response to the 'together' question, an intransitive verb is predicted to have a rating of 3.23, while a VP built from a transitive verb is predicted to have a rating of 3.67 – again, sizable and significant ($p < 0.0001$). As hypothesized, VPs built from transitives are less distributive, and more likely to allow a nondistributive understanding, compared to intransitives.

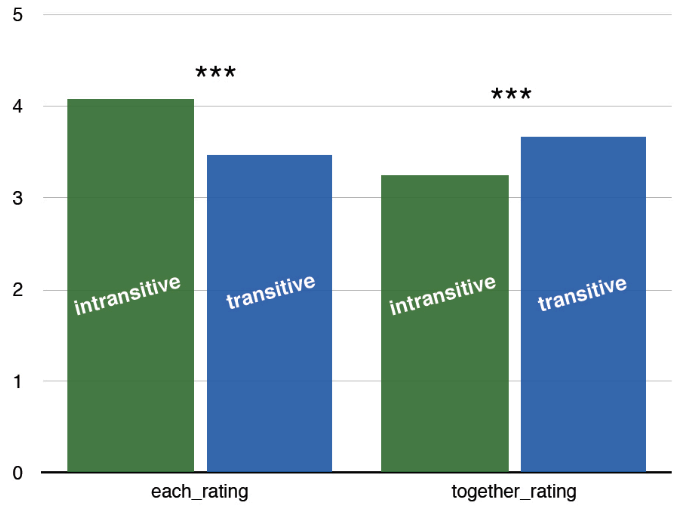

Figure 2: Predicted 'each' and 'together' ratings for transitive and intransitive verbs, from the models described in §3.1. VPs built from transitive verbs have systematically lower 'each' ratings, and systematically higher 'together' ratings, compared to intransitives.

While these findings are striking, it is much less clear how they could be explained. If the distributivity potential of a VP is grounded in the event it describes, as I claim, then why would it also be related to whether the VP involves an intransitive verb or a transitive one?

Perhaps it is because VPs built from intransitive verbs and transitive verbs describe different sorts of events. In particular, evidence converges from the acquisition literature (e.g., Naigles, 1990), the typology literature (e.g., Hopper & Thompson, 1980), and the lexical semantics literature (e.g., Dowty, 1991; Levin & Rappaport Hovav, 2005) to show that transitive verbs prototypically describe events in which an agent affects another entity in some way; while intransitive verbs describe events involving only one basic participant, which either acts autonomously or is affected by another entity that goes unmentioned. Thus, I suggest that the connection between argument structure and distributivity is an indirect one, driven by the types of events that tend to be described by transitive verbs versus intransitive ones.

The rest of the hypotheses that I lay out aim at more fine-grained aspects of VPs that shape their distributivity potential. Many of these hypothesis by their nature apply disproportionately to transitives or to intransitives, indirectly contributing to the observed asymmetry.

3.3 Body-Mind Hypothesis

Smile is distributive because it describes a facial action which a person can only carry out individually. The same reasoning should extend to other VPs describing the actions of an individual body or mind – both intransitive verbs such as *smile* and *meditate*, and transitive verbs describing bodily/mental actions such as *swallow a pill* and *see a photo*. Generalizing, we predict:[6]

(7) **Body-Mind Hypothesis:** Because individuals have their own bodies and minds, VPs describing the actions of an individual body or mind (*smile, jump, meditate, swallow a pill, see a photo, like a book*) are distributive.

[6]The Body-Mind Hypothesis faces apparent exceptions, such as *Alice's lips smiled (but her eyes didn't)*, which is understood nondistributively if the lips are imagined to create a smile jointly (Winter & Scha, 2015). But the Body-Mind Hypothesis assumes that each member of the subject has its own body or mind; so it is not surprising that the hypothesis no longer applies when that assumption is subverted.

To test the Body-Mind Hypothesis, the first step was to tag all of the VPs in the Distributivity Ratings Dataset that describe bodily or mental actions. Here and throughout the paper, verbs were tagged for such features by me, the author, using the Levin classes as a starting point. While it would admittedly be more objective to have multiple human taggers as well as a metric for inter-tagger agreement, I believe that using Levin's classification system (obviously unrelated to distributivity) already provides some independent support for these judgments. In total, 476 VPs were tagged as body-mind verbs (364 intransitive, 112 transitive; Table 3):

1. Verbs describing bodily actions:
   - Levin's 'verbs of assuming a position' (*kneel, bow, perch, slump, slouch ...*); 'verbs involving the body' (*squirm, sway, twitch, wiggle, faint, breathe, sweat, vomit, weep, kneel, curtsey, snore, swallow, hiccup, sniff, sob, sleep, wink, shrug …*); and 'verbs of grooming and bodily care' (*shower, exercise, shave ...*).
   - Levin's 'run' verbs (*canter, bounce, glide, hop, hurry, jog, run …*) and 'modes of being involving motion' (*tremble, waver, teeter, writhe …*).
   - Given that individuals have their own mouths/vocal tracts, I include Levin's 'verbs of ingesting' (*brunch, dine, graze, nosh, snack, swig, swallow …*); 'animal sound' verbs (*baa, bark, bay, bellow, bleat, cluck, coo ...*); some vocal 'performance verbs' (*sing, intone, hum*); vocal 'sound emission' verbs (*scream, screech, stutter, warble*); and contact verbs requiring specific body parts (*lick, bite, punch*).
2. Verbs of emotion and perception: Levin's 'psych' verbs with experiencer subjects (*admire, abhor, disdain, dislike, enjoy, envy ...*); and 'verbs of perception' (*recognize, glimpse, spy, spot, view ...*).

According to the models described above (§3.1; Figure 3), a regular intransitive is predicted to have an 'each' rating of 4.08, while a body-mind intransitive is predicted at 4.29 – a sizable and highly significant effect ($p < 0.0001$). The interaction between body-mind and transitivity was also significant ($p < 0.0001$); a regular transitive is predicted to have an 'each' rating of 3.46, while a body-mind transitive is predicted at 4.01, which is 0.24 points higher than if the effects of 'body-mind' and 'transitive' were kept separate. As for the 'together' model, a regular intransitive is predicted to have a 'together' rating of 3.23, while a body-mind intransitive is predicted at 2.62 ($p < 0.0001$). This time, the interaction between body-mind and transitivity was not significant; but (just based on the main effects of transitivity and body-mind) a regular transitive is predicted to have a 'together' rating of 3.67, while a body-mind transitive is predicted at 3.06 ($p < 0.0001$). In sum, body-mind VPs are <u>more</u> distributive and <u>less</u> nondistributive compared to others, consistent with the Body-Mind Hypothesis.

Moreover, the body-mind VPs are overwhelmingly intransitive: 76% (364 of 476) of them are intransitive, compared with 29% (671 of 2338) of the Distributivity Ratings Dataset overall. So the behavior of mostly-intransitive bodily and mental verbs contributes to the asymmetry between transitives and intransitives.

3.4 Multilateral Hypothesis

Like *smile*, the intuitive analysis of *meet* can be expanded. *Meet* is nondistributive because it describes a multilateral action which an individuals cannot do alone. Generalizing, we predict:[7]

[7] There are apparent exceptions to the Multilateral Hypothesis, such as Lasersohn's (1988) example, *The committees met*, which can be understood distributively (each committee meets alone). But this exception proves the rule; committees, comprising multiple members, can *meet* individually in a way that individual humans cannot.

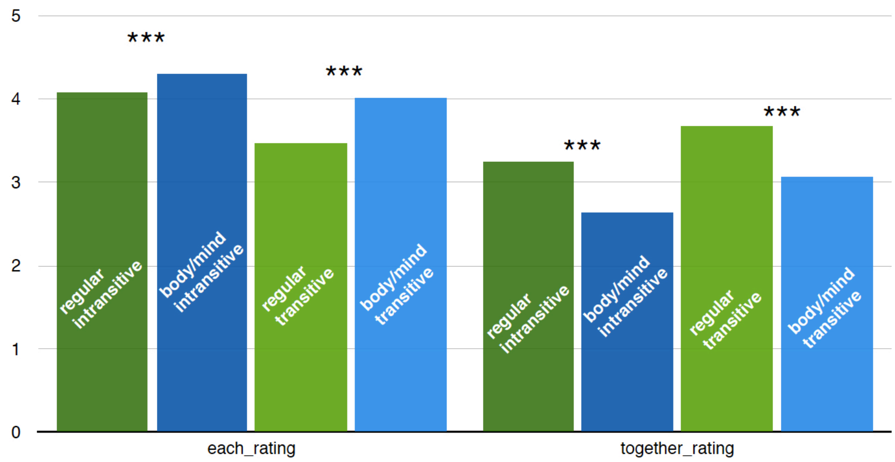

Figure 3: Predicted 'each' and 'together' ratings for body-mind vs. non-body-mind and transitive vs. intransitive verbs, from the models described in §3.1. Body-mind intransitives have systematically higher 'each' ratings, and systematically lower 'together' ratings, than other intransitives. In the same way, body-mind transitives have systematically higher 'each; ratings, and systematically lower 'together' ratings, than other transitives.

(8) **Multilateral Hypothesis:** Because individuals cannot carry out inherently multilateral actions alone, VPs describing such actions (*meet*) are understood nondistributively.

To test the Multilateral Hypothesis, the first step is to tag all the Distributivity Ratings Dataset verbs which describe inherently multilateral actions. There are some clear cases (Levin's *herd* verbs – *assemble, gather, congregate*), but also fuzzier cases. Of course, one person can only *marry* or *divorce* someone else; but if two people *marry*, they might do so together, or perhaps might each do so with some third party (e.g. *Alice and Bob married* when they each married someone else). The same goes for many other such verbs: *disagree, argue, date, elope, kiss, hug*, and so on. These verbs describe multilateral actions, but they might involve implicit third parties rather than the mutual action of the members of the plural subject.

There are also verbs describing actions for which it is difficult to say multiple participants are required or not (Winter, 2018). Does it really take two to *tango*, or can one *tango* (*waltz, foxtrot*) alone? What about *gossip, chitchat,* or *schmooze*? Given these uncertainties, it is not a simple matter to code verbs for whether they describe inherently multilateral actions or not.

To at least delineate the clearest cases, the following 91 verbs were coded as 'multilateral':

1. Levin's 'herd' verbs (*group, assemble, gather, herd, convene, congregate* ...).
2. Levin's 'meet' verbs (*meet, fight, battle, play* ...).
3. Levin's 'marry' verbs (*marry, divorce, date, court* ...).
4. Levin's 'chitchat' verbs (*chitchat, gossip, converse* ...).
5. Levin's 'correspond' verbs (*war, quibble, dispute, collaborate, compete, communicate, feud, banter* ...).

It may seem surprising that all of 91 these 'multilateral' verbs are intransitive. Several transitive verbs should arguably be considered 'multilateral' too (*share [a pizza]*, *coauthor [a book]*); but these were not tested because they unfortunately do not appear not among the Levin verbs.

There is also a large class of transitive, causative verbs describing events where the patient/object is required to have multiple parts: Levin's 'mix' verbs (*blend, combine, conjoin* ...), 'amalgamate' verbs (*interlock, interconnect*), and 'disassemble' verbs (*disconnect, unbuckle* ...). These verbs seem to describe multilateral actions on the part of their objects. However, given that I have defined distributivity here only in terms of the subject of a sentence, and given that causative verbs such as *blend* were tested only in their causative form (e.g., *blend a color* as opposed to the inchoative form, *the colors blended*), these verbs do not qualify as multilateral for the current study; a person can *blend a color* individually.

On the topic of argument structure, it is also worth noting that these 'multilateral' verbs undergo what Levin (1993) calls the 'understood reciprocal object alternation': *meet* can be transitive (*Alice met Bob*), in which case its subject can be a single individual; or intransitive, in which case its subject should be (morphologically and/or semantically) plural: *Alice and Bob met*. These syntactic realizations differ in their potential for distributivity: *Alice and Bob met Carol* is actually distributive (if two people meet someone, they each do), while *Alice and Bob met* is nondistributive. In the current study, these verbs were tested only in their intransitive form (*Alice and Bob met*).

According to the models described above (§3.1; Figure 4), a regular intransitive is predicted to have an 'each' rating of 4.08, while a multilateral intransitive is predicted at 3.68 ($p < 0.0001$). A regular intransitive is predicted to have a 'together' rating of 3.23, while a multilateral intransitive is predicted at 3.53 ($p < 0.0001$). It is surprising that the 'each' rating for such verbs is not lower, but this study's effect sizes may be dampened across the board by the fact that participants, uncertain of exactly how to interpret the questions, gravitated towards the intermediate responses '2=maybe no' and '4=maybe yes' rather than the poles.

While the relatively small effect size may be unexpected, at least the effects are sizable and significant in the predicted directions: multilateral verbs are <u>less</u> distributive and <u>more</u> nondistributive compared to others, consistent with the Multilateral Hypothesis.

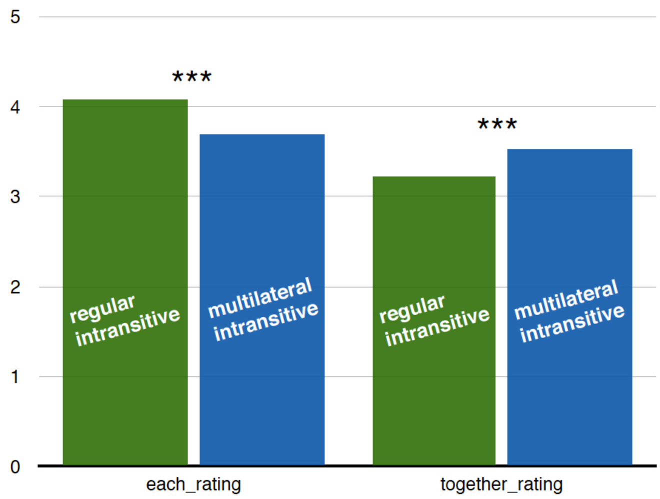

Figure 4: Predicted 'each' and 'together' ratings for multilateral verbs vs. other intransitives, from the models described in §3.1. Multilateral verbs have lower 'each' ratings and higher 'together' ratings than other intransitives.

In contrast to the Body-Mind Hypothesis, the Multilateral Hypothesis runs counter to the observed transitive/intransitive asymmetry. The 91 'multilateral' verbs (all intransitive, of 671 intransitives total), predicted to be nondistributive, are exceptions to the generalization that intransitive verbs tend to describe events that individuals carry out individually (distributive).

3.5 Causative Hypothesis

The next goal is to identify further VPs that behave like *open the/a window* in being understood in both ways. While *smile* is clearly distributive because it involves the body (§3.3), and *meet* is clearly nondistributive because it involves multiple parties (§3.4), it is much less obvious why *open the/a window* behaves the way it does, or which other VPs should pattern with it.

My proposal is that *open* is a causative verb (Dowty, 1979; Smith, 1970), describing an event in which the subject causes the object to change. By definition, causatives describe events in which the subject causes a change upon the object. I argue that this truism predicts their distributivity potential: as a general fact about causation, it is possible for multiple individuals' actions to jointly bring about a result without each individually doing so. So other causatives are predicted to behave like *open the/a window* in being able to be understood nondistributively:

(9) **Causative Hypothesis:** Because the nature of causation allows that multiple individuals' contributions may be jointly sufficient but individually insufficient to cause a result, VPs built from causatives (*open a window*) are predicted to allow a nondistributive understanding (in addition, perhaps, to a distributive understanding).

If a causative has an indefinite object, or if it has a definite object and describes an action that can be repeated on the same object (Table 1), then it can be understood distributively as well as nondistributively. With indefinite objects as used in the Distributivity Ratings Dataset, these VPs are all predicted to behave like *open a window*: understood in both ways.

To test the hypothesis that causatives and non-causatives differ in their distributivity potential, the first step was to tag verbs as 'causative' or non-'causative'. Of course, only transitive verbs can be considered causative in the sense of causing a change upon the object.

While there is no agreed-upon list of all the causative verbs, it seems clear that any verb undergoing the 'causative/inchoative alternation' (Smith, 1970, *break the vase/the vase broke*); should count as causative – encompassing for example Levin's long list of change-of-state verbs (*break, shatter, increase, boil*). Even non-alternating verbs can be considered causative if they entail that their object underwent a change of state: the 'remove' verbs (which entail that their object is removed in some way: *purge, void, confiscate*); similarly the 'put' verbs (which entail that their object ends up in a certain location: *pollute, soak, shroud*), and the 'psych' verbs describing events where the subject causes the object to feel some emotion (*annoy, frighten*). In total, 945 of the 1667 transitive verb-object combinations in the dataset were coded as causative.

According to the models described above (§3.1; Figure 5), a regular (non-causative) transitive is predicted to have an 'each' rating of 3.46, whereas a causative transitive is predicted at 3.28 ($p < 0.0001$). A regular transitive is predicted to have a 'together' rating of 3.67, whereas a causative transitive is predicted at 3.81 ($p < 0.0001$). Causatives are less distributive and more nondistributive than other transitives, consistent with the Causative Hypothesis.

With 945 (57%) of the 1667 transitive verbs in the dataset labeled as causative, this finding constitutes a far-reaching pattern. Moreover, because causatives as defined here are inherently transitive, the fact that causatives can be understood nondistributively helps to explain the observed tendency for VPs built from transitives to allow a nondistributive understanding (§3.2).

Of course, the Distributivity Ratings Dataset faces the limitation that every transitive verb is

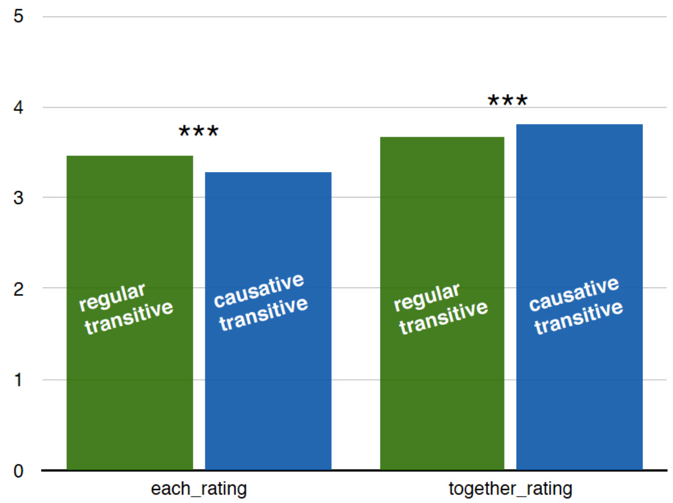

Figure 5: Predicted 'each' and 'together' ratings for causatives vs. other transitive verbs, from the models described in §3.1. Causatives (all transitive) have lower 'each' ratings and higher 'together' ratings than other transitives.

tested with a particular object, which may itself contribute to the inferences drawn about the VP's distributivity potential: perhaps *clean a house* differs from *wipe a skillet* not just because *clean* is causative and *wipe* is not, but also because houses are larger than skillets. But in the aggregate, the difference between *clean a house* and *wipe a skillet* should not matter. The procedure for choosing objects (§2.1) is not expected to give causatives and non-causatives systematically different sorts of objects in a way that would bias their distributivity potential. Moreover, by treating each VP as a random effect, the regression models that I conducted control for arbitrary differences between individual verb-object combinations. However *clean a house* differs from *wipe a skillet* in particular, the statistical analysis finds a robust difference between causatives and non-causatives in general, consistent with the Causative Hypothesis.

3.6 Discussion

There is of course more work to be done. The Body-Mind Hypothesis predicts all VPs describing the actions of individual bodies and minds to be understood distributively; but there are further non-body-mind VPs that behave that way too. In general, if two individuals are located at a particular place, then they are *each* located at that place (subparts share the location of the whole: if Bill is in Texas, then Bill's brain is in Texas; Schwarzschild, 1996, Chapter 5). Spatial verbs (*arrive, depart, enter, exit*) should therefore also be predicted to be distributive: if two people *arrive* or *enter a room*, they each do so.

Similarly, there are further VPs which behave like causatives in being understood nondistributively as well as distributively. *Rent* is not causative (renting something does not change that thing), and yet if two people *rent a car*, perhaps they each do so (distributive), or perhaps they do so jointly (nondistributive) – presumably because individuals can possess things individually or jointly (an explanation which extends to *buy, own, sell, lease*, and so on). *Eat* is not causative either,[8] but if two people *eat a pizza*, perhaps they each do so (different pizzas; 'distributive with covariation'), or perhaps they eat one pizza together (nondistributive), by each

[8]Causatives generally specify the result state of the causee/patient without specifying what the causer/agent did to effect it (e.g., Neeleman & van de Koot, 2012). In contrast, *eat* specifies an action of its agent (eating) and does not strictly entail a change in its patient: one could eat from a magical pizza which never decreases in size.

eating a different part of it. More generally, when a verb's object is construed as 'incremental' (Dowty, 1991; Krifka, 1992; Tenny, 1987) – where the parts of the object correspond to the parts of the event of affecting it – then multiple individuals might carry out the full VP event by each affecting a different portion of the object (each eating a different part of the pizza), only jointly affecting the whole, meaning that the VP can be understood nondistributively (Glass, 2018b).

Stepping back, it is hardly shocking that other body-mind VPs behave like *smile*, or that other multilateral verbs behave like *meet*. But we began with three VPs (*smile, meet, open the window*) and now systematically predict the distributivity of 1512 (476 body-mind VPs, 91 multilateral verbs, and 945 causatives), which is 64% of the 2338 total VPs: substantial progress.

4 Conclusion

This paper began from the longstanding observation that different VPs (*smile, meet,* and *open the window*) behave differently with respect to distributivity. To investigate which VPs go which ways, I used a large-scale dataset to predict the distributivity potential of 1512 Verb Phrases (see also Glass, 2018a, for an attempt to expand this type of investigation to adjectives). Other body-mind Verb Phrases act like *smile*; other multilateral verbs act like *meet*; other causatives act like *open the window*. Together, these patterns also indirectly explain why intransitive verbs tend to be distributive, while VPs built from transitives tend to allow a nondistributive understanding: many intransitives are body-mind verbs (distributive), while many transitives are causative (allowing a nondistributive understanding). More generally, the truism that a VP's distributivity potential 'depends on world knowledge about the event it describes' becomes predictive when transformed into a series of empirically tested hypotheses.

Appendix

R command for the model reported in Table 4

```
lmer(each_rating ~ trans_intrans
                 + bodymind
                 + multilateral
                 + causative
                 + bodymind * trans_intrans
                 + (1| ParticID)
                 + (1| full_pred),
                 data = d)
```

R command for the model reported in Table 5

```
lmer(together_rating ~ trans_intrans
                     + bodymind
                     + multilateral
                     + causative
                     + (1| ParticID)
                     + (1| full_pred),
                     data = d)
```

Meaning of terms used in Table 4 and Table 5

- **Predicted 'each' (or 'together') rating:** Predicted rating (along a 1-5 Likert scale treated as continuous) for the 'each' (or 'together') question for a predicate of the relevant type.

Calculated by adding or subtracting the relevant β coefficients from the intercept; for example (Table 4), a regular intransitive is predicted at 4.08; a transitive is predicted at -0.61 points less than that (=3.47).

- β **coefficient:** The number added or subtracted from the intercept to predict the 'each' or 'together' rating for the relevant type of predicate. For example (Table 4), the β score for a transitive verb is -0.61, which means we subtract that from the intercept (4.08) to get the model's predicted 'each' rating for a transitive verb.
- **Standard Error (SE):** A measurement of the accuracy of the model's predictions, defined as $\sqrt{\Sigma(Y - Y')^2/N}$, where Y is an actual predicate's 'each' (or 'together') rating, Y' is its predicted rating according to the model, and N is the number of pairs of Y, Y'. The closer the predicted values Y' are to the actual values Y, the lower the Standard Error.
- **Degrees of freedom (df):** The difference between the number of unique observations used as input into the analysis ('knowns') and the number of parameters that are uniquely estimated ('unknowns').[9]
- t **statistic:** The coefficient (β) divided by its standard error (SE). For the intercept in Table 4, the t statistic is (β=4.0845) / (SE=0.05277) = 77.39. (Note that Table 4 truncates these numbers to 4.08, 0.05, and 77.4.)
- **Significance level (**p**):** A p value is the probability of finding the observed results under the null hypothesis (that the independent variables have no effect on participants' ratings). Three stars (***) means $p < 0.0001$.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE (Institute of Electrical and Electronics Engineers) Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1007/978-1-4612-1694-0$_1$6

Barr, D. & Levy, R. & Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D. & Kliegl, R. & Vasishth, S. & Baayen, H. (2015b). Parsimonious mixed models. https://arxiv.org/abs/1506.04967

Carifo, J. & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116. https://doi.org/10.3844/jssp.2007.106.116.

Champollion, L. (2010). *Parts of a Whole: Distributivity as a Bridge Between Aspect and Measurement.* Ph.D. thesis, University of Pennsylvania.

Champollion, L. (2017). *Parts of a Whole: Distributivity as a Bridge Between Aspect and Measurement.*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198755128.001.0001

[9]From Wikipedia article titled 'Degrees of Freedom', accessed June 2018.

Champollion, L. (To appear). Distributivity, collectivity and cumulativity. In L. Matthewson, C. Meier, H. Rullmann & T. E. Zimmerman (Eds.), *Companion to Semantics*. Hoboken: Wiley.

Champollion, L. (2019). Distributivity in formal semantics. *Annual Review of Linguistics*, *5*. https://doi.org/10.1146/annurev-linguistics-011718-012528

Davies, M. (2008). *The Corpus of Contemporary American English (CoCA)*. 450 million words, 1990-present. Brigham Young University. http://corpus.byu.edu/coca/

Dotlačil, J. (2010). *Anaphora and Distributivity: A Study of Same, Different, Reciprocals and Others.* Ph.D. thesis, Utrecht University.

Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-9473-7

Dowty, D. (1987). A note on collective predicates, distributive predicates, and 'all'. In F. Marshall, A. Miller, & Z. Zheng-sheng (Eds.), *Eastern States Conference on Linguistics (ESCOL),* vol. 3 (pp. 97–116). Columbus: The Ohio State University.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. https://doi.org/10.1353/lan.1991.0021

Glass, L. & Jiang, N. (2017). *Distributivity Ratings Dataset*. Stanford University. https://doi.org/10.17605/osf.io/8953e

Glass, L. (2018a). Deriving the distributivity potential of adjectives via measurement theory. In P. Ferrell (Ed.), *Proceedings of the Linguistic Society of America*, vol. 3. Washington, D.C.: Linguistic Society of America. https://doi.org/10.3765/plsa.v3i1.4343.

Glass, L. (2018b). *Distributivity, Lexical Semantics, and World Knowledge.* Ph.D. thesis, Stanford University.

Hopper, P. & Thompson, S. (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251–299. https://doi.org/10.2307/413757

Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In I. Sag & A. Szabolsci (Eds.), *Lexical Matters* (pp. 29–54). Stanford: CSLI Publications.

Landman, F. (2000). *Events and Plurality: The Jerusalem Lectures*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-4359-2.

Landman, F. (1989). Groups, I. *Linguistics and Philosophy*, *12*(5), 27–69. https://doi.org/10.1007/bf00627774

Lasersohn, P. (1988). *A Semantics for Groups and Events.* Ph.D. thesis, The Ohio State University.

Lasersohn, P. (1995). *Plurality, Conjunction and Events*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-8581-1

Levin, B. (1993). *English Verb Classes and Alternations*. Chicago: University of Chicago Press.

Levin, B. & Rappaport Hovav, M. (2005). *Argument Realization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511610479

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, Use and Interpretation of Language* (pp. 127–146). Berlin: DeGruyter. https://doi.org/10.1515/9783110852820.302

Moltmann, F. (2004). The semantics of 'together'. *Natural Language Semantics*, *12*(4), 289–

318. https://doi.org/10.1007/s11050-004-6453-6

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374. https://doi.org/10.1017/s0305000900013817

Neeleman, A. & van de Koot, H. (2012). The linguistic expression of causation. In M. Everaert & M. Marelj & T. Siloni (Eds.), *The Theta System: Argument Structure at the Interface* (pp. 20–51). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199602513.003.0002

Roberts, C. (1987). *Modal Subordination, Anaphora, and Distributivity.* Ph.D. thesis, University of Massachusetts, Amherst.

Scha, R. (1981). Distributive, collective and cumulative quantification. In J. Groenendijk & M. Stockhof (Eds.), *Formal Methods in the Study of Language* (pp. 483–512). Mathematical Center Tracts. https://doi.org/10.1515/9783110867602.131

Schwarzschild, R. (1996). *Pluralities*. Dordrecht: Kluwer Academic Publishers.

Smith, C. S. (1970). Jespersen's 'move and change' class and causative verbs in English. In M. Jazayery & E. Polomé & W. Winter (Eds.), *Linguistic and Literary Studies in Honor of Archibald A. Hill (Volume 2: Descriptive Linguistics)* (pp. 101–109). DeGruyter. https://doi.org/10.1515/9783110800432.101

Syrett, K. (To appear). Distributivity. In C. Cummins & N. Katsos (Eds.), *Oxford Handbook of Experimental Semantics and Pragmatics* (in press). Oxford: Oxford University Press.

Tenny, C. (1987). *Grammaticalizing Aspect and Affectedness.* Ph.D. thesis, Massachusetts Institute of Technology.

de Vries, H. (2015). *Shifting Sets, Hidden Atoms: The Semantics of Distributivity, Plurality and Animacy.* Ph.D. thesis, Utrecht University.

de Vries, H. (2017). Two kinds of distributivity. *Natural Language Semantics*, *25*(2), 173–197. https://doi.org/10.1023/a:1008313715103

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. Tutorial from the University of California, Merced, 27–69. ttp://arxiv.org/pdf/1308.5499.pdf

Winter, Y. (2000). Distributivity and dependency. *Natural Language Semantics*, *8*(1), 27–69. https://doi.org/10.1007/s11050-017-9133-z

Winter, Y. & Scha, R. (2015). Plurals. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (pp. 77–113). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118882139.ch3

Winter, Y. (2018). Symmetric predicates and the semantics of reciprocal alternations. *Semantics and Pragmatics*, *11*(1), https://doi.org/10.3765/sp.11.1