Quantifying events and activities

Haley Farkas, Alexis Wellwood

1 Introduction

Expressions like *more* have been the target of early and sustained interest in formal semantics, from their occurrence as part of complex determiners (e.g., *more than three*; Barwise & Cooper 1981, Geurts & Nouwen 2007, etc.), as adjectival modifiers (e.g., *more intelligent*; Seuren 1973, Cresswell 1976, von Stechow 1984), and, more recently, as nominal and verbal modifiers (e.g., *more coffee*, *run more*). As the relevant empirical terrain has expanded, so are new questions raised about the relationship between quantification, broadly construed, and degree comparison. At the same time, research in cognitive psychology has revealed deep correspondences between comparative language and conceptualization, bringing to the fore certain foundational questions about how formal semantic analysis relates to language understanding. We examine the relationship between event structure (as encoded by verbs like *jump* and *move*) and conceptualization in the resolution of degree selection in comparatives.

The study of nominal and verbal comparatives has highlighted the general notion of 'measurement' in characterizing their meaning, where measurement is understood as a mapping μ_δ from an ordered set of entities E to degrees on a scale S_δ, where S_δ represents quantitative relationships along dimension δ that hold amongst the elements of E. In the case of *more coffee*, E is a set of portions of coffee ordered by inclusion, and it is a set of similarly-ordered stretches of running activity for *run more*. Which dimension for comparison δ is selected in

any given case depends on whether μ_δ preserves strict ordering relationships on E (see e.g. Schwarzschild, 2002, 2006; Wellwood et al., 2012). Thus, part of the meaning of *more* in its nominal and verbal occurrences is a variable μ ranging over measure functions,[1] whereas adjectival *taller* and adverbial *faster*/*more quickly* lexically specify particular measure functions. Importantly, then, the prevailing theory of comparatives with bare *more* targeting N or V is that the specific interpretation of *more*—which dimension it involves—depends on the ontological properties of N and V.

Such a theory could only ultimately be tested, though, given an independent grasp on the relevant ontology. To see this, consider a novel verbal comparative ϕ based on *more* V, for novel V. We should like to say not only (i) under what conditions ϕ should be judged true, but (ii) whether, for any given state of affairs s, speakers will in fact say that ϕ is true in s. (ii) is a challenge for our semantic theory precisely whenever we multiply entities but fail to specify when or whether s in fact provides those entities. For example, it is common enough to assume that a given object o and its constituent matter m are distinct in our semantic domain D (cf. Parsons 1990; Link 1983), and to leverage such a distinction to help explain the intuitive asymmetry between *more matter* and ?*more object* (cf. Wellwood 2018). More often than not, though, semanticists contend that such distinctions in D merely reflect what competent speakers of the language 'talk *as if*' there is (see Bach 1986; Pelletier 2011; Bach & Chao 2012 for explicit defense of this position; cf. Moltmann 2017), disregarding what our best physical, metaphysical, or cognitive theories might say about the relation between o and m. In the extreme, we're free to posit entities (or representations of entities) with properties that no plausible independent theory would endorse. In contrast, only rarely will a semantic analysis be taken seriously if it fails to conform to the structural expectations of our best syntactic theory.

[1] A major question raised by this view is whether it is right to assume that *more* as it occurs in *more than three books* or *more books* can be assumed to directly encode a cardinality function; see Wellwood (2018) for relevant discussion.

This chapter aims to correct, in small part, the theoretical retreat to mere 'talk *as if*.' We outline this perspective in somewhat more detail in the next section, along the way motivating our series of four experimental studies at the interface between language and vision.

2 Comparatives in language and mind

Semanticists often posit that the domain of entities to which we can refer or quantify over, D, distinguishes 'events' from 'activities,' analogously to its distinction between 'objects' and 'substances.'[2] Events, like objects, are importantly 'atomic', or indivisible for the purposes of reference and quantification; activities and substances are non-atomic, or divisible. Evidence for these distinctions is primarily drawn from the distributional profiles of particular Ns and Vs in concert with an intuitive characterization of the semantic field N or V invokes. In turn, these domain differences are put to semantic work, as in the noted analyses of nominal and verbal comparatives. Testing such theories, we contend, requires an explicit link between such theories and adjacent areas of cognitive psychology. Or at least, these are the points that this section aims to establish.

2.1 Ontology in semantic explanation

Semanticists often differentiate classes of NPs and VPs based on the kinds of relationships that hold (or fail to) between entities in their extensions.[3] Relevant data is typically drawn from (i) asymmetries in the intuitive naturalness/interpretability of NP/VP across a variety of grammatical environments, and (ii) the types of inferences (cumulativity, divisiveness,

[2]Depending on one's typology, 'activities' might be considered a subset of 'events'. Building on recent prior work testing the analogy *object : substance :: event : process* (Wellwood et al. 2018a,b), we understand the terms 'event' and 'activity'/'process' to be mutually exclusive.

[3]Our discussion is limited, for present purposes, to phrases that intuitively apply to concrete or 'basic level' entities like toys, mud, jumping, and moving.

etc.) that they support.

For example, it has long been observed that concrete nouns differ in whether they comfortably appear in grammatical contexts that impose differing demands with respect to 'countability'. The nouns properly called 'count' are perfectly comfortable in the singular and plural form, (1a-i); they straightforwardly support distributive quantification, (1a-ii), and counting language, (1a-iii); and, they surface naturally with *many* as opposed to *much*, (1b). The nouns properly called 'mass' have the opposite distribution, (2).

(1) a. i. Ann bought a toy/some toys.

ii. Each toy that Ann bought was shiny.

iii. She bought three toys.

b. i. Sue didn't buy many toys.

ii. ? Sue didn't buy much toy.

(2) a. i. ? Ann bought a mud/some muds.

ii. ? Each mud that Ann bought was blue.

iii. ? She bought three mud(s).

b. i. ? Sue didn't buy many muds.

ii. Sue didn't buy much mud.

Such nouns are also distinguished by the types of inference that they support. For example, the mass noun *mud* supports cumulativity inferences as in (3a), while the count noun *toy* does not, (3b).

(3) a. If this_1 is mud, and that_2 is mud, then this_{1+2} is mud.

b. ? If this_1 is a toy, and that_2 is a toy, then this_{1+2} is a toy.

One way of explaining these patterns goes as follows (see Gillon 2012 for a more detailed overview and discussion). The extension of a count noun like *toy* is a set of entities, no

proper subparts of which are in that same extension—i.e., a set of *atoms*.[4] In contrast, the extension of a mass noun like *mud* is a set of entities, any proper subpart of which *is* in that same extension. This difference in 'atomicity' can be understood in a couple of different ways; minimally, though, it has to do with whether the concept expressed by the noun supports non-arbitrary counts (see especially Koslicki 1997). That is, an array of four toys contains four toys no matter how they are arranged. But an array of portions of mud may, under rearrangement, end up being 8, or 2, or 17 portions. Singular and plural morphology, distributive quantifiers, and *many* impose a requirement for atomicity, which is plainly met by count nouns but not mass nouns.[5]

Similar observations and explanations have been made in relation to verbs, as well, though matters are more delicate here.[6] Certainly, it is straightforward to talk about single or multiple jumps, (4a-i), to pair jumps with commands to do so, (4a-ii), to count some jumps, (4a-iii), and to say that they weren't many in number, (4b-i). The pattern is different when we talk about what is happening *as* movement rather than *as* jumping, (5), though any jump is a movement of a particular sort. Movement *per se* can only be counted non-arbitrarily via, e.g., maximal episodes of continuous movement, or transitions from not moving to moving.[7]

(4) a. i. Ann jumped once/again and again.

ii. Ann jumped whenever Sue told her to.

iii. She jumped three times.

b. i. Sue didn't jump many times.

[4]Whether this is understood as a lexically-determined extension or that of *toy*+SG, for some zero singular morpheme, depends on the theory.

[5]At least, not by the sort of mass nouns that are under discussion. Superordinate mass nouns like *furniture* have atomic minimal parts in their extensions, e.g. the individual chairs, tables, etc. We do not consider these cases here, but we also do not suggest the view that mass syntax implies *anti*-atomicity; see Bale & Barner 2009, Gillon 2012.

[6]Testing the verb *qua* verb is challenging; in the text, we illustrate the hypothetically parallel patterns using semelfactive *jump* versus activity *move*, as these provide the clearest possible contrasts in English.

[7]It is important for our purposes that *move* in sentences like (5) not be read as in 'move house' or 'make a move', which we take to be related but independent from the sense of interest.

ii. ? Sue didn't jump much. [?distance]

(5) a. i. ? Ann moved once/again and again.

ii. ? Ann moved whenever Sue told her to.

iii. ? She moved three times.

b. i. ? Sue didn't move many times.

ii. Sue didn't move much. [✓distance]

Extended consideration of such data motivates the idea that the mass/count distinction in the nominal domain and the atelic/telic distinction in the verbal domain are semantically parallel. And indeed, atelic or 'unbounded' *move* supports cumulativity inferences, (6a), while telic or 'bounded' *jump* does not, (6b).[8]

(6) a. If Ann moved for 30 seconds, and then moved for 30 seconds, then she moved for a minute.

b. ? If Ann jumped in 30 seconds, and then jumped again in 30 seconds, then she jumped in a minute.

To be rendered compatible with plural morphology or pluractional phrases, it must simply be possible to find some way of bundling the stuff that mass nouns like *mud* apply to and the activity that verbs like *move* apply to, such that they support non-arbitrary counts in the context of evaluation. This property of a given N or V in particular—whether it directly supports non-arbitrary counts—has, in other literatures, been attributed to the kinds of concepts named by N or V: object and event concepts support such counts, regardless of the

[8]An anonymous reviewer suggests that (6) improperly tests the telicity profiles of modified VPs, rather than those of the embedded Vs as we have suggested. However, cumulativity inferences are always made on the basis of sentences which, whether we like it or not, contain functional material with the potential to mask the lexical implications of a given V. Such inferences are informative, then, only to the extent that we can carefully control the grammatical context so that it supports just what we think the verb semantics supports. Careful comparison of (6) with (3) should reveal that the prepositional phrases in (6) are doing no more or less than the unobjectionable indefinite singular or plural morphology in (3).

context, while substance and process concepts do not (see Rips & Hespos 2015; Wellwood, Hespos & Rips 2018b for recent discussion, and references). Whether located in (mental) concept or (physical) extension, the relevant asymmetries are attributed to whether the N or V has atoms in its extension, as required by plural morphology (overt *-s* with nouns, covert PL with verbs; see e.g. Ferreira 2005) among other grammatical devices.

When plural, of course, both count nouns and telic VPs support cumulativity inferences, (7). This reflects a shift to talk of pluralities, instead of their atomic minimal parts.

(7) a. If these$_1$ are toys, and those$_2$ are toys, then these $_{1+2}$ are toys.

b. If Ann jumped for 30 seconds, and then jumped again for 30 seconds, then she jumped for a minute.

2.2 The semantics of comparatives

These ontological distinctions matter for nominal and verbal comparatives, but not for adjectival and adverbial comparatives—at least, not so far as we have seen in the literature. We now quickly sketch the formal details of the semantics for comparatives that we will assume, drawing out where and how ontology matters. The basic set-up is that gradable adjectives and adverbs (GAs) lexically specify particular measure functions for their comparative forms (in the tradition following Seuren 1973 and Cresswell 1976), while measure functions are selected by the comparative morphology in an ontology-sensitive way with NPs and VPs.

Standard assumptions about the semantics of adjectival comparatives hold that GAs lexically introduce specific measure functions. On one popular way of formalizing this, the adjective directly expresses a measure function that is taken as an argument by a comparative operator; in this case, a sentence like (8a) would be interpreted as in (8b) (e.g., Cresswell 1976; Kennedy 1999)—true just in case the temperature of the coffee is greater than δ, the

degree contributed by the *than*-clause.[9,10] A simple extension of this approach to adverbial comparatives like (9a) would look as in (9b) (see Wellwood 2019, ch.2)—true just in case the speed of Ann's running is greater than δ'.

(8) a. The coffee is hotter than the soup is.

b. $\mathbf{hot}(c) > \delta$

(9) a. Ann ran faster than Betty did.

b. $\exists e(\mathbf{ag}(e, a)\ \&\ \mathbf{run}(e)\ \&\ \underline{\mathbf{fast}(e) > \delta'})$

Generally, GA-specified measure functions are not sensitive to whether their inputs have any interesting mereological structure.[11] Nominal and verbal comparatives are different. Here, the selection of measure functions varies, both within and across predicates, but the available measure functions are restricted to those that preserve ordering relations on their inputs. As described by Wellwood et al. (2012), building on important observations by Schwarzschild (2002, 2006) and Nakanishi (2007), comparatives targeting mass NPs and atelic VPs well demonstrate these properties: (10a) can be interpreted as a comparison by weight or volume, but not by temperature, while (11a) can involve distance or duration, but not speed. Subsequent works thus posit that part of the interpretation of *more* is a variable, call it μ, valued by the assignment function σ. Permissible values of $\sigma(\mu)$, applied to argument $\alpha \in D_P$, must be monotonic with respect to $\preccurlyeq_P$. Informally, this just means that permissible measure functions preserve strict ordering relations.

[9]Standard compositional assumptions would, in present terms, unpack δ in (8b) as $\mathbf{max}(\lambda d.\mathbf{hot}(s) \geq d)$, and δ' in (9b) as $\mathbf{max}(\lambda d.\exists e(\mathbf{ag}(e, b)\ \&\ \mathbf{run}(e)\ \&\ \mathbf{fast}(e) \geq d))$. See Bhatt & Pancheva 2004 for recent discussion of the syntax of the *than*-clause.

[10]A prominent alternative expands on the style of interpretation in (8b) to accommodate, in particular, scope-related phenomena; see Bartsch & Vennemann 1972; Heim 2000. On such a formulation, the interpretation of (8a) would look like $max(\lambda d.\mathbf{hot}(c) \geq d) > max(\lambda d.\mathbf{hot}(s) \geq d)$, which is truth-conditionally equivalent to (8b).

[11]This is not to say that no theory of GAs or GA comparatives is sensitive to ordering relations in the mapping to degrees. However, such theories are limited to considerations of 'base orderings' between individuals as introduced by the GA, and some homomorphic relationship between the base ordering and scalar structure; see for example Bale and Schwarzschild, this volume.

(10) a. Ann bought more coffee than Betty did.

b. $\exists e(\mathbf{ag}(e,a)\ \&\ \mathbf{buy}(e)\ \&\ \exists x(\mathbf{th}(e,x)\ \&\ \mathbf{coffee}(x)\ \&\ \underline{\sigma(\mu)(x) > \delta''}))$

(11) a. Ann ran more than Betty did.

b. $\exists e(\mathbf{ag}(e,a)\ \&\ \mathbf{run}(e)\ \&\ \underline{\sigma(\mu)(e) > \delta'''})$

Relatedly, the measurand must be drawn from a domain that has such structure, non-trivially. This claim explains the oddity of count NPs (?*more toy*) and perfective telic VPs (?*die (that time) more*) in the comparative (Wellwood, Hacquard & Pancheva 2012) in terms of the independently-motivated assumption that the extensions of such predicates are simply unordered sets of atomic entities. These 'flat' structures are presupposed by plural morphology, however, and any comparative targeting a plural XP (e.g., *toy-s*, or *jump*-PL) is fine and interpreted as a comparison by number. According to Bale & Barner (2009) and Wellwood (2018), this restriction is due to the fact that plural syntax introduces formally distinct structures—a set of pluralities ordered by a plural or individual part-of relation—against which the selection of $\sigma(\mu)$ can be determined. The shift in dimensionality, for example, from *more coffee* to *more coffees* can, in turn, be explained in terms of a shift in 'what is measured'. Table 1 summarizes one approach to this (cf. see Wellwood 2018).[12]

expression	**semantics**	**dimension**
toy	$\lambda x : \mathbf{at}(x)\ .\ \mathbf{toy}(x)$	-
toy-s	$\lambda X\ .\ [\forall x : X(x)\ \&\ \mathbf{at}(x)]\ \mathbf{toy}(x)$	number
coffee	$\lambda y\ .\ \mathbf{coffee}(y)$	volume, weight
coffee-s	$\lambda X\ .\ [\forall x : X(x)\ \&\ \mathbf{at}(x)]\ \exists y(x \lhd_m y\ \&\ \mathbf{coffee}(y))$	number
jump	$\lambda e : \mathbf{at}(e)\ .\ \mathbf{jump}(e)$	-
jump-PL	$\lambda E\ .\ [\forall e : E(e)\ \&\ \mathbf{at}(e)]\ \mathbf{jump}(e)$	number
move	$\lambda e'\ .\ \mathbf{move}(e')$	distance, duration
move-PL	$\lambda E\ .\ [\forall e : E(e)\ \&\ \mathbf{at}(e)]\ \exists e'(e \lhd_t e'\ \&\ \mathbf{move}(e'))$	number

Table 1: Hypothesized links between phrasal form, meaning, and dimensions for comparison. Bare occurrences of count nouns like *toy* and eventive verbs like *jump* are semantically restricted to atomic elements (**at**), which themselves are the minimal parts of pluralities.

[12]We use the colon in terms like $\lambda\alpha : \mathbf{at}(\alpha)$ and $\forall\alpha : (\alpha)$ to indicate domain restriction.

2.3 Theory evaluation

The theory of comparatives just sketched captures some interesting and important facts about semantic competence. Speakers allow the dimension for comparison with both nominal and verbal *more* to vary, but not without limit: both the observed variability and constraints are correlated with, if not explained by, abstract referential properties of what is targeted for measurement and comparison. Ultimately, of course, the explanatory power of a theory like this will depend on the extent to which it can predict new observations. In the present case, the theory will be predictive only once we have an independent theory of 'the domain' D—that is, some independent way of verifying whether we are right about the properties we've hypothesized for different sorts of entities in D. Generally, two options are considered: either our domain is identical with the world as described by physical or metaphysical theory, or with the way we represent the world, as described by cognitive scientists. In the former case, semantics interacts with metaphysics; in the latter, it interacts with systems of perception and reasoning.

That is, assume that predicate P introduces domain, D_P, with such and such properties. The theory will then say what the semantic significance of combining bare *more* with P should be, in light of the properties of D_P. This is alright, as far as it goes. A semantic theory describes a certain relation between linguistic objects and non-linguistic objects, and so it seems appropriate that explanations in semantics should sometimes depend, in part, on properties of those relata. On the morphosyntax side, theories of morphology and syntax provide independent checks on the theory; but what about on the non-linguistic side? Put differently, how can we determine the properties of D_P? So far, the independent evidence that we've considered for those properties comes from other areas of semantic analysis (e.g., the mass/count and telicity literatures).

Armed only with linguistic evidence for the properties of the relevant non-linguistic objects, the theory cannot predict, for example, how the combination of *more* and P will be

interpreted, for novel P. In such a case, there is no other linguistic data to check for the referential properties of P. We think that a semantic theory should be able to predict dimensional choices in such cases, since this is precisely the situation that young acquirers of English are plausibly regularly faced with, and, as Chomsky (1965) famously reminded us, an explanatory linguistic theory must be able to explain language acquisition.

While in many cases it may be appropriate to think about the domain in terms of what we 'talk *as if*' there is (e.g., Bach, 1986), to be predictive we must go beyond this. This is what we would like to do in the remainder of this chapter: to think explicitly about how subtle features of linguistic structure might align with representations and operations in non-linguistic cognition. While features of metaphysical reality might be relevant in explaining semantic competence at some point, we directly face the question of how linguistic and non-linguistic cognition are connected: the primary data of formal semantics—judgments of sentential truth and falsity in context—is output by minds, after all (cf. Pietroski, 2010).

In approaching this question from the perspective of cognitive science, there is a huge body of work in psycholinguistics, language acquisition, vision science, and conceptual development linking mass/count language (at least as applied to the concrete, or basic level categories, as we have assumed) to the conceptual distinction between object and substance (see Rips & Hespos 2015 for a broad overview). Recently, Wellwood, Hespos & Rips (2018b) have extended this type of research into the atelic/telic distinction, characterizing the event/process distinction as conceptual in nature, and parallel to the distinction between object/substance. In particular, they discovered that the same feature which cues people towards 'atomic' concepts in the static domain—namely, non-arbitrariness of form, in this case spatial—also cues people towards 'atomic' concepts in the dynamic domain, where (non-)arbitrariness is primarily temporally determined.

The right kind of test that *more* is sensitive to ontological features, and that these features are at least worked out independently of language, has been conducted for nominal

mass/count and conceptual object/substance by Barner & Snedeker (2004). These authors presented their research participants with novel stuff, the features of which were manipulated between those independently thought to influence categorization in terms of object versus substance concepts (amongst which are shape complexity, regularity, or repetition, etc.; again, see Rips & Hespos 2015). Next, they varied the sizes of the stuff/things and apportioned them in varying numbers, such that agent A had more if it by number and agent B had more of it by size. Now they could test whether ontological category influences dimensional choices, and they found that it did: participants strongly preferred to compare *more* N by number when given salient 'object' cues for N, whereas they strongly preferred area given 'substance' cues.

Such a test provides compelling evidence for the ontology-sensitivity of nominal *more* (as well as for the cognitive interpretation of that theory[13]). An appropriate test of verbal *more*, then, will have the same structure. Here, though, the independent evidence for a conceptual event/activity distinction is thinner on the ground, but this may be only because it hasn't yet been sought out in earnest. Here is how we think it best to proceed to the relevant test in a series of steps, of which the present contribution is only the first.

A prior study uses known verbs, and tests the semantic theory's claims about the interpretation of verbal and adverbial comparatives. Ideally, this test uses 'ambiguous' displays which can be described in various ways, and in which it is possible to compare along multiple competing dimensions (cf. Odic, Pietroski, Hunter, Halberda & Lidz, 2018). Here, different linguistic forms are evaluated against identical displays, making it possible in principle to attribute observed differences in evaluation to properties of the linguistic objects themselves. The second test constructs 'unambiguous' displays like those Barner & Snedeker used, which in this case are independently thought to suggest event versus activity categorization. Now,

[13]Or at least, it seems to us that the cognitive interpretation will directly support the linking hypotheses needed for Barner & Snedeker's result to count as evidence for the theory in the first place.

the strong prediction of the semantic theory can be tested: described with verb V, if some action independently suggests event categorization, then V *more* should be evaluated by number; if the scene suggests activity categorization, other dimensions should be permitted if not obligatory.

And so we proceed to the first step (Experiments 1 and 2), laying some of the preliminary foundations for the second (Experiments 3 and 4). Because we are only getting part of the way to the prediction that we ultimately aim to test, we can't answer here some of the interesting theoretical questions that would arise should the dynamic test go the way of Barner & Snedeker's static test. For example: what kind of a meaning is that of *more*? If we should consider that meaning as at least related to some 'domain-general' concept (e.g., Odic, 2018), what does that tell us about meaning in general, and about the cognitive power of language, specifically (cf. Spelke, 2003)?

For now, we turn to the experimental work.

3 Experiments

We investigate the evaluation of comparatives with event and activity verbs against dynamic displays that make multiple competing dimensions for comparison available. Testing adverbial comparatives that explicitly indicate the dimension for comparison (*more times*, *longer*, and *higher*), we can see how well participants are able to target and compare relations along each of the available dimensions. Testing verbal comparatives with bare *more*, we can see whether participants are sensitive to a target verb's event structure in selecting the dimension for comparison. Finally, we can test variants on the visual scene to see how that impacts dimensional selection when grammar leaves multiple options open.

As described below, we found that participants were highly accurate in their evaluation of adverbial comparatives along the stated dimensions, irrespective of the verb (Experiments

1 and 2). Moreover, they were highly consistent with a number-based evaluation for *jump more* (i.e., their responses were qualitatively the same as for the evaluation of *jump more times*; Experiment 1). In contrast, given the same dynamic scenes, our participants were far less likely to evaluate comparatives with *move more* by number, instead selecting number and distance at roughly the same rate (Experiment 2). Yet, 'undoing' certain features of our dynamic displays to make them more activity-like did not substantially impact participants' choice of number versus distance for *move more* (Experiments 3 and 4).

In what follows, we first describe the methodological similarities and differences between Experiments 1-4 taken together, followed by an overview of the results of the four experiments, and finally detailed presentation of the results of the individual experiments. While this order of presentation is non-standard, we believe it helps to eliminate much redundancy in the text and to ensure that the similarities and differences between the experiments can be quickly and easily grasped.

3.1 Overview

Our four experiments all involved asking comparative questions about two objects, A and B, where the extent to which each of their movements instantiated different dimensional values (number, height, and duration) was varied independently. In Experiments 1 and 2, these movement patterns were programmed to look, as much as possible, like jumps: the velocity of A or B's back-and-forth movement would change as the object approached the maximum point in its trajectory, and again as it made its return; and the object would pause briefly between one back-and-forth movement and the next. Experiment 1 tested the evaluation of comparatives with *jump* against these displays, where Experiment 2 used *move* and the same displays. Experiments 3 and 4 only tested the evaluation of comparatives with *move*. Experiment 3 eliminated velocity changes from the displays, and Experiment 4 further eliminated the pauses in between each back-and-forth movement.

We predicted that, presented with a comparative that explicitly states the intended dimension for comparison (i.e., *more times*, *higher*, and *longer*), participants would base their evaluation on the appropriate dimension (number, height, and duration, respectively), regardless of the verb. Presented with bare *more*, the lexical semantics of the verb should impact participants' dimensional choices: they should use number for *jump more* because *jump* is an event verb; however, because *move* is an activity verb that can be used in event VPs, participants in principle should be more flexible in their dimensional choices here. Since we anticipated that any of number, height, or duration could be viable options for quantification with *move more*, our experiments were designed so that we would be able to tease out which of these dimensions participants preferred to use.

Thus, Experiments 1 and 2 were designed to establish two things. First, that the lexical semantics of the verb plays no role in dimensional selection for explicit adverbial comparatives, and second, that the lexical semantics of the verb does play a role in the evaluation of comparatives with bare *more*. Experiments 3 and 4 inquired as to whether certain low level features of the visual scene could push around participants' selection of number versus height when evaluating *move more*, while leaving responses to the adverbial comparatives unchanged. The modifications we made to the dynamic displays of Experiments 1 and 2 that define Experiments 3 and 4 might, we reasoned, reduce the salience of a 'jump'-type parse of the scene, and thus lead participants to prefer the activity VP parse of *move more*. If so, this would decrease the proportion of number-based responses.

To preview our results, we found that people interpreted *jump more* the same as *jump more times*, but they did not interpret *move more* this way. Even for identical displays, *more* latches onto a different dimension depending on the event structure of the verb, as we expected. Yet, in contrast, we did not find that independent manipulations of the visual scene predicted which dimension people would land on for *move more*, which we found uniformaly supported both the resolution 'move more times' and 'move farther'.

3.2 Methodology

3.2.1 Participants

All of our participants were Northwestern University undergraduate students recruited through the Linguistics Department subject pool in accord with approved Institutional Review Board practices. Each received 1 lab credit for their participation, and each saw 240 trials. Study participation lasted 45 minutes on average. The total number of participants and observations by experiment are reported in Table 2.

Experiment	**Number of Participants**	**Total observations**
1	20	4800
2	20	4800
3	21	5040
4	21	5040

Table 2: Participant information for all experiments

3.2.2 Design

Participants evaluated comparative questions against scenes of a red star (object A in our shorthand) and a blue heart (object B) moving. Experiment 1 tested comparatives with *jump* and Experiments 2-4 tested comparatives with *move*. Each experiment manipulated 2 factors: COMPARATIVE and SIMULTANEITY. COMPARATIVE had 4 levels (*higher, longer, more times,* and *more*) which, combined with the verb tested in a given experiment, defined the questions that we asked participants; i.e. (12) and (13).

(12) Questions for Experiment 1:

a. Did the red star jump HIGHER than the blue heart?

b. Did the red star jump LONGER than the blue heart?

c. Did the red star jump MORE TIMES than the blue heart?

d. Did the red star jump MORE than the blue heart?

(13) Questions for Experiments 2, 3, and 4:

a. Did the red star move HIGHER than the blue heart?

b. Did the red star move LONGER than the blue heart?

c. Did the red star move MORE TIMES than the blue heart?

d. Did the red star move MORE than the blue heart?

COMPARATIVE was manipulated within subjects, in order to allow us to compare how the evaluation of bare *more* comparatives (the test conditions) compared with that of each of our adverbial comparatives (the control conditions) for each participant. The control for number was *more times,* the control for height was *higher*, and the control for duration was *longer*. Since participants responded with a simple 'yes' or 'no' to each question, we could compare responses to the test conditions by transforming these responses into each of three 'correct by' measures, e.g. a participant's 'yes' or 'no' response was coded as 1 for 'correct by number' if that response was predicted by a number-based comparison, and 0 otherwise. (See Section 3.2.3). Each set of 'correct by' measures could then be used to compare responses to test questions with *more* and the appropriate control question.

The factor SIMULTANEITY had 2 levels (sequential and simultaneous) and was manipulated within subjects. Half of the videos each participant saw showed the red star and the blue heart's movement patterns sequentially and the other half simultaneously. In sequential trials, the red star completed all of its back and forth movements before the blue star started moving, and in simultaneous trials the two started at the same time. Initial pilot studies showed that some dimensions were easier or harder to track depending on this presentation mode: duration (and sometimes height) was more difficult in the sequential presentation, whereas number was more difficult in the simultaneous presentation. By incorporating both modes, we can control for any effects of presentation mode on dimensional selection. In particular, we expect to see different 'correct by' measures impacted differently depending

on the presentation mode, in line with our pilot observations.

3.2.3 Stimuli

The stimuli for this experiment were created in Matlab version 8.6 using Psychophysics toolbox (Brainard 1997; Pelli 1997; Kleiner, Brainard, and Pelli 2007).

First, the background against which our objects moved differed based on the verb tested in the experiment (see Figure 1, Table 4), in order to differentially make description in terms of *jump* or *move* more felicitous.

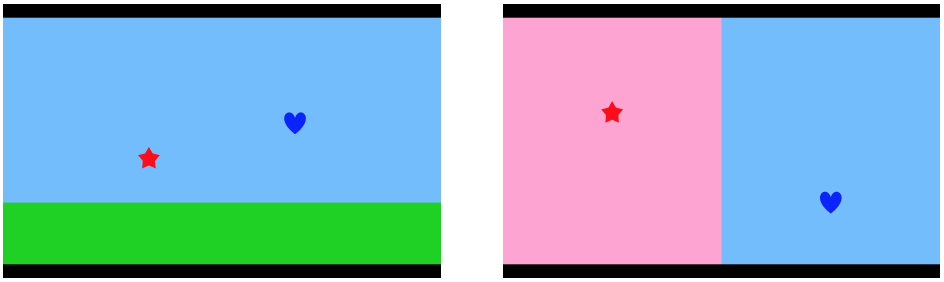

Figure 1: Screenshot of trial display for horizontal split (left) and vertical split (right)

In each of our dynamic displays, our two objects moved different numbers of times, reaching different heights at differing durations. The parameters determining each object's particular movements were drawn from the set in Table 3. (Due to a programming error, the possible heights in Experiments 3 and 4 differed from those of Experiments 1 and 2.) Each row in Table 3 defines a 'parameter set', indexing values along each of our three dimensions (e.g., 2, 600, 8 describes a parameter set of 2 moves to 600 pixels high for 8 seconds). Each experimental trial was defined by assigning non-identical parameter sets to the two objects, generating 30 unique pairs of parameter sets, or trial types.[14] On any given trial, object A "won" according to one or two of the parameters, and B won on the remainder. Thus, any two dimensions would agree on the 'winner' about 33% overall in Experiments 1 and 2; due

[14]There was one trial that differed from this design due to a programming error. The trial where A's parameters were the set 3, 800, 4, and B's were the set 2, 600, 8 was accidentally coded to have A's height to be 600 pixels.

to the programming error affecting the height values in Experiments 3 and 4, the level of agreement between pairs of parameters was 66%.

Number	**Duration** (seconds)	**Height** (pixels) E1, E2	E3, E4
2	8	600	480
2	6	800	280
3	8	400	680
3	4	800	280
4	4	600	480
4	6	400	680

Table 3: Parameter sets for dimensions of movement (including separate heights for Experiments 1 and 2 versus Experiments 3 and 4)

The specific manner in which the objects moved was the same for Experiments 1 and 2, but differed by design for Experiments 3 and 4 (see Table 4). Initially, we wanted the objects' movements to look as much like real jumping as possible. In Experiments 1 and 2, the objects' back-and-forth movement pattern followed a sine curve, such that they slowed down as they moved upwards and sped up as they returned to their starting position, and they paused briefly between each movement.[15] Experiment 3 subtracted the sine curve pattern, and Experiment 4 furthermore subtracted the temporal pauses between movements.

Experiment	**Sine curve**	**Pauses**	**Screen split**
1	✓	✓	horizontal
2	✓	✓	vertical
3	✗	✓	vertical
4	✗	✗	vertical

Table 4: Display and animation differences between experiments

[15]The temporal pauses contributed to our calculation of total duration. Due to a programming error, however, the pauses between jumps were 0.10 seconds instead of the intended 0.15 seconds, which puts the actual duration off by 0.05 - 0.15 seconds from the round numbers presented in Table 3.

3.2.4 Blocking

All of our experiments were blocked by SIMULTANEITY, with half of the participants seeing all of the sequential animations first and the other half seeing the simultaneous animations first. We made this choice in order to avoid any potential switching costs related to changes in the presentation mode from trial to trial. Within each of the two blocks, trials were furthermore divided into four sub-blocks, one for each level of the COMPARATIVE factor, also to avoid incurring any potential switching costs. The order of thesub-blocks was randomized, and each contained the full set of 30 unique trial types as discussed above. The order of these trials was randomized within each sub-block.

3.2.5 Procedure

The experiments were run on a computer using Matlab. Following the informed consent process, the participant began the study on an instructions screen, and pressed the space bar to proceed to the experiment. This began the first block, which started by telling the participant which question they would be asked for the first sub-block.[16] Pressing the space bar again advances to the first trial. For each trial, the participant saw a fixation cross, followed by the animation, and then they would see a response screen with a question and response key options (see Figure 2). Participants were asked to respond to each question with a key press, where F indicated 'yes' and J indicated 'no'. Pressing F or J advances to a new screen, which prompts the participant to press the space bar to advance to the next trial, or to advance to the next block or sub-block if they have completed the current sub-block.

[16]We hypothesized that asking the question first would guide participants' attention to their preferred dimension and thus reduce noise in the measurements. Asking the target question following the presentation of the stimulus would require participants to encode all relevant dimensional information; it may be interesting to know how many of these dimensions, and with what accuracy, can be faithfully recorded, but these questions are orthogonal for present concerns.

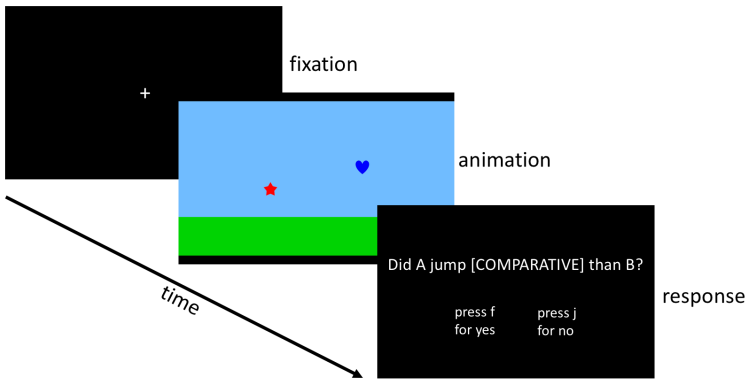

Figure 2: Trial structure with stimulus from Experiment 1

3.2.6 Data coding and analyses

For all experiments, we coded our participants' raw 'yes'/'no' responses using three 'consistency measures', one for each of the dimensions number, height, and duration (cf. Section 3.2.2). For example, on a given trial A may have exceeded B along the dimensions number and height, but not duration; thus, a 'yes' response on this trial would be counted as consistent for number and for height, but not for duration; a 'no' response would register as consistent for duration, but not for number or height. In our statistical analyses, we conducted three separate sets of analyses based on each of these consistency measures: for consistency with number, the set of responses in the *more* condition was compared with those for *more times*; for consistency with height, *more* was compared with *higher*; and for duration, *more* was compared with *longer*.

To conduct these statistical analyses for Experiments 1 and 2, we constructed distinct subsets of our data that overlapped in the set of responses to the test comparative *more*, and which differed otherwise in including responses to the control comparative relevant for a given consistency measure (e.g., checking consistency by number for *more* uses the 'correct by number' measure, and those responses are compared to *more times* responses using the same measure). For each of these subsets, we conducted generalized linear mixed effects model comparisons with maximal random effects structure, including random slopes and intercepts

by subject (Barr, Levy, Scheepers & Tily, 2013), and using the relevant consistency measure as the dependent variable (e.g., the responses to *more* and *more times* were compared based on the 'consistency by number' measure). In each analysis, we report model comparisons for our 2 factors, which were contrast coded: COMPARATIVE (two levels, different for each subset of the data analyzed) and SIMULTANEITY. The significance levels that we report for a given factor were calculated by comparing the relevant maximal model to a nearly identical model differing only in its exclusion of the relevant factor. All analyses were conducted using R's *lme4* package (Bates, Maechler, Bolker & Walker, 2014).

Additionally, we conducted paired two-sided t-tests in Experiments 2-4, as noted below, by comparing participant means for consistency with number and height in the *move more* condition.

3.3 Results

3.3.1 Overview of findings

The main results of these experiments are summarized in Figure 3, which plots the consistency of responses based on number versus height, as a function of the verb combined with bare *more* versus the relevant control adverbial. As the figure shows, our participants were highly consistent with the expected dimension for the control conditions with adverbial comparatives, regardless of the verb, but dimensional choices differed with *more* depending on the verb. As we predicted, participants were highly consistent with number for *jump more*, much more so than for *move more*. In the latter case, participants roughly equivocated between number and height. This finding is significant in that people make crucial reference to the event structure of the verb in resolving dimensional selection with *more*.

As Figure 4 shows, the changes in the visual scenes did not impact participants' ability to carry out number-based comparisons of the movements (i.e., performance with *more times*),

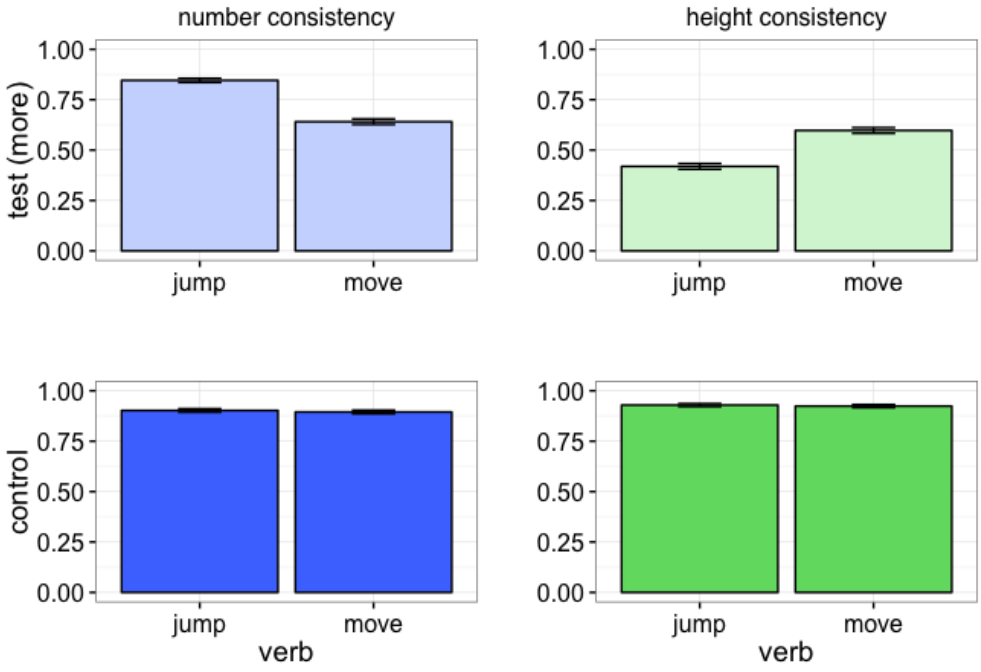

Figure 3: Consistency with number (left) and height (right) for test (top) and control (bottom) conditions for verbs *jump* (Experiment 1) and *move* (Experiment 2)

but those changes also failed to decrease participants' proportion of number-based choices with *move more*. In other words, participants persisted in choosing number and height equally often in our three *move* experiments. These findings suggest again a critical role for the linguistic information in fixing dimensional selection. However, before concluding that purely visual information is irrelevant in this respect, a number of additional questions must be addressed. We discuss this in more detail below.

3.3.2 Experiment 1: *jump*

In this experiment, we were interested in determining (i) the extent to which participants are able to estimate and compare number, height, and duration in simple dynamic scenes, as determined by their responses to our control comparatives, and (ii) whether they choose number as the relevant dimension for evaluating *jump more*. Our participants were asked all of the four questions that can be formed from (14) in separate blocks (once for simultaneous

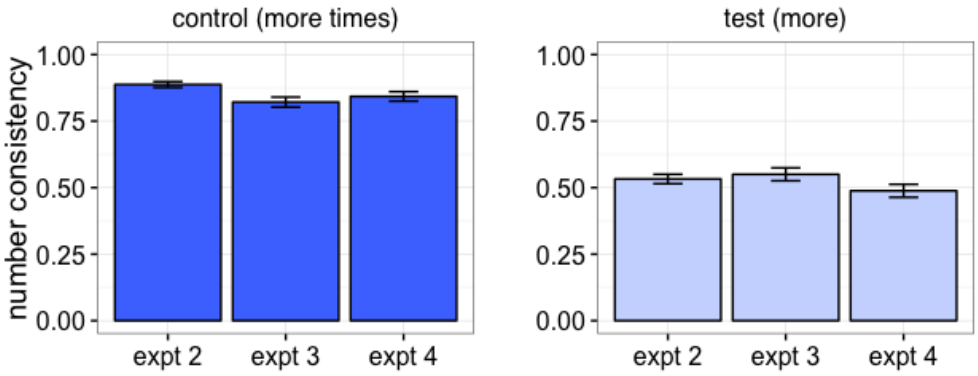

Figure 4: Consistency with number for trials that are unambiguous between number and height for Experiments 2, 3, and 4 for control condition (*more times*, left) and test condition (*more*, right)

presentation, and again for sequential).

(14) Did the red star **jump** HIGHER/LONGER/MORE TIMES/MORE than the blue heart?

In overview, we found for (i) that participants were highly accurate at tracking each of the dimensions for comparison that we varied, roughly equally so for each dimension, and for (ii) that participants' responses to *jump more* questions were nearly as highly consistent with number as were their responses to *jump more times* questions (see Figure 5).

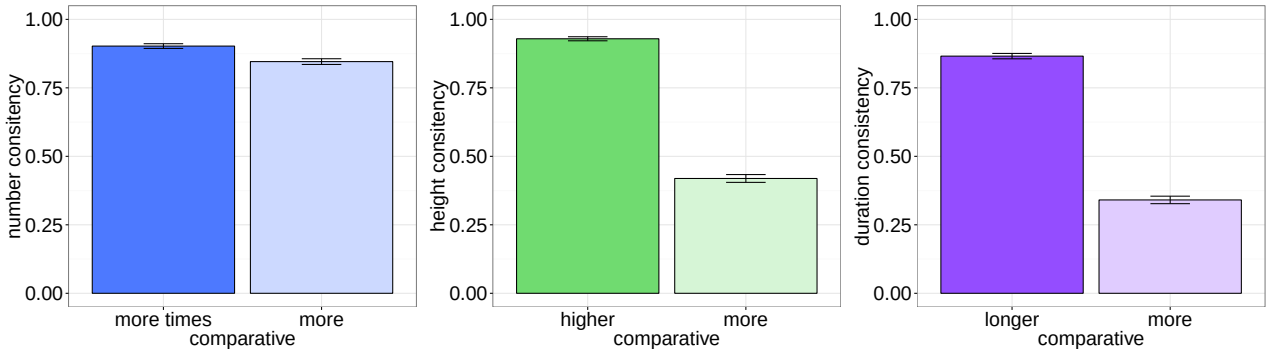

Figure 5: Experiment 1 responses coded for consistency with number (left), height (center), and duration (right). For each graph, the left bar plots the control comparative, and the right bar plots the test comparative *more*.

Table 5 shows the results of the generalized linear mixed effects model comparisons for our

two factors (see Section 3.2.6). First, we found that responses to *more* were only marginally different from responses to *more times* with respect to consistency by number ($p = 0.07$), while those responses were massively different from the relevant controls for consistency with height (*higher*) and consistency with distance (*longer*). Inspecting the means in Table 5, these results support the conclusion that *jump more* was judged roughly equivalently to *jump more times*, but completely differently from *jump higher* or *jump longer*.

As can also be seen in Table 5, we found that SIMULTANEITY had a significant effect on all three consistency measures, in the directions we expected based on our pilot studies. Consistency with number declined in the simultaneous mode relative to the sequential mode; in contrast, consistency with height and consistency with duration declined in the sequential mode relative to the simultaneous mode. These results validate our inclusion of both presentation modes to test dimensional specification with *more*. We found no interaction effects between SIMULTANEITY and COMPARATIVE with two of our consistency measures (number: $\chi^2 < 1, p = 0.35$; height: $\chi^2 < 1, p = 0.25$), but a marginal effect on the third (duration: $\chi^2 = 3.5, p = 0.06$).

Measure	**Factor**	**Level**	**Mean**	β	SE	χ^2	p
number	COMPARATIVE	*more times*	0.90	0.57	0.30	3.3	0.07
		more	0.85				
	SIMULTANEITY	seq	0.90	0.99	0.25	14.5	<0.001
		simul	0.85				***
height	COMPARATIVE	*higher*	0.93	3.12	0.23	48.4	<0.001
		more	0.42				***
	SIMULTANEITY	seq	0.66	-0.30	0.13	5.3	0.02
		simul	0.69				*
duration	COMPARATIVE	*longer*	0.87	2.66	0.20	46.5	<0.001
		more	0.34				***
	SIMULTANEITY	seq	0.56	-0.30	0.11	6.9	<0.01
		simul	0.62				**

Table 5: Experiment 1 generalized linear mixed effects model output for factors COMPARATIVE and SIMULTANEITY

These results were as expected based on the semantic theory. Our participants evaluated

our dynamic scenes based on the dimension specified explicitly by the control comparatives, and based on number when bare *more* was paired with *jump*. These results thus establish a baseline for 'ceiling' performance at evaluating number, height, and duration for our displays, and a basis for comparison of the effects of lexical semantics on that evaluation. If we replace the event verb *jump* with the activity verb *move*, while keeping the visual stimuli unchanged, will we see different dimensional choices with bare *more*?

3.3.3 Experiment 2: *move*

In this experiment, we wanted to determine (i) the extent to which people are able to estimate and compare the three possible dimensions available in our dynamic scenes in the control conditions, and (ii) whether they are flexible in their choice of dimension for evaluating *move more*. Participants were asked all of the four questions formed from (15) in separate blocks (once for simultaneous presentation, and again for sequential).

(15) Did the red star **move** HIGHER/LONGER/MORE TIMES/MORE than the blue heart?

In overview, we found for (i) that participants were highly accurate when tracking the appropriate dimension for comparison and (ii) that participants' responses to *move more* were flexible in the chosen dimension, choosing either number or height to evaluate the comparative (see Figure 6).

Table 6 shows the results of the generalized linear mixed effects model for our two factors. For COMPARATIVE, we see a significant effect for all three consistency measures. This result, combined with the mean consistency for the controls, suggest that participants were quantifying by the expected dimension for adverbial comparatives. It additionally shows that participants were not fully consistent with one potential dimension when evaluating *move more*. Instead, results show that participants are using both number- and height-based quantification at a level above what we would expect from potential overlapping dimensions,

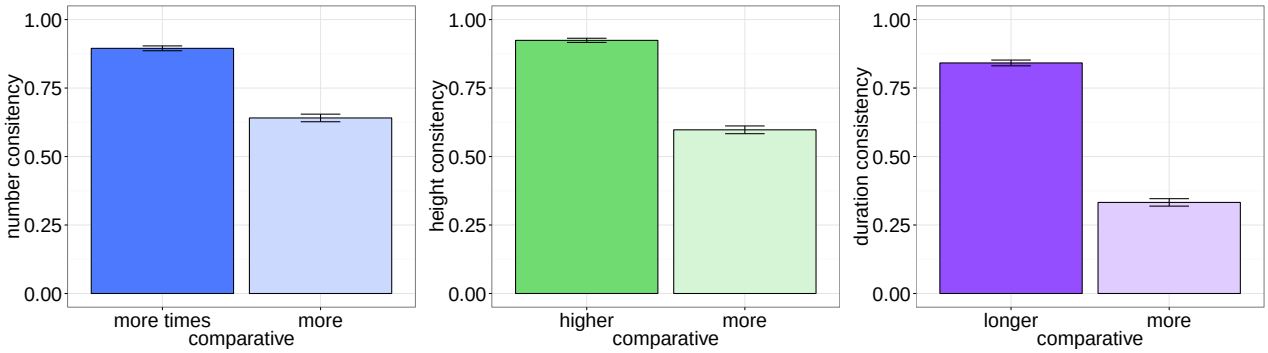

Figure 6: Experiment 2 responses coded for consistency with number (left), height (center), and duration (right). For each graph, the left bar plots the control comparative, and the right bar plots the test comparative *more*.

while they are not using duration-based quantification to evaluate *move more*.

We see in Table 6 that there are also significant main effects of SIMULTANEITY on both consistency with number and duration. As in Experiment 1, this result is not surprising given the relative difficulty of tracking different dimensions in different animation types. The results here are in line with our predictions.

Measure	**Factor**	**Level**	**Mean**	β	SE	χ^2	p
number	COMPARATIVE	*more times* *more*	0.90 0.64	1.85	0.20	34.4	<0.001 ***
	SIMULTANEITY	seq simul	0.79 0.75	0.51	0.19	6.7	<0.01 **
height	COMPARATIVE	*higher* *more*	0.92 0.60	2.52	0.30	30.2	<0.001 ***
	SIMULTANEITY	seq simul	0.74 0.78	-0.04	0.28	0.01	0.91
duration	COMPARATIVE	*longer* *more*	0.84 0.33	2.81	0.24	43.4	<0.001 ***
	SIMULTANEITY	seq simul	0.57 0.61	-0.54	0.21	6.1	0.01 *

Table 6: Experiment 2 generalized linear mixed effects model output for factors COMPARATIVE and SIMULTANEITY

While two of our consistency measures showed no interaction between COMPARATIVE and SIMULTANEITY (number: $\chi^2 < 1, p = 0.32$; height: $\chi^2 < 1, p = 0.56$), there was a sig-

nificant interaction for the consistency with duration measure (β=-1.21, SE=0.32, χ^2=11.7, $p < 0.001$). This interaction is such that only the COMPARATIVE level *longer* was largely impacted by the SIMULTANEITY (*longer*, sequential: 0.79; *longer*, simultaneous: 0.89; *more*, sequential: 0.34; *more*, simultaneous: 0.33). This result is easily explained by the fact that participants chose not to evaluate *move more* based on duration. Because they did evaluate *move longer* in this way, this evaluation was selectively impacted by trials where duration was harder to track.

Table 6 shows that number was used to evaluate *move more* 64% of the time, while height was used to evaluate *move more* 60% of the time. We found that there was no significant difference between the means (t(19)=0.75, $p = 0.46$).

Participants used both number and height to evaluate *move more*, which the semantic theory permits. It is unclear to us why our participants didn't use duration, however. They were able to track this dimension, as shown by performance on *longer* (both in Experiments 1 and 2), and duration should be a viable option. It is possible that our visual stimuli made two of the available dimensions more salient than the others, masking what might otherwise be the use of duration. We speculate that, if we made the animations look less event-like, and so more activity-like, that would lead participants to decrease their consistency with number, and make greater use of the available continuous dimensions.

3.3.4 Experiments 3 and 4

We conducted two alternate versions of Experiment 2, each 'undoing' a salient feature of the visual displays that we initially included to make the animations more jump-like (and so, potentially, more liable to event-based categorization and number-based quantification). Experiment 3 undid the changes in speed from the previous experiments, instead using a constant speed, and Experiment 4 furthermore undid the pauses between individual movements. In this section, we hone in on participants' responses to *move more*, and compare

those responses for their consistency with number with their consistency with height. We consider these two measures in particular in order to see if participants continue to use number and height equally, as they did in Experiment 2. Because the means for consistency with duration fall within the region of overlapping with another dimension in both experiments, we have no evidence that the experimental manipulations encouraged use of duration, and thus did not include this measure in this analysis.

Our results suggest that these differences in the manner of movement did not lead to different preferences for the evaluation of *move more* at all. For example, Experiment 2 showed equal use of number and height, while use of duration was in the region expected by overlapping dimensions (consistency with... number: 0.64; height: 0.60; duration: 0.33). These results are roughly equal to Experiment 3 (number: 0.81; height: 0.78; duration: 0.49), and to Experiment 4 (number: 0.78; height: 0.79; duration: 0.51).

Experiment 2 found that participants used number and height equally often to evaluate *move more*. If our present experimental manipulations decreased number-based responses, we would expect the means for consistency with number and consistency with height to be different. Yet, they appeared to be roughly equivalent both in Experiment 3 (number: 0.81, height: 0.78) and in Experiment 4 (number: 0.78, height: 0.79), and statistical analyses confirm this equivalence. We found that there was no significant difference between these means for Experiment 3 ($t(20)=1.28$, $p = 0.21$) or Experiment 4 ($t(20)=-0.36$, $p = 0.72$). This result suggests that, as in Experiment 2, participants were using number and height equally as often to evaluate *move more*, which further suggests that the experimental manipulations had no impact on their choice of dimension.

So far at least, we have no evidence that the visual scene pushes around dimensional preferences when grammar makes multiple ones available, as it does with *move more*. Future research is needed to know whether and when we might expect the visual scene to have this impact.

4 General Discussion

We showed that the semantic theory of adverbial and verbal comparatives makes good predictions when evaluated in a formal experimental setting. Comparatives with bare *more* and *jump* are evaluated by number, while those with *move* are more flexible. Furthermore, while we observed that dimensional selection was sensitive to verb semantics with *more*, the verb had little impact on that selection when the dimension to be used was specified explicitly (*more times*, *higher*, *longer*).

Experiment 1 showed that participants were able to quantify based on the expected dimensions for comparison in control conditions (i.e., those with adverbial comparatives), and that they also chose the predicted dimension of number for *jump more*. Experiment 2 confirmed that the dimensions chosen for quantification of adverbial comparatives does not differ by verb, while also showing that, even when the visual display is identical, the change in verb impacts quantification of the verbal comparative. Our participants used both number and height equally to evaluate *move more*. Perhaps surprisingly, we did not detect participants using duration for such comparisons, though their performance in the *move longer* condition suggests that this dimension was available to them.

With Experiments 3 and 4, we wanted to further investigate if we could push down the proportion of number-based responses by making the visual displays intuitively less event-like. However, these changes did not lead us to observe different preferences: our participants still used number and height equivalently often. Minimally, these results suggest that the lexical semantics of the verb was more important to dimensional selection (compare Experiments 1 and 2) than was the visual scene (compare Experiments 2, 3, and 4). However, it is still possible that visual properties could impact dimensionality—we just may not have uncovered the proper visual cues to manipulate.

This, we contend, suggests the need for follow-up studies that can tell us more about

how dynamic scenes are parsed, independently of language. In other words, it requires a cognitive psychology of the event/activity distinction.

References

Bach, Emmon. 1986. Natural language metaphysics. *Logic, Methodology and Philosophy of Science* 7. 573–595.

Bach, Emmon & Wynn Chao. 2012. Semantic types across languages. *Semantics. An International Handbook of Natural Language Meaning* 10. 2537–2558.

Bale, Alan & David Barner. 2009. The interpretation of functional heads: Using comparatives to explore the mass/count distinction. *Journal of Semantics* 26(3). 217–252.

Barner, David & Jesse Snedeker. 2004. Mapping individuation to mass-count syntax in language acquisition. In Kenneth Forbus, Dedre Gentner & Terry Regier (eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, 79–84. Chicago IL.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.

Bartsch, Renate & Theo Vennemann. 1972. *Semantic structures: A study in the relation between semantics and syntax*. Frankfurt am Main: Athenaum.

Barwise, John & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4. 159–219.

Bates, Douglas, Martin Maechler, Benjamin M. Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. `http://CRAN.R-project.org/package=lme4`.

Bhatt, Rajesh & Roumyana Pancheva. 2004. Late merger of degree clauses. *Linguistic Inquiry* 35(1). 1–46.

Brainard, David H. 1997. The Psychophysics Toolbox. *Spatial Vision* 10. 433–436.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

Cresswell, M. J. 1976. The semantics of degree. In Barbara Hall Partee (ed.), *Montague grammar*, 261–292. New York: Academic Press.

Ferreira, Marcelo. 2005. *Event quantification and plurality*. Boston MA: Massachusetts Institute of Technology dissertation.

Geurts, Bart & Rick Nouwen. 2007. 'At least' et al.: the semantics of scalar modifiers. *Language* 533–559.

Gillon, Brendan S. 2012. Mass terms. *Philosophy Compass* 7(10). 712–730.

Heim, Irene. 2000. Degree operators and scope. In Brendan Jackson & Tanya Matthews (eds.), *Proceedings of SALT X*, 40–64. Cornell University, Ithaca, NY: CLC Publications.

Kennedy, Chris. 1999. Gradable adjectives denote measure functions, not partial functions. *Studies in the Linguistic Sciences* 29(1). 65–80.

Kleiner, Mario, David Brainard & Denis Pelli. 2007. What's new in psychtoolbox-3? Perception 36 ECVP Abstract Supplement.

Koslicki, Katherin. 1997. Isolation and non-arbitrary division: Frege's two criteria for counting. *Synthese* 112(3). 403–430.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In Rainer Baeuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 302–323. Berlin, Germany: DeGruyter.

Moltmann, Frederike. 2017. Natural language ontology. Oxford Research Encyclopedia of Linguistics. `10.1093/acrefore/9780199384655.013.330`.

Nakanishi, Kimiko. 2007. Measurement in the nominal and verbal domains. *Linguistics and Philosophy* 30. 235–276.

Odic, Darko. 2018. Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science* 21(2).

Odic, Darko, Paul Pietroski, Tim Hunter, Justin Halberda & Jeffrey Lidz. 2018. Individuals and non-individuals in cognition and semantics: The mass/count distinction and quantity representation. *Glossa: a journal of general linguistics* 3(1).

Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. Cambridge, Massachusetts: MIT Press.

Pelletier, Francis Jeffry. 2011. Descriptive metaphysics, natural language metaphysics, sapir-whorf, and all that stuff: Evidence from the mass-count distinction. *Baltic International Yearbook of Cognition, Logic and Communication* 6(1). 7.

Pelli, Denis G. 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10. 437–442.

Pietroski, Paul. 2010. Concepts, meanings, and truth: First nature, second nature, and hard work. *Mind & Language* 25(3). 247–278.

Rips, Lance J. & Susan J. Hespos. 2015. Divisions of the physical world: Concepts of objects and substances. *Psychological Bulletin* 141(4). 786–811.

Schwarzschild, Roger. 2002. The grammar of measurement. In B. Jackson (ed.), *Proceedings of SALT XII*, 225–245. Cornell University, Ithaca, NY: CLC Publications.

Schwarzschild, Roger. 2006. The role of dimensions in the syntax of noun phrases. *Syntax* 9(1). 67–110.

Seuren, Pieter A. M. 1973. The comparative. In Ferenc Kiefer & Nicolas Ruwet (eds.), *Generative Grammar in Europe*, 528–564. Dordrecht: D. Reidel Publishing Company.

Spelke, Elizabeth S. 2003. What makes us smart? Core knowledge and natural language. *Language in mind: Advances in the study of language and thought* 277–311.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1). 1–77.

Wellwood, Alexis. 2018. Structure preservation in comparatives. In Sireemas Maspong, Brynhildur Stefánsdóttir, Katherine Blake & Forrest Davis (eds.), *Semantics and Linguistic Theory (SALT) 28*, 78–99. CLC Publications.

Wellwood, Alexis. 2019. *The meaning of* more. Studies in Semantics and Pragmatics. Oxford UK: Oxford University Press.

Wellwood, Alexis, Valentine Hacquard & Roumyana Pancheva. 2012. Measuring and comparing individuals and events. *Journal of Semantics* 29(2). 207–228.

Wellwood, Alexis, Susan J. Hespos & Lance Rips. 2018a. How similar are objects and events? *Acta Linguistica Academica* 15(2-3). 473–501.

Wellwood, Alexis, Susan J. Hespos & Lance Rips. 2018b. The *object : substance :: event : process* analogy. In Tania Lombrozo, Joshua Knobe & Shaun Nicholas (eds.), *Oxford Studies in Experimental Philosophy*, vol. II, chap. 8, 183–212. Oxford UK: Oxford University Press.