

Individuation, reliability, and the mass/count distinction*

Peter R. Sutton and Hana Filip
Heinrich Heine University Düsseldorf

ABSTRACT

Counting in natural language presupposes that we can successfully identify what counts as *one*, which, as we argue, relies on how and whether one can balance two pressures on learning nominal predicates, which we formalise in probabilistic and information theoretic terms: INDIVIDUATION (establishing a schema for judging what counts as one with respect to a predicate); and RELIABILITY (establishing a reliable criterion for applying a predicate). This hypothesis has two main consequences. First, the mass/count distinction in natural language is a complex phenomenon that is partly grounded in a theory of individuation, which we contend must integrate particular qualitative properties of entities, among which a key role is played by those that rely on our spatial perception. Second, it allows us to predict when we can expect the puzzling variation in mass/count lexicalization, cross- and intralinguistically: namely, exactly when the two learning pressures of INDIVIDUATION and RELIABILITY conflict.

Keywords:
mass/count distinction, probabilistic semantics, individuation, reliability, semantic learning, information theory, context-sensitivity, Type Theory with Records

1

INTRODUCTION

This paper attempts to combine state of the art research on the mass/count distinction in formal semantics with the cutting edge research in Type Theory with Records that provides a unified representation of cognitive, perceptual, and linguistic information. This allows

*This research was funded by the German Research Association (DFG), CRC 991, project C09. We would like to thank the attendees of the TYTTLES workshop at ESSLLI 2015 and the CLASP research seminar. In particular, Robin Cooper, Simon Dobnik, and Shalom Lappin for many helpful discussions.

us not only to unify two largely separate strands of research and enrich both with our novel contributions, but also, and most importantly, to further our understanding of the concept of *individuation* (what counts as one) relative to a predicate, which, as we argue, is the fundamental concept in countability research. The account proposed here covers a number of the complex and puzzling data that pertain to cross- and intralinguistic mass/count variation, which resist an adequate account within the extant theories of the mass/count distinction in formal semantics, to the best of our knowledge.

The outline of this paper is as follows. In Section 2, we introduce the basis for our semantic formalism: Type Theory with Records (TTR, Cooper 2012) and probabilistic type theory with records (prob-TTR Cooper *et al.* 2015). In Section 3, we outline some of the most influential recent theories of the mass/count distinction in formal mereological semantics, which are largely driven by the concept of individuation (what counts as one). Sections 4–7 focus on our new proposal. In Section 4, we enrich prob-TTR with mereological assumptions (probM-TTR). We then show how this formalism can represent, in detail, both qualitative and quantitative criteria for the application of nominal predicates (inspired by Krifka (1989)).

Novelly, building on Dobnik *et al.* (2012), we model how representations of spatial perceptual information in a given context can inform and affect judgements about what counts as an individual (as one) relative to a predicate. In Section 5, we relate the quantitative and qualitative criteria to probabilistic learning and argue that the ability to successfully individuate entities relative to a predicate, and thereby establish a basis for counting, is essentially tied to how one balances two learning pressures. The first is to establish a disjoint individuation schema for a predicate (INDIVIDUATION), the second is to establish reliable criteria for applying a predicate (RELIABILITY). In Section 6, we give a schema for the lexical entries of concrete nouns.

In Section 7, we show how mass and count encoding arises from the balancing of individuation and reliability with respect to nominal predicate learning. One of the advantages of our proposal over merely mereological ones is that the interaction of these two learning pressures also allows us to delimit the range of cases where one should expect to find cross- and intralinguistic mass/count variation in natural language.

2 BACKGROUND: PROBABILISTIC TYPE THEORY WITH RECORDS

2.1 *Type Theory with Records*

TTR integrates rich lexical semantic frame-based representations in the sense of Fillmore (1975, 1976) (and elsewhere) with a compositional semantics in the Frege-Montague tradition. As such, it is an ideal theoretical framework to investigate the lexical semantics of nouns as well as compositional (formal) properties of complex expressions. TTR also integrates the insights of situation semantics insofar as situation types (RECORD TYPES) are taken to be true of situations (RECORDS), rather than being true at possible worlds. The idea is that an agent judges whether a situation s , is of type T . Such judgements correspond to type theoretic objects, namely *Austinian propositions* (Barwise and Etchemendy 1987) inspired by Austin's (1950/1979) idea that to say something true (in indicative, non-generic cases at least) is to refer to a particular situation with one's utterance that is of the type expressed by the sentence one uses. Sentences can be used to express situation types, and utterances can refer to particular situations.

In the following, we present TTR as presented in Cooper (2012). Record types, which are displayed in tabular format (1), are sets of fields, i.e., ordered pairs, whose first member is a LABEL (to the left of the colon in (1)) and whose second member is a TYPE (to the right of the colon in (1)):

$$(1) \quad \left[\begin{array}{l} x \quad : \quad T_1 \\ \dots \\ y \quad : \quad T_n \end{array} \right]$$

Records, which are displayed in tabular format (2), are sets of fields, i.e., ordered pairs, whose first member is a LABEL (to the left of the '=' in (2)) and whose second member is a VALUE for this label (to the right of the colon in (2)):

$$(2) \quad \left[\begin{array}{l} x \quad = \quad v_1 \\ \dots \\ y \quad = \quad v_n \end{array} \right]$$

An example of a record type is given in (3), which represents the type of situation in which a cat purrs.

$$(3) \quad \left[\begin{array}{l} x \quad : \quad Ind \\ s_{cat} \quad : \quad \langle \lambda v. \text{cat}(v), \langle x \rangle \rangle \\ s_{purr} \quad : \quad \langle \lambda v. \text{purr}(v), \langle x \rangle \rangle \end{array} \right]$$

In the record type in (3), the first field contains a label x and a basic type Ind . In TTR, there is a set of basic types, such as Ind for *individual*, $Time$, and Loc (location). Predicates are functions. This can be seen in the second field. The label s_{cat} is a label for a situation, and the type contains a predicate which is a function from a value v to the type of situation in which this value is a cat. It is important to note that these predicates do *not* take labels as arguments, but rather the *values* of labels. Labels function as pointers to values. We will use an abbreviated conventional notation in this work, as illustrated in (4) (Cooper 2012, p.11):

$$(4) \quad \left[\begin{array}{l} x \quad : \quad Ind \\ s_{cat} \quad : \quad \text{cat}(x) \\ s_{purr} \quad : \quad \text{purr}(x) \end{array} \right]$$

Records are the entities of which record types are true or false ('proofs' of propositions in type-theoretic terminology, 'witnesses' in a natural language setting). For example, (5) specifies a situation in which there is an individual, *felix*, and two witnesses p and q . Witnesses can be thought of as situations or parts of the world that make type judgements true or false.

$$(5) \quad \left[\begin{array}{l} x \quad = \quad \text{felix} \\ s_{cat} \quad = \quad p \\ s_{purr} \quad = \quad q \end{array} \right]$$

Individuals in the domain are typed. The record in (5) will make the proposition in (3)/(4) true iff:

felix is of type Ind ,
 p is a witness of $\text{cat}(\text{felix})$,
 q is a witness of $\text{purr}(\text{felix})$.

An important aspect of TTR, however, is that it is a semantics that reserves a role for judgements made by agents (see Section 2.2). If the

type in (3)/(4) is T_1 and the situation represented by (5) is s , an agent may judge s to be of type T_1 , ($s : T_1$). This judgement will be true iff the above conditions hold.

Finally, natural language predicates denote properties of type $[x : Ind] \rightarrow RecType$, a function from records containing individuals, to a record type. For example, a simplified representation of *cat* would be as in (6):

$$(6) \quad \lambda r : [x : Ind]. \left(\begin{array}{l} r.x \quad : \quad Ind \\ s_{cat} \quad : \quad cat(r.x) \end{array} \right)$$

In (6), $r.x$ means that the value to be supplemented is the value of x in r . Hence, if provided with a record $[x = felix]$, (6) would, via β -conversion, yield the proposition that *felix is a cat* in (7):

$$(7) \quad [s_{cat} \quad : \quad cat(felix)]$$

2.2 Probabilistic Type Theory with Records

The following two subsections summarise some of the key details of probabilistic TTR from Cooper *et al.* (2014, 2015). The principal difference between prob-TTR and TTR is that judgements are graded as opposed to categorical. Instead of a judgement $s : T$, which is true or false, judgements hold with a probability $p(s : T) = k \in [0, 1]$, or the probability that an agent will judge a situation s to be of type T . The reason for introducing probabilities is to be able to model the inherent gradience in the ways in which we classify parts of the world when we apply predicates of natural language to them. This type of gradedness is directly represented by the inherent gradedness in metalinguistic uncertainty within the range $[0, 1]$.

A major advantage of prob-TTR is that one can model how probability values can be assigned, in a cognitively plausible way, and how probability distributions can be updated via observation and semantic learning. Witnessing how language is used provides localised information that, over time, helps language learners build a probability distribution that guides them how to use language and that approximates the ‘true’ distribution which underpins how competent speakers use language. This is in contrast to top-down probabilistic approaches that assign (global) probability distributions over sets of possible worlds. This issue is discussed in more depth in Cooper *et al.* (2015).

Meet types and join types in prob-TTR respect the Kolmogorov axioms for probability (Kolmogorov 1950):

$$(8) \quad p(a : T_1 \wedge T_2) = p(a : T_1) \times p(a : T_2 | a : T_1)$$

$$(9) \quad p(a : T_1 \vee T_2) = p(a : T_1) + p(a : T_2) - p(a : T_1 \wedge a : T_2)$$

2.3 Bayesian learning in prob-TTR

Assuming that agent-learners continually receive new evidence with respect to how to correctly apply types to aspects of the world, this can be modelled, in line with Bayesian approaches to cognition, as continuous updates of the probability distributions in the light of new evidence. The probability distributions of learners will gradually come to be close to those of competent speakers. The way this is modelled in prob-TTR is that agents maintain judgement sets.

In simple and intuitive terms, when an agent makes a judgement about a given situation, an entry in the agent's judgement set is made. Entries in judgement sets record the probability that the encountered situation is of some type. Members of judgement sets are what Cooper *et al.* (2015) refer to as *probabilistic Austinian propositions*.¹ For example, the probabilistic Austinian proposition involving a cat purring, judged with a probability of 0.9 would be:

$$(10) \quad \left[\begin{array}{l} \text{sit} \\ \text{sit-type} \\ \text{prob} \end{array} \right] = \left[\begin{array}{l} s_1 \\ \left[\begin{array}{l} x : \text{Ind} \\ s_{cat} : \text{cat}(x) \\ s_{purr} : \text{purr}(x) \end{array} \right] \\ 0.9 \end{array} \right]$$

Proposition (10) records a situation, a type and a probability value for the judgement that the situation is of that type. The set of such judgements, i.e., the set of probabilistic Austinian propositions, for

¹ Cooper *et al.* (2015) present a naive Bayesian learning model, a picture that is highly simplified. (They state an intention to develop a more sophisticated learning model.) On the simple model, agents are presented with discrete situations, probabilistic judgements are made and the situation, situation type and judgement value are recorded in the judgement set. A more plausible model would have to incorporate the dynamic development of situations and how judgements will very often be implicit.

an agent A will be the set \mathfrak{J} . For a type T , \mathfrak{J}_T is the set of Austinian propositions j such that $j.\text{sit-type} \sqsubseteq T$:

$$(11) \quad \mathfrak{J}_T = \{j \mid j \in \mathfrak{J}, j.\text{sit-type} = T\}$$

The sum of probabilities associated with that type T in \mathfrak{J} is:

$$(12) \quad \|T\|_{\mathfrak{J}} = \sum_{j \in \mathfrak{J}_T} j.\text{prob}$$

Priors are calculated from sums over entries in the judgement set:

$$(13) \quad \text{prior}_{\mathfrak{J}}(T) = \frac{\|T\|_{\mathfrak{J}}}{\sum(\mathfrak{J})} \quad \text{if } \sum(\mathfrak{J}) > 0 \text{ and } 0 \text{ otherwise.}$$

It is worth contrasting this approach with more top-down models in terms of possible worlds. What this system provides, in contrast to more top-down models, is an explanation of how priors are set. In probabilistic possible worlds-based approaches such as that of Eijck and Lappin (2012), one must assume a set of priors over possible worlds from which priors for particular sentences/propositions are calculated. On top of issues of computational tractability (discussed in Cooper *et al.* (2015)), this approach leaves it unexplained on what basis the priors for possible worlds are calculated.

The judgement set also provides a simple way to estimate conditional probabilities such as likelihoods and posteriors. For example, suppose \mathfrak{J} were to contain the record in (10) and the record in (14):

$$(14) \quad \left[\begin{array}{l} \text{sit} \\ \text{sit-type} \\ \text{prob} \end{array} \right] = \left[\begin{array}{l} = s_2 \\ = \left[\begin{array}{l} x \\ s_{cat} \end{array} \right] : \left[\begin{array}{l} Ind \\ cat(x) \end{array} \right] \\ = 0.9 \end{array} \right]$$

We could then calculate the probability of there being a situation in which something purrs given it is a cat. $p_{\mathfrak{J},A}$ is a probability function with respect to a judgement set \mathfrak{J} , and an agent A . Conditional probabilities are calculated in terms of a type theoretic version of Bayes' Rule:

$$(15) \quad p_{\mathfrak{J},A}(s : T_1 \mid s : T_2) = \frac{\|T_1 \wedge T_2\|_{\mathfrak{J}}}{\|T_2\|_{\mathfrak{J}}}$$

For the case in hand, this yields:

$$(16) \quad P_{\exists, A}(s : \begin{bmatrix} x & : Ind \\ s_{cat} & : cat(x) \\ s_{purr} & : purr(x) \end{bmatrix} \mid s : \begin{bmatrix} x & : Ind \\ s_{cat} & : cat(x) \end{bmatrix}) = \frac{0.9}{0.9 + 0.9} \\ = 0.5$$

which follows given that:

$$\begin{bmatrix} x & : Ind \\ s_{cat} & : cat(x) \\ s_{purr} & : purr(x) \end{bmatrix} \sqsubseteq \begin{bmatrix} x & : Ind \\ s_{cat} & : cat(x) \end{bmatrix}$$

and that if $T \sqsubseteq T'$, then $p(a : T \wedge T') = p(a : T')$. Intuitively, if two situations involving cats are observed with equal probability, but only one is a situation involving purring, then the probability of a situation involving a purring cat, given it involves a cat should be 0.5.

3 BACKGROUND: VAGUENESS, OVERLAP, MASS/COUNT VARIATION AND INDIVIDUATION

In this section, we briefly introduce some of the state of the art semantic accounts of the mass/count distinction in concrete nouns.² The accounts we discuss here are all based on enriching formal semantics with mereology (first proposed by Link (1983)). In mereological semantics, domains of entities form a Boolean semilattice closed under sum (\sqcup). That is to say: domains of entities are populated not just with individuals (a, b, c), but also with sums of entities ($a, \sqcup b, a \sqcup b \sqcup c$ etc.) which are of the same semantic type. The use of mereology has, since Link (1983), proved highly fruitful in analysing both plurality and the count/mass distinction.

We highlight two key factors that have been proposed to play a role in the mass/count distinction as part of a mereological semantics: (i) vagueness (understood in terms of a kind of context-sensitivity) (Chierchia 2010); (ii) disjointness vs. overlap at a context in a noun's denotation (Rothstein 2010; Landman 2011).

²For a more thorough critical analysis, see Sutton and Filip (2016a,b).

3.1 *Context sensitivity: vagueness and overlap*

Chierchia's (2010) main claim is that mass nouns are *vague* in a way that count nouns are not. Mass nouns are uncountable, because they lack *stable atoms*. Stable atoms are the entities in the denotation of a noun that are atoms in every context relative to a *ground context*. Ground contexts determine the entities that are denotations of a noun at every context (but are not necessarily atoms). If a noun lacks stable atoms (has only unstable individuals), then there is no entity that is an atom in the denotation of the predicate at all contexts.

Nouns such as *rice* are vague in so far as the minimal entities in the denotation of *rice* may vary depending on context. Sometimes they are sums of a few grains, sometimes single grains, half grains, or even rice flour dust. Thus these quantities of rice are in the vagueness band of *rice*. Chierchia (2010) models this vagueness with a supervaluationist semantics. At some total precisifications of a ground context, c , single grains are rice atoms. Where $c \propto c'$ means that c' precisifies c ; then at some c' such that $c \propto c'$, half grains are rice atoms. At some c'' such that $c' \propto c''$, rice dust particles are rice atoms. There is, therefore, no entity that is a rice atom at every total precisification of rice. The denotation of *rice* lacks stable atoms, but counting is counting stable atoms, and so *rice* is mass.

Rothstein (2010) focuses on providing a formal model of how nouns such as *fence* and *wall* – which fail to denote entities with “natural units” in the sense of Krifka (1989) – nonetheless behave like ordinary count nouns.³ Rothstein (2010) coins the term *counting context*, and defines count nouns as typically distinct from mass nouns. Mass nouns are of type $\langle e, t \rangle$. Count nouns, which are of type $\langle e \times k, t \rangle$, denote sets of entity-context pairs (the entity denoted and the context in which it counts as one). To take Rothstein's example, suppose that a square field is encircled by fencing. The question ‘How many fences are there?’ has no determinate answer, but rather the answer depends on what counts as one disjoint fence in a given context. In some con-

³However, Filip and Sutton (2017) argue that nouns such as *fence* are not bona fide count nouns, since they are felicitous as bare singulars in some measure constructions, for example, “*You will need about 150 yards of fence per acre*” (BNC). Bona fide count nouns are much less felicitous in such constructions. For example, *#I read 10cm of book*.

texts, it would be natural to answer ‘four’: namely, one for each side of the field. In other contexts, it would be more natural to answer ‘one’: namely, one fence encircling the whole field. By indexing count nouns to contexts, Rothstein can explain how there can be one single answer to the question of how many fences there are in any particular context, despite *fence* lacking natural atoms in its denotation, i.e., atoms that are independent of counting-contexts. Countability, in Rothstein’s view, is a matter of what might be dubbed *disjointness at a counting context*.

In Landman (2011), counting is a matter of *non-overlap in a given context*. He defines a set of generators which contains “the things that we would want to count as one” (Landman 2011, p. 26) relative to a context. Formally, generator sets generate a noun’s whole denotation under sum. If the elements in the generator set are non-overlapping, as in the case of count nouns, then counting is sanctioned: counting is counting elements in the generator set and there is only one way to count. However, if generators overlap, as in the case of mass nouns, counting goes wrong, because it leads to a number of different simultaneous counting results. Formally, this is modelled as maximally disjoint subsets which generate the superset under sum (variants). In the above, “a number of different simultaneous counting results” equates to a variation in cardinality across variants. One of Landman’s innovations is to provide a new delimitation of the two cases when this happens: *mess* mass nouns like *mud*, and *neat* mass nouns like *kitchenware* (a.k.a. ‘object’ or ‘fake’ mass nouns). A noun is a *mess* mass noun if, at every world, its intension determines a regular generator set whose set of minimal elements is overlapping. A noun is a *neat* mass noun if its intension at every world specifies a regular generator set whose set of minimal elements is non-overlapping.

Landman offers an ingenious solution to the perennial problems posed by mass nouns like *kitchenware*, *furniture*, *silverware* and the like. Let us take his paradigm example *kitchenware*:

“The teapot, the cup, the saucer, and the cup and saucer all count as kitchenware and can all count as one simultaneously in the same context. ... In other words: the denotations of *neat nouns* are sets in which the distinction between *singular individuals* and *plural individuals* is not properly articulated.” (Landman 2011, pp. 34–35)

The key idea here is that there are contexts which allow overlap in the denotation of a noun N with respect to what counts as ‘one N’. In other words, there are contexts in which, either one simply does not apply an individuation schema, or, alternatively, the individuation schema one applies fails to resolve overlap; in either case, overlap is not made ‘irrelevant’, and therefore counting goes wrong.

3.2 *The puzzle of mass/count variation*

All three of the aforementioned analyses of the mass/count distinction make significant advances in accommodating the puzzling data that display cross- and intralinguistic mass/count variation. However, each account taken individually cannot accommodate the full range of such data. We take five broad classes of nouns as cases in point. Two of these, the prototypical cases, pose no problems for most accounts of the mass/count distinction:

Prototypical objects: Examples in English are *cat, car, boy, chair*. These nouns show a very strong intra- and crosslinguistic tendency towards being count.⁴ They are not vague in Chierchia’s sense, not counting context-sensitive in Rothstein’s sense, and not overlapping in Landman’s sense.

Substances: Examples in English are *mud, air, blood, slime*. These nouns show a very strong intra- and crosslinguistic tendency towards being mass.⁵ They are vague in Chierchia’s sense, not indexed to counting contexts in Rothstein’s semantics, and have overlapping minimal generators (are mess mass) in Landman’s sense.

Granulars: Examples in English are *lentils, rice, oats, beans*. These nouns show a significant amount of variation in mass/count encoding such as in (17) and (18):

- (17) *lentil-s*_{+C,PL}; *linse-n*_{+C,PL} (German); *lešta*_{-C} (Bulgarian);
*čočka*_{-C} (Czech).
- (18) *oat-s*_{+C,PL}, *oatmeal*_{-C}; *kaura*_{-C} (Finnish);
*kaurahiutale-et*_{+C,PL} (Finnish, lit. oat. flake-s).

⁴There are some languages, such as Brazilian Portuguese, which also license a non-coerced mass reading of many count nouns. See Pires de Oliveira and Rothstein (2011) for discussion.

⁵There are some languages, such as Yudja, which also license a non-coerced count reading of many or even all mass nouns. See Lima (2014) for discussion.

Granulars are vague in the sense of Chierchia (2010), but if vagueness were the only factor in mass/count encoding, these data could not be accommodated.⁶ Rothstein's account can introduce a typal difference between, for example, *rice* and *lentils*, but does not have the formal tools to explain why a typal distinction should arise commonly for these nouns, but not for, say, prototypical count nouns. Landman (2011) faces a challenge, given that it is unclear why the English *lentil*, for example, should be count, whereas its Bulgarian counterpart *lešta* ('lentil') would presumably come out as either a neat or a mess mass noun (depending on how Landman's theory is applied to this case). That is, it is unclear – given Landman's account – why granulars should license non-overlapping generators in some languages but overlapping generators in others.

Collective artifacts: These nouns, examples of which in English are *furniture*, *kitchenware*, *footwear*, *equipment*, show a significant amount of variation in mass/count encoding, such as we see in (19) and (20):

(19) *furniture*_{-C}; *huonekalu*-t_{+C,PL} (Finnish);
meubel-s_{+C,PL}, *meubilair*_{-C} (Dutch).

(20) *kitchenware*_{-C}; *Küchengerät*-e_{+C,PL} (German, lit. kitchen device-s).

Collective artifacts are recognised to be exceptions to a vagueness based analysis of the mass/count distinction and as requiring a separate source for their mass/count encoding (Chierchia 2010, pp. 136–139). For Landman, collective artifacts constitute the key data points for developing his theory, and to this goal, he focuses on Dutch examples like (19).

Although Landman's theory is not explicitly intended to account for cross-linguistic variation in mass/count encoding, it could be extended to do this job too. A possible line one could then adopt is that mass/count variation is only licensed for neat nouns which can have overlapping generators 'simultaneously in the same context'. If neat mass nouns have overlapping generators simultaneously in the same context, and overlap means MASS, then we may ask why the count noun counterparts of neat mass nouns are count nouns. In Landman's analysis, they have non-overlapping generator sets. However, presum-

⁶Chierchia (2010) is aware of this problem; however it is only informally addressed (Chierchia 2010, p.140).

ably, at different contexts, exactly what counts as one can vary. For example, in some contexts a vanity counts as one *huonekalu* (the Finnish count noun counterpart of the English neat mass *furniture*); in other contexts, it counts as at least two (the mirror and the table etc.).

If the count noun counterparts of neat mass nouns are context sensitive with respect to what counts as one across contexts, then arguably, the count nouns in (19) and (20) are counting context-sensitive in the sense of Rothstein (2010) (just like *fence*). A possible extension to Rothstein's account is that one tends to find cross- and intralinguistic mass counterparts of count nouns that are counting context sensitive. Indeed, the link between these two classes suggests, to us, that the explanation of why count/mass variation is found within them should have a common explanation. We develop these lines of thought in Section 7 in the light of the formal analysis we develop in Sections 4-6.

Non-bounded objects: Examples in English are *fence*, *wall*. These nouns are usually count in their morphologically simple form, but frequently have derived mass counterparts:

(21) *fence*_{+C} - *fencing*_{-C}; *wall*_{+C} - *walling*_{-C}

Chierchia (2010) argues that the count versions of these nouns are not vague (with respect to their minimal countable entities), given that the ground context is fixed (Chierchia 2010, pp. 122–123). As such, the mass counterparts provide a challenge to a vagueness-only based account.

Overlap/non-overlap based accounts may fare better when it comes to *non-bounded objects*. Indeed, Rothstein's and Landman's accounts could be extended in a similar way just outlined for collective artifacts. Namely, the mass counterparts of count non-bounded object nouns are neat mass (licensing a count counterpart), and the count counterparts (*fence*) of mass non-bounded object nouns (*fencing*) are counting context sensitive (licensing a mass counterpart). Indeed, the link between the non-bounded objects and collective artifacts classes suggests, to us, that the explanation of why count/mass variation is found within them should have a common explanation.⁷ We pursue this in Section 7.

⁷ See Sutton and Filip (2016a) for in-depth discussion.

In summary, two ways in which context is important emerge from these three accounts. First, the extension of a noun may vary across contexts with respect to its atomic elements (Chierchia 2010). Second, the entities that ‘count as one’ in the denotation of a noun may vary across contexts thereby yielding either a disjoint set of individuated entities (Rothstein 2010), or an overlapping set in which all possible individuated units appear simultaneously (Landman 2011).

3.3 *Individuation and two criteria of applicability for nouns*

Here we briefly review how both qualitative and quantitative criteria for the application of noun predicates have been highlighted as important for the semantics of the mass/count distinction and individuation. Specifying these two criteria originates in the work of Krifka (1989), but echoes of it percolate through his later work and that of others. The majority of responses to Krifka’s work have focused on improving his representation of the quantitative criteria for the application of count predicates. We will also detail how these qualitative and quantitative criteria come together to feed into an account of individuation in the form of mereotopological properties (Grimm 2012).

Krifka (1989) proposed that the semantic representation of (concrete) count nouns involves two criteria of applicability: one qualitative, and one quantitative. For example, *one/a cow* has the following semantic representation: $\lambda n.\lambda x.[\text{COW}(x) \wedge \text{NU}(\text{COW})(x) = n]$. Intuitively, the quantitative criterion yields what counts as one ‘natural unit’ in the denotation of a given predicate, and is represented by means of NU, standing for a natural unit measure function. Natural unit functions are instances of extensive measure functions and are used to form quantized predicates from cumulative ones.⁸ The qualitative criterion of applicability, which is represented by COW, qualitatively distinguishes cows from, say, cats, dogs and other entities. In contrast, the semantic representation of (concrete) mass nouns only contains the qualitative criterion of application. For example, the semantic representation for *water* is: $\lambda x.[\text{WATER}(x)]$. This amounts to the claim that there is a typical distinction between mass and count nouns, such that only count nouns involve NU a natural unit function. This is motivated by the fact that singular

⁸ $\forall P[\text{QUA}(P) \leftrightarrow \forall x, y[P(x) \wedge P(y) \rightarrow \neg x < y]]$

count nouns, Krifka argues, have quantized reference, whereas mass nouns do not.

The main responses to Krifka's proposal (to be detailed below) have focused on criticisms of his NU function; however something akin to a distinction between qualitative and quantitative criteria remains in most leading accounts (even if the quantitative criterion is often given at a pretheoretical level). The position we will argue for, in line with Grimm (2012), is that a more satisfactory account of individuation requires specifying mereotopological properties.

Zucchi and White (1996, 2001) criticise Krifka's claim that count nouns are semantically quantized. Take, for example, *fence*, *twig*, *line*. They have entities in their denotation whose proper parts also fall under the denotation of *fence*, *twig*, *line*, hence they fail to be quantized. They have a solution in terms of a "maximal participant" relative to situation and a time. On their *Maximal Participant Approach*, determiners such as *a/an* encode a requirement that the entity bound by the existential quantifier is the largest sum individual in the denotation of the V predicate at the event time. On this view, *Alex broke a twig* translates loosely as, *a breaking event whose patient is maximal among the individuals in the denotation of twig broken at the reference time*. Crucially, this does not require that the maximal twig entity is maximal for other events and reference times. The effect, in simple terms, is to make sure that the denotation of the noun is quantized relative to an event and a time.

Whereas Zucchi and White (1996, 2001) emphasise the maximal participant relative to an event and reference time, Rothstein (2010) emphasises that what counts as one varies with *counting context* (Section 3.1). However, on Rothstein's account, what is 'one' is not defined in terms of maximality. Take, for example, fencing around a square field, where what counts as one *fence* need not be the whole enclosure, in each context. Furthermore, *fence* does not denote natural units, since what counts as one varies with context. However, how exactly the set of entities that can count as one are to be delimited remains at a pretheoretical level.

Similarly, as per Landman's account, as we saw in Section 3.2, the only formal restriction on the set of entities that count as one for a predicate is that this set generates the noun's whole denotation. But this means that the criteria deciding the membership of

the set of entities that count as one also remain at a pretheoretical level.

What matters the most for our proposal is that these accounts converge on one key and valuable insight: namely, there is a non-trivial concept ‘what counts as one’ that underlies the mass/count distinction, albeit treated as pre-theoretical.⁹ This insight in fact takes centre stage in Grimm (2012). Grimm argues that mereology is insufficient to define the notion of individual, and that mereology, therefore, must be enriched with topological notions. Mereotopological properties of concrete objects include their part-whole structure, spatial proximity, size, disjointness, adjacency, and shape. Grimm’s *mereotopological* theory uses mereotopological predicates in the lexical entries of nouns. For example, for *dog*:

$$(22) \quad \llbracket \text{dog} \rrbracket := \lambda x_o [R(x_o, \text{Dog}) \wedge \text{MSSC}(x_o)]$$

This states that entities in the denotation of the singular count noun *dog* are *Maximally Strongly Self-Connected* (MSSC). x_o is an object variable (as opposed to a kind variable). MSSC is a mereotopological property. An mereological individual “is Maximally Strongly Self-Connected relative to a property if (i) every (interior) part of the individual is connected to (overlaps) the whole (Strongly Self-Connected) and (ii) anything else which has the same property and overlaps it is once again part of it (Maximality)” (Grimm 2012, p. 135).

Our account takes inspiration from Grimm (2012), but we will connect mereotopological properties more directly to formal accounts of perception. In particular, we will address the problematic data of *granulars* like *rice* and *lentils* and argue that the conceptualisation of mereotopological properties can arise out of more domain general perceptual processes.

Instead, building on the suggestion in Krifka (1989) that the application criteria of nouns consist of both a qualitative and a quantitative criterion, we propose that qualitative criteria involve perceptual

⁹Although, arguably, Chierchia (2010) tries to derive ‘counting as one’ from his supervaluationist semantics, he still assumes a pre-theoretical setting of the ‘ground context’ which, among other things, ensures that nouns such as *mountain* and *fence* have stable atoms. On Chierchia’s (2010) account, different answers to the question ‘How many fences are there?’ is attributed to their being different ground contexts (Chierchia 2010, pp. 122–123).

properties of objects, which subsume Grimm’s mereotopological properties, and functional properties of objects. This is not to say that there are not other properties relevant to individuation, but perceptual and functional properties form the most salient aspects of entities in the denotation of concrete nouns. Here we focus on perceptual properties that concern the spatial organisation entities in the world, and as a case study we take granulars, since granulars present problems for previous theories (Section 3.2). To model this with probM-TTR, we take as a foundation work done by Dobnik *et al.* (2012), because they link spatial knowledge gained by perception with semantic knowledge in a single TTR representation.

4 PROPOSAL: COUNTABILITY AND PROBABILISTIC MERELOGICAL TYPE THEORY WITH RECORDS

4.1 *Probabilistic mereological Type Theory with Records*

Thus far, the structure of objects of basic types has been left unspecified. We assume a domain for physical entities that is structured as a Boolean semi-lattice closed under sum. A part of such a domain is given in Figure 1. As is standard in mereological semantics, we assume the operation \sqcup and the relations $<, \leq$.¹⁰

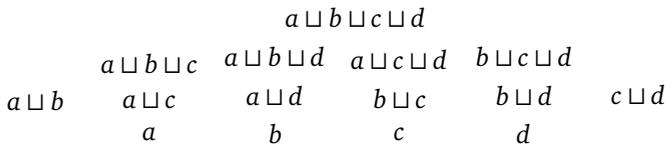


Figure 1:
Boolean semi-lattice
closed under join

This means that, formally, our enrichment of (prob-)TTR regards the structure of the domain. The principal divergence from TTR and prob-TTR is that we do not assume a basic type *Ind*, but instead only a basic type of *Stuff* for physical entities and individuals; i.e. the type for the whole physical domain. In terms of mereological semantics, this is comparable to adopting the approaches of Krifka (1989) and Landman (2016) who assume a domain unspecified for atomicity (non-atomic). This contrasts with Link’s (1983) two domain approach (an atomic

¹⁰Part relation: \leq , where $x \leq y \leftrightarrow x \sqcup y = x$. Proper part relation $<$, where $x < y \leftrightarrow x \leq y \wedge \neg y \leq x$.

domain for count nouns and a non-atomic domain for mass nouns), and also with, for example, Chierchia (2010) and Rothstein (2010) who assume a single atomic domain.

Upward closures of types are defined recursively:

Definition: $*T$ (The upward closure of a type T under sum)

Where **Type** is the set of types:

1. for any $T \in \mathbf{Type}$, $*T \in \mathbf{Type}$
2. for any $T \in \mathbf{Type}$, $a : *T$ iff:
 - (i) $a : T$
 - (ii) or there is some $b, c : *T$ such that $b \sqcup c = a$

For example, if $a, b : T$, then, by (i), $a, b : *T$, and by (ii), $a \sqcup b : *T$.

The advantage of using the tools of probM-TTR is that they will allow us to provide a more nuanced proposal of what it means to be an individual relative to a predicate than those which are found in most mereological approaches to the mass/count distinction. Individuals relative to a predicate are what count as one relative to that predicate (see Section 3.3). TTR provides us with the sufficient tools to combine perceptual, functional, spatial and semantic information within the same representational framework. This will allow us, for example, to show how the same entities can count as a plurality, an aggregate, or even be judged to count as an individual (as one). Such subtle cognitive and perceptual details at the level of our representations allow us to give a formal characterisation of individuation that captures intuitions which are left at the pre-theoretical level in other approaches.

4.2 *Qualitative types*

The qualitative criteria for applying concrete noun concepts will vary greatly from noun to noun not only in values for predicates (like COLOUR, SHAPE, SIZE) used to capture the criteria related to their perceptual properties, or in values for predicates (like USED-FOR-GRINDING) related to their functional properties, but they will also vary with respect to which kinds of criteria are relevant for their application in the first place. Take, for instance, the contrast between natural objects like apples, leaves, trees on the one hand, and artifacts like cars, chairs, buildings, on the other. Whereas *perceptual* properties (from the senses) may be relevant for identifying both natural objects

and artifacts, *functional* properties will play a far bigger role in identifying artifacts. A pile of cushions can count as a chair and a cardboard box can function as a table if, in context, that pile of cushions can aptly function as a chair and the cardboard box can aptly function as a table. In contrast, it is harder to imagine a situation in which some natural object that is not a carrot could count as a carrot even if it fits the same functional role as a carrot does. For instance, even if one uses beetroot or courgette instead of carrot to moisten a cake, one has not, thereby, still made a carrot cake.¹¹

We represent such perceptual and functional properties in terms of an all-encompassing type as the one schematised in (23) and exemplified for *rice* in (24). In this respect, we build on a previous proposal in Sutton and Filip (2016b). We assume a basic type *Stuff* that does not distinguish between substances and individuals. Entities of this type may or may not be a clearly demarcated and countable entity, i.e., an individual in our sense.

$$(23) \quad \left[\begin{array}{l} x \quad : \quad Stuff \\ s_{pptys} \quad : \quad P_{pptys}(x) \end{array} \right]$$

$$(24) \quad \left[\begin{array}{l} x \quad : \quad Stuff \\ s_{rice_{pptys}} \quad : \quad rice_{pptys}(x) \end{array} \right]$$

Instances of the predicate P_{pptys} are placeholders for a wider number of predicates that specify perceptual information such as colour, texture, and, especially for artifacts, functional information (e.g. what activities these items are used for).

Here we wish to expand somewhat on what kind of information the predicate P_{pptys} is a placeholder for, especially with respect to granular nouns such as *rice* and *lentil(s)*. To this end, for the rest of this section we will focus on how qualitative *perceptual* – in particular mereotopological – properties of concrete objects facilitate their classifications under concrete noun predicates (Grimm 2012). We will show how this information can be included in TTR frames, and to this goal, we will also make use of the work done in Dobnik, Cooper and

¹¹ It is quite plausible that our discussion of perceptual and functional properties mirrors distinctions made in Pustejovsky (1995). For example, Pustejovsky's *constitutive* and *formal* roles seem to approximate our perceptual properties and his *telic* and *agentive* roles seem to approximate our functional properties.

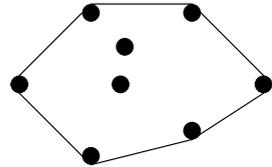
Larsson (Dobnik *et al.* 2012) on the linking of semantic, perceptual and world knowledge in a single TTR representation. Their focus is on the interaction of the inputs from robot perceptual sensors with higher level semantic representations of the robot’s environment.

The robot uses a sensor to build a map of points where each point has been classified as being at a particular location in the robot’s environment (a point map). Points, in this context, are minimal readings that the robot’s sensor makes; the robot builds up a map of its environment by taking point readings. Point maps are represented along the lines of the schema in (25) (Dobnik *et al.* 2012, p. 54).

$$(25) \quad \textit{PointMap} = \left[\begin{array}{l} p_1 \quad : \quad \textit{Point} \\ \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \\ p_n \quad : \quad \textit{Point} \end{array} \right]$$

Such point maps are then used as the inputs to functions to define bounded regions or volumes which envelop points. These are known as convex hulls, i.e., regions or volumes, and are classified as individuals. The convex hull of a set of points is the smallest convex region containing that set of points. A simple representation of a 2D convex hull of points is given in Figure 2.

Figure 2:
Simple example
of a 2D convex hull of 8 points



Formally, this is represented in (26) (Dobnik *et al.* 2012, p. 55).

$$(26) \quad f : \lambda r : \textit{PointMap} \left(\left[\begin{array}{l} a \quad \quad : \textit{Ind} \\ p_1 \quad \quad : r.\textit{Point} \\ \cdot \quad \quad \quad \cdot \\ \cdot \quad \quad \quad \cdot \\ p_n \quad \quad : r.\textit{Point} \\ c_{reg} \quad \quad : \textit{region}(a) \\ c_{inc} \quad \quad : \textit{includes}(a, \langle p_1, \dots, p_n \rangle) \\ \textit{conv-hull} : \langle p_i, p_j, p_k \rangle \\ c_{hulled} \quad : \textit{hulled}(\langle p_1, \dots, p_n \rangle, \textit{conv-hull}) \end{array} \right] \right)$$

It turns out that the above insights from Dobnik *et al.* (2012) allow us to analyse the problematic data of granulars such as *rice*, *lentils*, *peas*, which pose thorny problems for mereological accounts (Section 3).

The basic idea we pursue here is that stuff in the world can be conceptualised in different ways based in part on its perceptual properties. Entities such as, for example, grains of rice can be conceptualised in different ways; this reflects different ways of individuating or otherwise grouping stuff with the relevant rice properties. We highlight three such ways. Granular entities can be (i) individuated in terms of single grains; (ii) grouped in terms of *aggregates* of grains (of some amount or another); (iii) grouped in terms of *bounded aggregates* of grains (portions of grains that form a discrete bounded region or volume in space). Substances such as mud, in contrast, cannot be individuated in terms of anything like grains (mud does not come in clearly perceptible units such as grains). However, stuff like mud can, similar to aggregation, be *amassed* (stuff of some amount with the relevant properties) and conceived of in terms of *bounded amassments* (stuff with the relevant properties that forms some discrete bounded region or volume in space).

We now outline how aggregation and bounded aggregation can be represented in mereological TTR using representations inspired by the work of Dobnik *et al.* (2012). We use rice as a working example. Aggregates with respect to a predicate *rice* involve identifying some plurality of entities each of which has the relevant properties for being grains of rice and judging them to be an aggregate. Unlike Dobnik *et al.*'s convex hulled regions, aggregates need not be grouped into a single discrete region. This is outlined in (27).

$$(27) \quad f : \lambda r : \left[\begin{array}{l} x_1 : Stuff \\ \vdots \\ x_n : Stuff \end{array} \right] \left(\left[\begin{array}{l} c_{rice_agg} : rice_agg(a) \\ r.x_1 : Stuff \\ c_{1_col} : white(r.x_1) \\ c_{1_shape} : grain_shaped(r.x_1) \\ \vdots \\ r.x_n : Stuff \\ c_{n_col} : white(r.x_n) \\ c_{n_shape} : grain_shaped(r.x_n) \\ c_{agg} : aggregate(a) \\ c_{inc} : includes(a, \langle r.x_1, \dots, r.x_n \rangle) \end{array} \right] \right)$$

The predicate *aggregate* is specified as containing some quantity of entities each of which has some relevant properties (such as white colour or being grain-shaped). This collection is then judged as being a rice aggregate ($rice_{agg}$). In other words, we can recast mereological sums in terms of an aggregation of, in this case, entities with the requisite rice-grain properties.

Alternatively, we can add extra restrictions on aggregates by requiring that aggregate entities form ‘hulled regions’. That is to say that we use the notion of a hulled volume or region as a means of representing mereotopological sum entities. Our novel proposal is that something akin to hulling, namely, carving out chunks or regions out of the parts of the world and judging this region to be a *bounded aggregate* (or alternatively a *bounded amassment* for substance denoting nouns) could model a process of individuation that relies on the spatial (mereotopological) properties of concrete objects: namely, properties having to do with their spatial proximity, disjointness, adjacency, size and shape. This proposal not only capitalises on some insights in Grimm (2012), but is also reminiscent of the longstanding proposals in cognitive semantics (Jackendoff 1991; Talmy 2000) which emphasise spatial notions in the analysis of lexicalization patterns of mass and count nouns.

In the frame in (28), we assume that mereotopological sums of bounded entities may be represented via a similar mechanism to ‘region hulling’, namely, aggregating identified portions of stuff each of which have certain physical properties.

$$(28) \quad f:\lambda r: \left[\begin{array}{c} x_1:Stuff \\ \cdot \\ \cdot \\ x_n:Stuff \end{array} \right] \left(\begin{array}{l} c_{rice_agg} : rice_{bounded}(a) \\ r.x_1 : Stuff \\ c_{1_col} : white(r.x_1) \\ c_{1_shape} : grain_shaped(r.x_1) \\ \cdot \\ \cdot \\ r.x_n : Stuff \\ c_{n_col} : white(r.x_n) \\ c_{n_shape} : grain_shaped(r.x_n) \\ c_{reg} : region(a) \\ c_{inc} : includes(a, \langle r.x_1, \dots, r.x_n \rangle) \\ conv_agg : \langle r.x_i, r.x_j, r.x_k, \dots \rangle \\ c_{b_aggd} : agg(\langle r.x_1, \dots, r.x_n \rangle, conv_agg) \end{array} \right)$$

We can perceptually identify the collection of grains of rice as being comprised of individual grains; however, at the same time this collection can be viewed as a bounded entity in a manner akin to hulling. Importantly, just as with hulling collections of perceptual points to identify entities, there will be restrictions on what kinds of collections of entities will be identified as bounded aggregates. One such restriction will be that the entities that form the aggregate cannot be too dispersed and so must be relatively clustered together. Such intuitions also motivate how some of the mereotopological restrictions on granular and collective aggregates in Grimm (2012) can be represented.¹²

The function in (28) mirrors that in (26); however the function is from a record of a type of having more than one physical entity (or bit of stuff), rather than in terms of perceived points. The function determines a bounded aggregate then yields a new entity judged to be of type $rice_{\text{bounded}}$ that ‘collates’ the physical entities in this region. The bits of stuff have properties such as colour and shape. c_{inc} labels a function that selects which of these is to be included in the region. The *conv-agg* tuple determines the entities around which the boundaries of the convex aggregate will be ‘drawn’. c_{b_agg} is the condition that all the entities in the region are within the bounds of the boundary.

So, similarly to defining a convex hull in terms of perceived points, this function defines a *bounded aggregate* in terms of entities that have already been classified as physical entities.¹³ Intuitively, if a situation contains many small entities (such as lentils or grains of rice) that are in close proximity to one another, this function picks them out as a convex aggregate – an aggregate falling within what is judged to be a certain bounded area of space – and then classifies this as $rice_{\text{bounded}}$.

For substance denoting nouns such as *mud* or *blood*, a similar function could be defined. However, instead of aggregating grains into bounded aggregates, it would hull stuff with the relevant properties into a bounded amassment.

¹² Examples of collective aggregates in Grimm (2012) are names for insects (found in swarms or groups) and berries. Examples of granular aggregates are *rice* and *sand*.

¹³ It will be that something akin to (26) is also needed to classify what, in a perceptual field, is to be identified as of the type *Stuff*, albeit at a ‘lower’ level.

4.3 Quantitative functions relative to a predicate

Given a representation of the qualitative properties of an entity or some collection of entities, we may assign some quantity value to the entities of that type. This is the role of a quantitative function, which is of the type in (29), a function from record types specifying qualitative criteria for applicability to real numbers.

$$(29) \quad f_{p_{quant}} : \left(\begin{array}{l} x \quad : \quad Stuff \\ s_{p_{pptys}} \quad : \quad P_{pptys}(x) \end{array} \right) \rightarrow \mathbb{R}$$

Quantitative functions are relative to predicates (different functions are defined for different predicates), since the same entity or entities may count as ‘one’ relative to one predicate, but not another. For example, 52 playing cards could be judged to have a quantity of 1 with respect to a predicate *deck of cards*, but not with respect to the predicate *card* (Link 1983). However, because we are not assuming a pre-theoretical notion of individuation, how some stuff will be quantified may depend on what counts as an individual relative to that predicate. Our strategy is to derive individuation from a special case in which a quantitative function outputs 1. Competing schemas for individuation will be represented as competing quantitative functions. These competing quantitative functions differ with respect to what perceptual and functional properties are required to measure 1 (count as one). For example, take the record type in the right hand side of (27) as compared to the one below in (30).

$$(30) \quad \left[\begin{array}{l} x_1 \quad : \quad Stuff \\ c_{1_col} \quad : \quad white(x_1) \\ c_{1_shape} \quad : \quad grain_shaped(x_1) \end{array} \right]$$

There is more than one possible way to try to individuate the stuff or collections of stuff with rice-like properties (being white, grain shaped etc.). We give three cases by way of example.

Case 1: One possible quantitative function would output a value 1 for the type in (30). This function would individuate single grains. Applied to something like the type in (27), which contains multiple entities with the requisite properties that have been judged to be an aggregate, this function could use, for example, the approximate size of the aggregate to output an approximate quantity value. We do not assume that these functions are mere cardinality functions. In fact, it

would be cognitively implausible to do so. Take the case of some collections of rice grains. For larger collections that have been judged to be aggregates (Section 4.2), we do not assume that the quantity value will reflect the number of rice grains exactly, since we are not in a position to know this without explicitly counting. However, for small numbers of grains, such as an aggregate entity comprised of three grains, this function could return a value where the output number of the function equals the number of grains. Whether the output of the quantitative function reflects the exact number of grains or some approximation, we suggest, could be grounded in the distinction found in psychology between the *approximate number system* (ANS) and the *parallel individuation system* (PI) (Hyde and Spelke 2011, and references therein). In brief, both of these systems are supposed to be developed pre-linguistically. The difference between them is that PI operates accurately in individuating entities, but is severely limited in terms of number. It operates accurately up to about four entities, and is assumed to involve the representation of all entities individually. ANS works on much larger numbers of entities but is assumed to represent entities as collections, not individually. ANS works effectively as a way of discerning differences in number between collections, but not as an accurate representation of cardinality. If the quantitative function is constrained by these systems, its numerical output would be an accurate measure of cardinality of pluralities up to about four entities, but only an approximation of cardinality for larger collections (about 5, about 10, ..., about 50, about 100). What the output of this function is, we suggest, could be modeled in relation to factors such as the size and density of the aggregate identified in the situation. So an output of e.g. 1 would indicate *exactly 1*, but an output of e.g. 10, would indicate *approximately 10*.

Case 2: An alternative quantitative function could individuate, not single grains, but clusters, such that any (sufficiently large) aggregates containing entities that are individually white, grain shaped, etc., would measure 1. This function, applied to the type in (27) would output 1, but applied to the type for a single grain in (30) would measure either a value less than one (or, alternatively, could be undefined). This would allow for the possibility that overlapping collections of rice grains could each be judged to be an aggregate and so each be measured by the quantitative function as one.

Case 3: Another alternative quantitative function could individuate, not single grains or aggregates, but bounded aggregates: any clusters of entities that are individually white, grain shaped etc., but also form a discrete bounded region would measure 1. This function, applied to the type in (28) would output 1. The additional boundedness condition, in effect, treats any discrete regions filled with rice grains as entities to be counted.

These cases represent different ways of individuating rice. The first quantitative function ‘finds’ individual grains, and if more than one is present approximates a quantity. The second function ‘finds’ collections of grains and groups them as an aggregate entity. The third function ‘finds’ bounded regions or clusters of grains and groups them as a bounded aggregate. In Sections 5–7, we will argue that the fact that there can be competing individuation schemas can be used in conjunction with information theoretic requirements, to explain count/mass lexicalization patterns cross- and intralinguistically.

The special case for the application of a quantitative function will therefore be where the output is 1. In the case of *cat* this would indicate a type of individual cats. For granular nouns such as *rice* or *lentils*, this could be the type for individual grains of rice or individual lentils.¹⁴ In this sense, the special case where a quantitative function returns a value of 1 marks the INDIVIDUATION SCHEMA for a predicate. We introduce the following notational convention for the special case to act as both an abbreviation and as an mnemonic for this individuating role:

$$(31) \quad \left[\begin{array}{l} s_{cat-ind} : \left[\begin{array}{l} s_{cat-qual} : [s_{cat-pptys} : cat_{pptys}(x)] \\ f_{cat-quant} : ([s_{cat-pptys} : cat_{pptys}(x)] \rightarrow \mathbb{R}) \\ f_{cat-quant}(s_{cat-qual}) : \mathbb{R}_1 \end{array} \right] \end{array} \right] \\ = [s_{cat-ind} : cat_{Ind}(x)]$$

In other words, the type of situation in which for some predicate *P*, a physical entity (or sum) is judged to have a quantity of ‘one’ is a type of situation in which one judges that thing to be a *P*-individual. To emphasise, this means that we do not take being an individual as a basic notion, but rather as a classification task. We assume a basic

¹⁴We discuss other nouns in detail in Section 7.

type of physical entities (and the upward closure of this type), but which of these (collections of) physical entities are individuals is both relative to a predicate and a non-trivial question.

An important restriction for any P_{Ind} predicate associated with a count noun is that the entities of this type are disjoint (do not overlap mereologically). In other words, an entity that is judged to be ‘one’ P -individual cannot also be judged to be of some larger quantity value with respect to P under the same individuation schema if that individuation schema is to form the basis for grammatical counting.

4.4 *Individuation schemas and learning to apply predicates in context*

It is important to note that there may be cases where individuating, relative to a predicate and a quantitative function, may not always guarantee felicitous application of the predicate. Chierchia (2010) points out that for many mass nouns, whether or not some entity falls under the denotation of that noun can depend on the context. For example, take a collection of around ten grains of rice. In the context of cooking dinner, one can truly say “We have no rice” when the ten grains are all that remains of a once full packet, and a child can felicitously say “I have eaten all the rice” when only ten grains remain on the plate. However, when around ten grains have fallen in the same context, we can felicitously and truthfully state “I spilled some rice on the floor”. Stating “There is rice in this dish” is felicitous and truthful, when uttered by someone with a severe rice allergy, for example, even if it contains only about ten grains of rice.

Assuming a quantitative function (labeled $f_{P_{quant}}$) for a predicate P , this kind of context sensitivity can be represented as the calculation of conditional probabilities of the form in (32). The learner-agent A must identify which qualities (specified in the qualitative criterion type) and which quantities of these entities (relative to an individuation schema) maximise the probability of applying the predicate relative to her judgement set \mathfrak{J} .

$$(32) \quad P_{\mathfrak{J},A}(r : \left[\begin{array}{l} x : Stuff \\ s_P : P(x) \end{array} \right] \mid r : \left[\begin{array}{l} s_{P_{qual}} : \left[\begin{array}{l} \text{qualitative} \\ \text{criterion type} \end{array} \right] \\ f_{P_{quant}} : \left(\left[\begin{array}{l} \text{qualitative} \\ \text{criterion type} \end{array} \right] \right) \rightarrow \mathbb{N} \\ i : \mathbb{N} \\ f_{P_{quant}}(s_{P_{qual}}) : \mathbb{N}_i \end{array} \right])$$

We assume the data for these judgements come from, in part, witnessing competent speakers' judgements with regard to applying predicates. We now give a simple example for *rice*.

Suppose that a learner is exposed to two situations in which an adult speaker provides her with evidence for how to make *rice* judgements. Furthermore, the agent is employing a schema/quantitative function that individuates single grains. In $s_{\text{making dinner}}$, there are, by the agents' estimations, approximately 10 grains (say at the bottom of a packet). The adult speaker may say, looking at the packet, *Oh no! We have no rice left*. This constitutes evidence that the small quantity of around 10 grains is not sufficient to count as *rice*. In s_{allergy} , a similar quantity of grains falls into the soup. The soup is for someone with a rice allergy and the adult speaker says *Oh no! Rice fell into the soup*. This constitutes evidence that the small quantity of around 10 grains is sufficient to count as *rice*.

In terms of learning data, these utterances in context provide conflicting information as to whether or not a collection of around ten grains counts as *rice*. For the case we just informally described, this could result in a judgement set containing the probabilistic Austinian propositions in Figure 3. The figure itself contains judgements for different situations (labelled *sit*). These situations are meant to represent the contexts just described. The idea is that the quantity of grains

Figure 3:
Possible (partial)
judgement set for around
10 grains of rice.

$$\begin{aligned}
 j_1 = & \left[\begin{array}{l} \text{sit} \quad =s_{\text{making dinner}} \\ \text{sit-type} = \left[\begin{array}{l} x \quad : \text{Stuff} \\ s_{\text{rice}} \quad : \neg \text{rice}(x) \\ s_{\text{rice}_{\text{qual}}} \quad : \left[\begin{array}{l} s_{\text{rice.col}} : \text{white}(x) \\ \dots \end{array} \right] \\ f_{\text{rice}_{\text{quant}}}(s_{\text{rice}_{\text{qual}}}) : \mathbb{N}_{10} \end{array} \right] \\ \text{prob} \quad = 0.9 \end{array} \right] \\
 j_2 = & \left[\begin{array}{l} \text{sit} \quad =s_{\text{allergy}} \\ \text{sit-type} = \left[\begin{array}{l} x \quad : \text{Stuff} \\ s_{\text{rice}} \quad : \text{rice}(x) \\ s_{\text{rice}_{\text{qual}}} \quad : \left[\begin{array}{l} s_{\text{rice.col}} : \text{white}(x) \\ \dots \end{array} \right] \\ f_{\text{rice}_{\text{quant}}}(s_{\text{rice}_{\text{qual}}}) : \mathbb{N}_{10} \end{array} \right] \\ \text{prob} \quad = 0.9 \end{array} \right]
 \end{aligned}$$

to count as *rice* could vary across these contexts. The situation types assign a measure to stuff with the relevant rice properties (such as colour) and a condition that the stuff is rice. The probability values represent the extent to which the situations are of that type.

We assume that the learner has learned these judgements directly from the competent speaker and so attributed a high value (0.9) to all of them.¹⁵ Using the prob-TTR version of Bayes’ rule (15), these judgements then allow the calculation of the probability of something being rice, given that it has a quantitative function value (relative to *rice*) of 10. This is shown in (33).¹⁶

$$\begin{aligned}
 (33) \quad & p_{\exists A}(s : [s_{rice} : rice(x)] \mid s : \left[\begin{array}{c} s_{rice_{qual}} \\ \vdots \\ s_{rice_{col}} : white(x) \end{array} \right] \\
 & \left[\begin{array}{c} s_{rice_{qual}} \\ f_{rice_{quant}}(s_{rice_{qual}}) : \mathbb{N}_{10} \end{array} \right] \\
 & = \frac{\left\| \left[s_{rice} : rice(x) \right] \wedge \left[\begin{array}{c} s_{rice_{qual}} \\ \vdots \\ s_{rice_{col}} : white(x) \end{array} \right] \right\|_{\exists}}{\left\| \left[\begin{array}{c} s_{rice_{qual}} \\ \vdots \\ s_{rice_{col}} : white(x) \end{array} \right] \right\|_{\exists}} \\
 & = \frac{0.9}{0.9 + 0.9} \\
 & = 0.5
 \end{aligned}$$

Given the judgements in Figure 3, the result is 0.5. A gloss on the importance of this value is that the learner has as much reason to believe that the predicate *rice* can be applied to around 10 grains of rice as she does for not thinking so, given her judgement set. By “around 10”, we mean that the output of the quantitative function may only approximate actual numbers of grains for quantities above that which can be directly quantified via the PI (parallel individuation) cognitive system (Section 4.3).

¹⁵In a more sophisticated model, reflected in the probability value should be that the evidence for $s_{making\ dinner}$ is indirect (making $s_{making\ dinner} : \neg rice$ less certain). A promising route would be to adopt something akin to Lassiter’s representation of indirect evidence in terms of Bayesian networks (Lassiter 2016).

¹⁶We assume, following Cooper *et al.* (2015), that negation is classical. Instances of $[x : Stuff]$ have been suppressed for brevity.

However, across contexts in which there are much larger quantities of grains present, the judgement set may be far more consistent. For example, a learner will rarely experience a whole packet of rice not being judged as rice, i.e., not falling under the predicate *rice*, and so is far more confident about classifying larger amounts of rice as *rice*. In short, in this way we capture the observation by Chierchia (2010) that when it comes to classifying with granular nouns like *rice*, quantity matters. Independently of context, one may not be safe to classify ten grains as *rice*, but one could, with high confidence, classify a packet of rice as *rice*.

This means that, although an individuation function for *rice* that identifies single grains does succeed in identifying disjoint (potentially countable) entities, it is not wholly reliable when applied across contexts to establish, with a high degree of certainty, when to apply the predicate *rice*. A more reliable schema could be found by opting for an individuation schema that picks out larger collections of rice grains. Such a move could end up failing to properly individuate disjoint entities suitable for counting, however (since larger collections overlap). In Section 5, we will show how this tension can be formally captured within probM-TTR.

5 THE LEARNING PRESSURES OF RELIABILITY AND INDIVIDUATION

Given the insights of formal (mereological) theories (Section 3), we propose that in addition to identifying a reliable criterion for applying a predicate, learners also seek to identify an individuation schema for a predicate. This means that pressures on nominal predicate learning will be at least twofold:

- (i) RELIABILITY: to establish with a high degree of certainty when to apply a predicate.
- (ii) INDIVIDUATION: to establish (if possible) an individuation schema for a predicate.

In some cases, these two pressures will operate in unison, for example, for *cat*, accurately judging a situation to contain one or more cat-individuals is a very good ground to judge those entities as falling under the number neutral predicate *cat*. However, as we shall argue,

this is not always the case for other predicates. For example, individual rice grains are the clearly individuable units for counting rice, but the presence of a single grain is not a reliable criterion for applying *rice* since there are many contexts in which a single grain is not a sufficient quantity to count as *rice*. Furthermore, we argue in Section 7 that tensions between these two pressures generate exactly the cases where we find cross- and intralinguistic mass/count variation.

5.1 *Formalising the requirement of reliability*

Reliability is a pressure on a learner to find a set of properties that reliably predict when to apply a predicate. We have proposed that these properties include both qualitative and quantitative criteria. Reliability itself is therefore a balance between using a P_{Ind} predicate, the upward closure of which ($*P_{Ind}$) includes too much in P ; and using a P_{Ind} predicate, the upward closure of which ($*P_{Ind}$) includes too little in P . Of course, the ideal balance means using a P_{Ind} predicate that neither includes too much nor too little in P . In other words, the most reliable individuating predicate will be one that maximises the conditional probabilities in (34) and (35).¹⁷

$$(34) \quad \text{Max}_j(p(s : [s_P : P(x)] \mid s : [s_{P-ind} : *P_{Ind_j}(x)]))$$

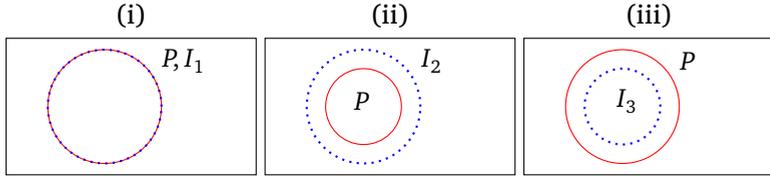
$$(35) \quad \text{Max}_j(p(s : [s_{P-ind} : *P_{Ind_j}(x)] \mid s : [s_P : P(x)]))$$

Maximising the probability in (34) means that being a P -individual or a sum of P -individuals is a very strong indicator of being a P . This militates against the over-inclusivity of P_{Ind_j} . Maximising the probability in (35) means that being a P is a very strong indicator of being a P -individual or a sum of P -individuals. This militates against the under-inclusivity of P_{Ind_j} . Balancing these two (optimally maximising both probabilities) should result in as close an approximation of P and $*P_{Ind}$ as possible. In the trivial case, this would just be to use P as P_{Ind} . However, in most cases, doing this would fail to individuate any entities at all.

To make this clearer, take the three simple cases which are graphically represented in Figure 4. (i) This represents the case where the application conditions for the predicate are perfectly matched to the application conditions for the upward closure of the P_{Ind} predicate,

¹⁷ The specifications of $x : \text{Stuff}$ here and further below are omitted for brevity.

Figure 4:
Maximising both
conditional
probabilities vs.
maximising one.



$$P = [s_P : P(x)] \text{ (solid line) and}$$

$$I_j = [s_{P-ind} : *P_{Ind_j}(x)] \text{ (dotted line)}$$

therefore both conditional probabilities (34) and (35) are maximised. (ii) This represents the case where there are some things which are P individuals or sums thereof that are not correctly judged as P . This means the the probability in (35) is maximised, but the probability in (34) is not. (iii) This represents the case where there are some things which are correctly judged as P which are not P individuals or sums thereof. This means that the probability in (34) is maximised, but the probability in (35) is not.

5.2 Formalising the requirement of individuation

The pressure that can push in the opposite direction to reliability is *individuation*. This pressure can be derived from a more general pressure towards informativeness (Piantadosi *et al.* 2011). The main idea in the context of countability is that disjoint individuation schemas/predicates P_{Ind} have minimum entropy with respect to determining counting results compared with predicates that are not disjoint. The reason for this, building on Landman’s (2011) insights, is that when we have an overlapping set of entities, there are multiple answers to the question ‘how many?’. Uncertainty over how many things (relative to a predicate) there are equates to a higher level of entropy compared with a single answer.

In order to formally capture the pressure towards individuation, we will define a probabilistic notion of disjointness of a type, and then relate this to minimising entropy (with respect to the disjoint variants of a type). The (probabilistic) notion of disjointness which will be used below is a condition for the maximal individuation of P_{Ind} predicates. This follows the standard mereological notion of disjointness, but adds the condition that the only relevant entities are those that are of the

relevant type with sufficient amounts of certainty. We formalise ‘sufficient degree of certainty’ using a probability threshold θ .

A type T is mereologically pairwise disjoint relative to a probability threshold θ iff:

$$\forall x, y [(p(x : T) \geq \theta \wedge p(y : T) \geq \theta) \rightarrow \neg \exists z [z \leq x \wedge z \leq y]]$$

In words, any two entities, taken pairwise, that are of a type with a probability above the threshold cannot share a part with one another.

Disjoint types have only one maximally disjoint subtype (akin to variants in Landman (2011)), namely the type itself. For types that are not disjoint, one can form, possibly multiple, maximally disjoint subtypes. For example, if $a, b, a \sqcup b : T$, then, relative to $a, b, a \sqcup b$, there are two maximally disjoint subtypes v_1 and v_2 such that $a, b : v_1$ and $a \sqcup b : v_2$, but where $a \sqcup b \not\leq v_1$ and $a, b \not\leq v_2$.

The pressure of individuation can be modelled as pushing towards the use of a disjoint P_{Ind} type. At a first pass, we could, therefore, suggest that the pressure of individuation is a requirement merely to minimise entropy as in (36).

$$(36) \quad \text{Min}_j \left(- \sum_{v_i \in V} p(v_i | P_{Ind_j}) \times \log p(v_i | P_{Ind_j}) \right)$$

Here, entropy values give the average amount of information needed to determine a specific counting result. For example, assuming an equal distribution over variants, and a base-2 logarithm, numbers of variants and entropy values would be as follows:¹⁸

Number of variants	1	2	4	8	16
Entropy	0	1	2	3	4

The effect is that minimising entropy pushes towards a disjoint P_{Ind} predicate because disjoint predicates have an entropy of 0.

However, the definition in (36) misses some details. As we have seen, there are nouns such as *fence*, *twig*, *line* which display context-sensitivity with respect to what counts as a single individual (focused on by Zucchi and White (1996, 2001) and Rothstein (2010)). If the

¹⁸ For example, if there are four variants such that $p(v_1 | P_{Ind_j}) = p(v_2 | P_{Ind_j}) = p(v_3 | P_{Ind_j}) = p(v_4 | P_{Ind_j}) = 0.25$, then the surprisal for each variant equals $0.25 \times \log_2 0.25 = -0.5$ and entropy equals $-(-0.5 + -0.5 + -0.5 + -0.5) = 2$.

Rothstein-type analysis for such nouns is correct, then, at every (default) context, there will only be a disjoint set of fence entities, even if across contexts, these entities overlap. The denotations of prototypical count nouns, such as *cat* and *chair*, intuitively, have inherently individuated denotations, unlike the denotations of count nouns such as *fence* that require context to identify countable entities in their denotation. We should, therefore, include, in the calculation for minimising entropy, some cost C that increases entropy in relation to the number of admissible (disjoint) P_{Ind} predicates. This is given in (37).

$$(37) \quad \text{Min}_j \left(- \left(\sum_{v_i \in V} p(v_i | P_{Ind_j}) \times \log p(v_i | P_{Ind_j}) \right) + C \right)$$

To give an example, let us compare three cases. Context general P_{Ind} predicates are predicates that can be applied to correctly individuate P s across all contexts.

(i) A context general, disjoint P_{Ind} predicate. This will apply to nouns with individuation schemas that pick out naturally atomic, clearly disjoint objects. In this case, there is only one variant, P_{Ind} itself, so $\log p(v_i | P_{Ind_j}) = 0$, and there is only one P_{Ind} predicate, so $C = 0$. Applying (37) gives an entropy value of 0.

(ii) A context general, not-disjoint P_{Ind} predicate with, for instance, four variants. This will apply to nouns such as *furniture*, which, following the analysis in Landman (2011), have denotations in which what counts as one overlaps in the same context. If the four variants are equally probable, conditional on P_{Ind} , this would make $\sum_{v_i \in V} p(v_i | P_{Ind_j}) \times \log p(v_i | P_{Ind_j}) = 2$. There is only one P_{Ind} predicate, so $C = 0$. The total entropy value will thus be 2 (see footnote 18).

(iii) A collection of two disjoint P_{Ind} predicates that each only apply in specific contexts. This will apply to nouns, such as *fence* and *huonekalu* ('furniture', Finnish, count), where what counts as one varies from context to context. Each context-specific P_{Ind} predicate is disjoint and has only one variant (itself), so $\sum_{v_i \in V} p(v_i | P_{Ind_j}) \times \log p(v_i | P_{Ind_j}) = 0$ for each j . However, given that there are, in this simple example, two P_{Ind} predicates (one for some contexts, the other for the other contexts), one of which must be chosen in each context, there is some cost added. The final result will be the value of C .¹⁹

¹⁹ Arguably, C should itself be sensitive to the probability that a context selects

On its own, individuation pushes towards finding a single disjoint P_{Ind} predicate and sticking with it across contexts. However, if individuation were the only factor, then any arbitrary disjoint P_{Ind} predicate would suffice as an individuation schema, since there would be no requirement for this schema to reliably identify Ps . Clearly, the upward closure of this predicate type should predict, with reasonable certainty, when to apply P . This is what the competing pressure of *reliability* ensures.

5.3 *Summary of reliability and individuation*

Considered independently, the pressures of *individuation* and *reliability* are each insufficient to capture a good criterion for applying a predicate P . Reliability alone would not ensure adequate informativeness for individuation (entropy would be high, on the assumption made here that disjoint individuation schemas/predicates P_{Ind} have minimum entropy compared to predicates that are not disjoint). Individuation on its own would not ensure adequate reliability (what counts as one P in some contexts is not a reliable indicator for what counts as one P in another, hence a single individuation schema will not reliably indicate what counts as P across contexts).

In Sections 6–7, we will see that some nouns allow a ready balance between these pressures. These nouns will turn out to be those that are fairly stably lexicalized as count cross-linguistically e.g. *cat* and *chair*. Other nouns will exemplify how these two pressures can be in direct conflict. Resolution of this conflict can only come by prioritising one pressure or the other. In these cases, the count/mass encoding of the noun will reflect which pressure is prioritised, and most importantly, as we argue, predicts the variation in mass/count encoding across different languages (e.g., the mass noun *furniture* and the Finnish count noun *huonekalu* ('furniture')) as well as within a particular language. In yet a third case, we will argue that individuation cannot really be satisfied. This case motivates the 'stubborn' encoding of nouns as mass, as we find with prototypical mass nouns, such as *mud* and *blood*.

a particular P_{Ind} predicate. For example, there should, reasonably, be a lower cost for a situation where the same individuation schema is selected 99 percent of the time, versus the situation in which two schemas are equally probable. We leave inclusion of such factors to further work.

6 THE SEMANTICS OF CONCRETE NOUNS IN
PROBM-TTR

Some recent theories of the mass/count distinction propose to represent lexical entries of nouns as pairs where one projection of the pair determines the standard denotation of a noun and the other determines the counting base and/or individuation schema to apply to that noun (Rothstein 2010; Landman 2011, 2016; Sutton and Filip 2016a). In keeping with this general idea, we represent lexical entries of common nouns as a complex frame in which there are two record types.

One specifies the situation type for the predicate being learned, for instance, $\text{cat}(x)$, which is a number neutral predicate a learner associates with situations in which competent speakers make judgements that something is a cat or that some things are cats. We label this s_{pred} . The learner simultaneously learns how such number neutral predicates correspond to the perceptual and functional properties of the objects they witness competent speakers referring to.

Representations of such perceptual and functional properties are encoded in the other part of the lexical entry which we label s_{c_base} . This specifies the counting base for this predicate and includes both the quantitative and qualitative criteria of application for the number neutral predicate in the sense of Krifka (1989). The disjointness of this type is what enters into calculations regarding individuation (Section 5.2). The upward closure of this type is what enters into the reliability calculations (Section 5.1). The motivation for this bipartite lexical structure is that a learner requires both kinds of information in order to learn how to use natural language predicates such as *cat*.

A schema for a noun entry is given in (38). In the frame-based representation offered by TTR, these two record types feature as parts of the same complex type and so can be abstracted over so as to receive the same values when applied to a record containing some physical entity (e.g. an entity such as *felix* labelled x in the record r in (38)).

$$(38) \lambda r : [x : \textit{Stuff}]. p\left(\begin{array}{l} s_{pred} : \left[\begin{array}{l} \text{Rec. type for predicate} \\ (\text{contains label } r.x) \end{array} \right] \\ s_{c_base} : \left[\begin{array}{l} \text{Rec. type for counting base} \\ (\text{contains label } r.x) \end{array} \right] \end{array} \right)$$

We also adopt the basic structure for a common noun lexical entry in prob-TTR given by Cooper *et al.* (2015)²⁰ for which the result of applying a record of the requisite type is the probability of the relevant record type.

The simplest case is that of prototypical count nouns such as *woman*, *cat*, *chair*, *car*. They are associated with a record type that includes the relevant number neutral predicate, and a record type for the counting base which contains the relevant P_{Ind} predicate. For example, the entry for *cat* is given in (39), and the entry for *cats* is given in (40).

$$(39) \quad [[cat]] = \lambda r : [x : Stuff].p\left(\begin{array}{l} s_{pred} : [s_{cat} : cat(r.x)] \\ s_{c_base} : [s_{cat_{Ind}} : cat_{Ind}(r.x)] \end{array} \right)$$

$$(40) \quad [[cats]] = \lambda r : [x : Stuff].p\left(\begin{array}{l} s_{pred} : [s_{cat} : cat(r.x)] \\ s_{c_base} : [s_{cat_{Ind*}} : *cat_{Ind}(r.x)] \end{array} \right)$$

This structure for the semantics of concrete nouns also encodes our conception of the semantic learning of these nouns as being guided by the establishment, if possible, of a counting base, namely the type labeled s_{c_base} which may then serve as reliable criteria for applying the type labelled s_{pred} . For cases such as *cat*, this is relatively straightforward, because being of the type of a cat individual is a reliable criterion for applying the number neutral predicate *cat* (39).

Reliability: On the assumption that a suitably accurate individuation schema for cats can be found, the upward closure of this predicate will be a highly reliable indicator of when to use the predicate *cat*. That is to say, there are very few instances of things having the requisite properties of being a cat individual or of a cat sum for which the judgement $cat(x)$ would be inappropriate (recall that we assume that the predicate type $cat(x)$ is number neutral). Likewise, there are relatively few mispredications (relatively few *cat* judgements made by competent speakers to refer to entities without the properties of being cat individuals (or sums thereof)). Hence, such a cat_{Ind} predicate would yield high conditional probabilities of the sort in (41) and (42).

$$(41) \quad p(r : [s_{cat} : cat(x)] \mid r : [s_{cat-ind} : *cat_{Ind}(x)]) = \text{high}$$

$$(42) \quad p(r : [s_{cat-ind} : *cat_{Ind}(x)] \mid r : [s_{cat} : cat(x)]) = \text{high}$$

²⁰Cooper *et al.* use pTTR.

Individuation: Furthermore, whatever schema yields the highest balance of these probabilities will be a predicate type that is disjoint (it will be the type for individual cats). This means that it will satisfy the entropy minimisation requirement as in (43).²¹

$$(43) \quad - \sum_{v_i \in V} p(v_i | \text{cat}_{md}) \times \log p(v_i | \text{cat}_{md}) = 0$$

With reliability and individuation acting in unison, the lexical entry in (39) is predicted. It has a disjoint counting base, which constitutes a part of the lexical entry for a count noun. This account further predicts that lexicalisations for the *cat* predicate will be stably count.

However, there are nouns for which it is less straightforward to establish a counting base, because the learning pressures of reliability and individuation conflict. In these cases, we argue, there are two ways of conceptualising a referent, one in which individuation is paramount, in which case the result is a count noun, and another in which reliability is paramount, in which case the result is a mass noun. We connect the source of this conflict between reliability and individuation to the types of context sensitivity that play a key role in the theories of the mass/count distinction in Chierchia (2010), Rothstein (2010) and Landman (2011), which we discussed in Section 3. In the next section, we offer a probM-TTR proposal of how context sensitivity can impact the weighting of individuation and reliability in generating predictions about variation in mass/count lexicalization patterns.

7 MASS/COUNT VARIATION: THE EFFECTS OF CONTEXT SENSITIVITY ON INDIVIDUATION AND RELIABILITY

Contemporary mereological theories of the mass/count distinction converge on the idea that the concepts on which the distinction is based are context-sensitive in one way or another. However, the proposals differ with respect to the degree and nature of the relevant context-sensitivity. In particular, Rothstein (2010) and Landman (2011) emphasise disjointness and overlap, which we argue can be interpreted as generating a conflict between the learning pressures of

²¹We omit the cost C when $C = 0$.

reliability and individuation (or, alternately, as generating competing requirements on the nature of lexical predicate’s meaning). Another source of conflict between the learning pressures of reliability and individuation stems from what Chierchia (2010) calls ‘vagueness’. However, we suggest this would be more aptly considered a form of context-sensitivity (in a sense we explain below).

Crucially, we argue that the way these conflicts are resolved tracks differences in count/mass lexicalisation of nouns, and hence the specific resolution strategies could serve as a motivation for the lexicalisation of nouns as mass or count, both within a particular language and crosslinguistically.

Let us first consider the notion of disjointness. As in Landman (2011), we assume that there is a grammatical counting function which is sensitive to disjointness. In our account, the counting function is of the type in (44) and applies to the record type in a lexical entry labeled s_{c_base} , which captures the idea that what is counted are the entities of the type in the counting base. Hence, for a counting function f_{count} and probability threshold θ , we propose a type restriction:

$$(44) \quad f_{count,\theta} : (RecType \wedge Disj_{\theta} \rightarrow \mathbb{R})$$

This type restriction means that the counting function is only defined for types that are disjoint relative to some probability threshold and outputs a real number.

The notion of context used in Rothstein (2010) is that of a ‘counting context’. Counting contexts are subsets of the domain, i.e., a set of entities that count as atoms, as one, in a particular context, which are then intersected with the root noun denotation of a noun to form a disjoint set, in ‘default’ cases. As a formal device, this representation has the right motivation and effect for nouns such as *fence*. However, there are two weak points of Rothstein’s account that we improve upon. First, both counting contexts and individuation remain at a pre-theoretical level in Rothstein’s account. Second, Rothstein (2010) effectively subsumes prototypical count nouns like *cat* as a special case of context-sensitive count nouns like *fence*, for which there just happens to be no variation in what counts as one across contexts. Intuitively, what is one cat is stable across all contexts, but what is one fence varies with context. Our formalism improves on Rothstein’s idea of counting contexts by modelling them as contexts in which some in-

dividuation schema is used, where the different individuation schemas that are licensed in a given context are grounded in our account of semantic learning. As a result, we can explain and better motivate why there is only one licensed individuation schema for nouns such as *cat*, but multiple such schemas for nouns such as *fence*.

7.1 *Mass/count variation in collective artifacts*

The term ‘collective artifacts’ here refers to nouns which denote collections of entities like, for example, furniture and kitchenware. The main point of this section is to show that the way in which nouns for collective artifacts denote in context allows for two distinct ways of forming concepts. One is compatible with counting, and the other is not. Therefore, we should expect to find variation in mass/count encoding in this class. A good example is the mass noun *furniture* in English and the count noun *huonekalu*-*t*_{+C,PL} (‘furniture’) in Finnish. What ultimately motivates whether a noun, say, *furniture* is mass in English, and not count, may well be wholly conventional (guided by, for example, etymological factors). Here, we focus on the issue of variation in mass/count encoding, which is separate from the question why, say, *furniture* is mass in English, which we leave aside here.

When it comes to collective artifacts, we propose that it is reasonable to assume that what counts as one item of furniture will largely be derived from the function of the relevant item and to a lesser degree from its perceptual qualities. A vanity, formed of a mirror and a table, has a joint function *qua* item of furniture, so plausibly counts as a single item. However, the mirror and the dressing table each have its own function and each can stand and be used as an individual item of furniture in its own right. Likewise, we see similar patterns crosslinguistically, even when collective artifacts are lexicalized as count nouns. For example, what counts as one for the count noun *huonekalu* (‘furniture’, Finnish) varies with context in the same way as for the English *furniture*.

The kind of contextual variation that we observe with respect to what counts as one for *furniture* or for *huonekalu* presents a learning challenge to our basic picture of learning an individuation schema for a given noun predicate. Recall that individuation schemas are modelled as the type for which a quantitative function outputs 1. They apply to functionally – and also perceptually – characterised situation

types (of the qualitative criterion type) and are used to individuate the entities within a situation (record) relative to a predicate. A learning challenge arises because the same entity, the vanity, plausibly counts as one item of furniture in one schema and as two items of furniture in another. But that means that no single function should be able to output both of these results for a situation (record) containing a vanity. Put simply, witnessing the same kind of item being treated as one thing in one situation, and two (or more) things in other situations, is evidence that there is more than one felicitous individuation schema for *furniture*.

In other words, we have a case such as the one outlined in Section 5, in which a learner has evidence for multiple context-specific individuation schemas. This generates a conflict between the two learning pressures of individuation and reliability. Take *furniture*, for instance. Opting for a single individuation schema would keep stable what is individuated as one item of furniture in every situation, but it is not a reliable way of individuating, since one single schema will wrongly individuate in some situations. For example, a schema that individuates a table and a mirror as two entities will be incorrect in circumstances where they should count as one item of furniture, a vanity.

Faced with this challenge, a learner has two strategies available: namely, either to learn to apply a different schema depending on the situation, or to form a single complex join type based on all licensed schemas. For example, if P_{Ind_i} is a licensed individuation schema, a more generally applicable type would be the join type formed of all such schemas. This is shown in (45).

$$(45) \quad P_{Ind_{join}} = P_{Ind_1} \vee P_{Ind_2} \vee \dots \vee P_{Ind_n}$$

The kind of context sensitivity that affects what counts as one for collective artifacts gives rise to two alternative ways of encoding the semantics of a *furniture*-like noun. We now detail these with respect to the formal characterisations of reliability and individuation.

First, one can opt for a complex join schema, and so have a reliable indicator of when to apply the noun predicate in most (possibly all) contexts. This is indicated by the high conditional probabilities in (46) and (47). Given an individuating predicate that is the join of all contextually specific ones, the upward closure of this predicate will closely track how one should apply the predicate *furn*.

$$(46) \quad p(r : [s_{furn} : furn(x)] | r : [s_{furn-ind} : *furn_{Ind_{join}}(x)]) = \text{high}$$

$$(47) \quad p(r : [s_{furn-ind} : *furn_{Ind_{join}}(x)] | r : [s_{furn} : furn(x)]) = \text{high}$$

Obviously, an individuating predicate that is the join of all contextually specific ones cannot individuate or yield a disjoint type which can support determinate counting results in a specific situation, due to overlaps of the individuals picked by different individuation schemas of that join individuating predicate. This captures Landman's (2011) intuition that overlapping entities can "simultaneously in the same context" count as one. As shown in (48), the fact that there are a number of different ways to resolve overlap in respect to what counts as one item of furniture leads to a comparatively high level of entropy.

$$(48) \quad - \sum_{v_i \in V} p(v_i | furn_{Ind_{join}}) \times \log p(v_i | furn_{Ind_{join}}) = \text{high}$$

Second, one can make the selection of one's individuation schema context sensitive, i.e. apply individuation schemas that may vary from situation to situation. However, as shown in (49) and (50), this has a negative effect on reliability. Although the probability of *furn* is high given the upward closure of any specific individuating predicate $furn_{Ind_{c_i}}$, the inverse conditional probability is lower than in the complex join individuation schema, because most particular schemas will exclude those bits of furniture that are parts of that which count as one under a different individuation schema. For example, if an agent has an individuation schema that classifies a vanity (table and mirror) as one item of furniture, then this will exclude its parts from counting as one. However, this means that the probability of applying this particular schema given a *furn* judgement will be lower than for the context general join-type case, since the context specific schema (for vanity) will not be reliable for situations in which the table and mirror should count as two items of furniture. This lowers the conditional probability in (50).

$$(49) \quad p(r : [s_{furn} : furn(x)] | r : [s_{furn-ind} : *furn_{Ind_{c_i}}(x)]) = \text{high}$$

$$(50) \quad p(r : [s_{furn-ind} : *furn_{Ind_{c_i}}(x)] | r : [s_{furn} : furn(x)]) = \text{lowish}$$

However, the entropy value is arguably lower than in the single join schema case. That is, in the case of a single join schema, every specific schema will yield an entropy value of 0, since each one is a disjoint predicate, but since there are many such schemas, the task of determining the right one in context incurs a cost C (see case (iii) in Section 5.2). This is shown in (51). Provided that the cost value is lower than the entropy value for the join type in (48), the specific case will fare better at minimising entropy (maximising individuation).

$$(51) \quad -\left(\sum_{v_j \in V} \sum_{c_i \in C} p(v_j | \text{furn}_{\text{Ind}_{c_i}}) \times \log p(v_j | \text{furn}_{\text{Ind}_{c_i}})\right) + C = C$$

In summary, if one tries to maximise reliability, the context general join-type individuation schema wins out over adopting a number of context specific ones. However, if one minimises entropy, selecting a context specific schema wins out over the context general join-type individuation schema.

This creates a tension. One simple outcome is merely to prioritise either reliability (and thereby encode a context general join-type individuation schema), or prioritise individuation (and thereby encode a context specific schema). Choosing the former strategy results in a counting base that is not disjoint. This means, as per our Landman-inspired account of the mass-count distinction, that the resulting lexical entry is one for a mass noun. Choosing the latter strategy results in a counting base that is disjoint. This means that the resulting lexical entry is one for a count noun, as in the case of the Finnish singular count noun *huonekalu*. The results of these two strategies are given in (52) and (53).

$$(52) \quad \llbracket \text{furniture} \rrbracket^{c_i} = \lambda r : [x : \text{Stuff}] . p\left(\begin{array}{l} s_{\text{pred}} : [s_{\text{furn}} : \text{furn}(r.x)] \\ s_{c_base} : [s_{\text{furn}_{\text{Ind}_{\text{join}}}} : \text{furn}_{\text{Ind}_{\text{join}}}(r.x)] \end{array}\right)$$

$$(53) \quad \llbracket \text{huonekalu} \rrbracket^{c_i} = \lambda r : [x : \text{Stuff}] . p\left(\begin{array}{l} s_{\text{pred}} : [s_{\text{furn}} : \text{furn}(r.x)] \\ s_{c_base} : [s_{\text{furn}_{\text{Ind}_{c_i}}}} : \text{furn}_{\text{Ind}_{c_i}}(r.x)] \end{array}\right)$$

The lexical entry for *furniture* in (52) has a generalized schema as the base, but this will not yield a noun suitable for counting, since $\text{furn}_{\text{Ind}_{\text{join}}}$ will not be a disjoint type (both the vanity and the mirror and table that comprise the vanity will be of this type). However, the lexical entry for *huonekalu* in (53) will yield a count noun, since every single individuation schema $\text{furn}_{\text{Ind}_{c_i}}$ will be disjoint.

7.2 *Mass/count variation in non-bounded objects*

The explanation we have just used to motivate the variation in the mass/count encoding of collective artifacts across different languages (e.g., *furniture* (mass) versus *huonekalu* (count) ‘furniture’, Finnish) can also be applied to motivate the intralinguistic variation exhibited by pairs such as *fence* (count) and *fencing* (mass), which constitute a well-defined semantic subclass we dub here ‘non-bounded objects.’ The entry for *fence* is given in (54) and the entry for *fencing* is given in (55).

$$(54) \quad \llbracket \text{fence} \rrbracket^{c_i} = \lambda r : [x : \text{Stuff}] . p \left(\begin{array}{l} s_{pred} : [s_{fence} : \text{fence}(r.x)] \\ s_{c_base} : [s_{fence_{ind}} : \text{fence}_{ind_{c_i}}(r.x)] \end{array} \right)$$

$$(55) \quad \llbracket \text{fencing} \rrbracket^{c_i} = \lambda r : [x : \text{Stuff}] . p \left(\begin{array}{l} s_{pred} : [s_{fence} : \text{fence}(r.x)] \\ s_{c_base} : [s_{fence_{ind}} : \text{fence}_{ind_{join}}(r.x)] \end{array} \right)$$

Across situations, *fence* is interpreted relative to a context specific individuation schema $fence_{ind_{c_i}}$ where c_i is selected depending on the situation. From this it follows that individuation is maximised, but not reliability. In a given context, $fence_{ind_{c_i}}$ is disjoint, and so defined for counting, which leads to the desirable prediction that the exact result of counting the same stretch of fencing may result in different answers across situations.

In contrast, *fencing* applies the same individuation schema across situations, namely, one that is defined in terms of a join individuation schema type $fence_{ind_{join}}$, which consists of a number of individuation schemas. But this means that it is not disjoint. Take, for example, Rothstein’s square field example, where the sum of four fence sides is of type $fence_{ind_{join}}$, but so too are the four fence-sides taken individually, whereby the former overlaps with the latter. But this means that the question ‘How many fences are there?’ has two different possible answers: ‘one’ or ‘four’. In this sense, non-disjoint types are not countable, and so *fencing* is mass (Landman 2011).

7.3 *Mass/count variation in granulars*

In Section 5, we outlined how, for small quantities of rice, an agent may be left with a high degree of uncertainty whether or not to judge it as satisfying the predicate *rice*. If, as observed by Chierchia (2010), quantity of grains is a major factor affecting the applicability of a predicate like *rice* to a collection of entities, then we should also expect

the amount of uncertainty along an axis of quantity to be relatively smooth (graded).

For nouns that are context sensitive in this way, a learning challenge arises. We argued, in Section 5, that semantic learning for concrete nouns is largely governed by two pressures. One is to ascertain a consistent and reliable criterion for the application of a noun; the other is to establish what, if anything, the individuable units in the noun's denotation are, which is a prerequisite for counting. In simple cases, identifying the individuable units in a noun's denotation is to identify what the minimal entities are to license applying a predicate. This is the case for prototypical count nouns such as *cat*. If one has either a single cat, or a sum of single cats, one can correctly use the noun *cat* of them.

It is precisely the context sensitivity of granular nouns, such as *rice* and *lentils*, which provides a compelling argument in support of reliability and individuation as two pressures on semantic learning implicated in the acquisition of the mass/count distinction, because learning of granulars pushes reliability and individuation in opposite directions. The denotations of granulars contain perceptually individuable units (e.g., single rice grains, or single lentils). However, having either a single grain of rice or a single lentil does not always license that the noun *rice* or *lentil(s)* can be felicitously applied to them. This is because there are many contexts in which single grains of rice or single lentils, or even small quantities of rice grains or lentils are insufficient in quantity to count as *rice* or *lentil(s)*. Hence, individuating in terms of grains loses reliability.

One way to increase reliability is to make the quantitative function one that identifies aggregates of entities with the requisite properties such as colour, shape, etc., especially if the sizes of these aggregates are those most predictive of the appropriate conditions for using the relevant predicate. The most diagnostic sizes of aggregates will be those that are frequently encountered and have a high correlation with correct application of the relevant predicate. For example, if, say, spoonfuls, bowlfuls and packets of lentils are the most frequently encountered aggregates of lentils and almost always get judged to be lentils by competent speakers, then if an aggregate of lentils is a spoonful, a bowlful or a packet of lentils in size, then one has very good reason to apply *lentils* to that aggregate. Furthermore, if someone has

used the term *lentils* one has good reason to expect it to refer to such a frequently encountered aggregate size. Doing this would satisfy the pressure to establish a reliable criterion; however, it would do so at the expense of satisfying the individuation pressure.

We label the individuating predicate that is based on such single-grain properties $lentil_{Ind_gr}$. This schema will not provide good results for one aspect of reliability such as the conditional probability in (56), because small quantities of lentils are not good predictors for when to make a *lentil* judgement. In many contexts, larger quantities are required to count as *lentils*. With respect to the inverse conditional probability (57), it fares better, since the upward closure of the predicate which picks out single lentils will match the conditions for applying *lentil* in all but the cases where sub-grain parts of lentils count as *lentils*.

$$(56) \quad p(r : [s_{lentil} : lentil(x)] | r : [s_{lentil-ind} : *lentil_{Ind_gr}(x)]) = \text{lowish}$$

$$(57) \quad p(r : [s_{lentil-ind} : *lentil_{Ind_gr}(x)] | r : [s_{lentil} : lentil(x)]) = \text{highish}$$

The individuating predicate $lentil_{Ind_gr}$ maximises individuation, however. It applies to single lentils, which are disjoint. This means that there is only one variant, the predicate itself, hence entropy is 0.

$$(58) \quad - \sum_{v_i \in V} p(v_i | lentil_{Ind_gr}) \times \log p(v_i | lentil_{Ind_gr}) = 0$$

So, adopting $lentil_{Ind_gr}$ maximises individuation, but does so at the expense of reliability.

The alternative strategy is to choose a schema that is more reliable, namely in terms of aggregates (which were formally characterised in Sections 4.2-4.2). Instead of individuating only in terms of single grains, one could instead use a schema that identifies the sizes of aggregates of grains that are most diagnostic of when to apply the predicate *lentil* (a join type of the most diagnostic *lentil* aggregate sizes). Call this $lentil_{join_agg}$. As formalised in (60), the $lentil_{join_agg}$ predicate may also miss out on some cases where very small collections of lentils count as *lentils* (which indicates that our representation is missing some element of further context-sensitivity for granulars). However, it will do better with respect to predicting when to apply

lentil, given the schema as shown in (59). This is because the cases where there are insufficient amounts of lentils to make a *lentil* judgement will also be cases where it is insufficient to make a *lentil*_{join_agg} judgement.

$$(59) \quad p(r : [s_{lentil} : lentil(x)] | r : [s_{lentil-ind} : *1lentil_{join_agg}(x)]) = \text{high}$$

$$(60) \quad p(r : [s_{lentil-ind} : *1lentil_{join_agg}(x)] | r : [s_{lentil} : lentil(x)]) = \text{highish}$$

However, the aggregating strategy fares badly with respect to individuation. A join individuating predicate that identifies aggregates of some minimum sizes is not disjoint because e.g. spoonful sized aggregates form proper parts of e.g. bowlful sized aggregates. Such join types have multiple maximally disjoint variants. Therefore, each context specific schema yields a higher entropy value than in the *lentil*_{Ind_gr} case:

$$(61) \quad - \sum_{v_i \in V} p(v_i | lentil_{join_agg}) \times \log p(v_i | lentil_{join_agg}) = \text{high}$$

Neither of the two alternatives for individuation schemas can satisfy both pressures of individuation and reliability. *Lentil*_{Ind_gr} minimises entropy, thereby maximising individuation, but does not maximise reliability. *Lentil*_{join_agg} maximises reliability, but does not maximise individuation. As in the *furniture* and *fence* cases, this tension can result in two kinds of lexical entries involving the same number-neutral type *lentil*. Equation (62) uses *lentil*_{Ind_gr}, has a disjoint counting base, and so is the entry for a count noun such as the English *lentil*. Equation (63) uses *lentil*_{join_agg}, does not have a disjoint counting base, and so is the entry for a mass noun such as the Czech *čočka* ('lentil').

$$(62) \quad [[lentil]] = \lambda r : [x : Stuff]. p \left(\begin{array}{l} s_{pred} : [s_{lentil} : lentil(r.x)] \\ s_{c_base} : [s_{lentil-ind} : lentil_{Ind_gr}(r.x)] \end{array} \right)$$

$$(63) \quad [[čočka]] = \lambda r : [x : Stuff]. p \left(\begin{array}{l} s_{pred} : [s_{lentil} : lentil(r.x)] \\ s_{c_base} : [s_{lentil-ind} : lentil_{join_agg}(r.x)] \end{array} \right)$$

7.4 Mass/count stability in substances, liquids and gasses

When it comes to mass nouns like *mud*, *blood*, and *air*, similarly to granulars, the quantity of a substance has an impact on the applica-

bility of the noun in a way that varies with context. For example, a speck of mud on one's shoes could count as *mud* in a scientific clean room context, but not in a context where one is entering a garden shed. The principal difference between substances and granulars is in the perceptual properties of their references. Whereas in the granular case, there are clearly individuable entities which could be judged to be of some P_{Ind} type (e.g. $lentil_{Ind_gr}$), substances lack any such thing.

This means that, of the strategies so far considered, there is only one type individuation schema one might try to use, namely an amassment, the substance noun counterpart to an aggregating schema, that individuates in terms of a join of amassments of stuff with mud properties which are, jointly, the best indicators of when to apply mud. In a similar vein to the granulars case we have considered, the amassment schemas would fare well with respect to reliability ((64) and (65)).

$$(64) \quad p(r : [s_{mud} : mud(x)] | r : [s_{mud-ind} : *mud_{join_amass}(x)]) = \text{high}$$

$$(65) \quad p(r : [s_{mud-ind} : *mud_{join_amass}(x)] | r : [s_{mud} : mud(x)]) = \text{highish}$$

Individuation is militated against with such a schema, however, since there is a high number of admissible (disjoint) variants and presumably none of them will be particularly weighted over the others:

$$(66) \quad - \sum_{v_i \in V} p(v_i | mud_{join_amass}) \times \log p(v_i | mud_{join_amass}) = \text{high}$$

On the face of it, it may look as though this strategy is the only viable one. It maximises reliability, but does so at the expense of individuation. This leads us to expect most languages to develop a lexical entry for mud with an overlapping counting base, thus lexicalized with a mass noun. This is the case in English as in (67).

$$(67) \quad [[mud]]^{ci} = \lambda r : [x : Stuff]. p \left(\begin{array}{l} s_{pred} : [s_{mud} : mud(r.x)] \\ s_{c_base} : [s_{mud-ind} : mud_{join_amass}(r.x)] \end{array} \right)$$

Our account, therefore predicts relative stability in the mass lexicalization of substance, liquid and gas denoting nouns crosslinguistically.

However, we might ask if there is any way one could boost individuation, even for noun concepts which denote substances such as mud and blood. A clue for what kind of strategy might do this comes from languages like Yudja as reported in Lima (2014, 2016,

a.o.). In Yudja, different sizes/portions of substances such as blood can be directly counted provided that they are contextually disjoint. Lima’s (2014) analysis of Yudja relies on mereotopological concepts from Grimm (2012), and specifically on the concept of Maximal Self Connectedness (MSC) which is the property of countable entities. Informally, “an entity is self-connected means that whenever we partition this entity into two parts, these two parts are connected to each other.” (Lima 2014, p. 140)

We formalise this in terms of *bounded amassments* (Section 4.2), namely, identifying, at a perceptual level, distinct bounded regions formed from stuff with the requisite properties. For example, an individuation schema such as $blood_{bounded}$ applies to stuff with blood properties that also forms a bounded region; namely, a disjoint part of space containing blood. As such, the $blood_{bounded}$ predicate will individuate as there will not be multiple variants (e.g. a drop of blood will not be formed of disjoint bounded drops of blood). Namely, we have zero entropy as shown in (68).

$$(68) \quad -\left(\sum_{v_j \in V} p(v_j | blood_{bounded}) \times \log p(v_j | blood_{bounded})\right) = 0$$

However, although being a bounded region of e.g. blood may be a reliable indicator for applying $blood$ (69), being blood may not be a reliable indicator for being a bounded region of blood or a sum thereof, since blood (and other substances) do not always come in bounded portions. This translates into a lowering of the conditional probability in (70).

$$(69) \quad p(r : [s_{blood} : blood(x)] | r : [s_{blood-ind} : *blood_{bounded}(x)]) = \text{high}$$

$$(70) \quad p(r : [s_{blood-ind} : *blood_{bounded}(x)] | r : [s_{blood} : blood(x)]) = \text{not high}$$

Yudja does not have a rich lexicalized measurement system (aside from loan words (Lima p.c.)). The result is that the only way to quantify stuff (be it intuitively individuated or not) is by direct counting.²² Languages with such relatively rare characteristics could therefore be ones which adopt a strategy of individuating any bounded, disjoint amounts of stuff with the relevant perceptual (or functional) properties (e.g. colour, consistency, etc.). This strategy, applied across the

²²We are grateful to S. Rothstein for raising the possibility of this connection.

board to substance denoting nouns, could result in there being no genuine mass nouns in such languages, as is reported to be the case in Yudja (Lima 2014).

Substance denoting noun entries would, therefore, look like that for *apeta* ('blood', Yudja), as in (71), and would be count.

$$(71) \llbracket \text{apeta} \rrbracket^{ci} = \lambda r : [x : \text{Stuff}], p \left(\begin{array}{l} s_{pred} : [s_{blood} : \text{blood}(r.x)] \\ s_{c_base} : [s_{blood-ind} : \text{blood}_{bounded}(r.x)] \end{array} \right)$$

8 CONCLUSION AND SUMMARY

The formalism we have developed as a mereological enrichment of prob-TTR can be justified independently of issues surrounding the mass/count distinction. With respect to probabilistic semantics, there is increasing recognition that semantic, pragmatic, and knowledge representations, in order to be cognitively plausible, should be able to reflect gradience in judgements, and be consistent with a tractable account of semantic learning. Mereology is widely used in semantics for modelling plurality, tense, and aspect as well as the mass/count distinction. Using these formal tools, we tried to flesh out the intuition of Krifka (1989) that applying nouns involves both qualitative and quantitative criteria. We sketched how some properties, such as the size and boundedness of an aggregate of rice grains, could be modelled in a manner inspired by work on linking TTR representations to perceptual inputs, and how spatial perception is one factor in guiding the quantitative process of individuating entities. We have also shown how probM-TTR naturally accommodates cutting edge ideas on the semantics of the mass/count distinction, and, significantly, we are able to offer a unified explanation of why some classes of nouns display a wide amount of cross and intralinguistic mass/count variation while others do not; namely, as the result of balancing the pressures of individuation and reliability in semantic learning. Sometimes these pressures align (*prototypical objects*), sometimes they do not (*collective artifacts, non-bounded objects, and granulars*) and sometimes individuation cannot easily be prioritised at all (*substances*).

This yields four semantic classes of nouns which pattern differently with respect to the distribution their nouns have over the two grammatical properties MASS and COUNT. These are summarised in Table 1 and elaborated on below.

Noun class	Properties of counting base	Mass/count variation
Prototypical objects	Disjoint, single individuation schema across contexts	Rare
Collective artifacts & non-bounded objects	Multiple, disjoint, context specific schemas OR a single multiplicity of overlapping schemas	Common
Granulars	Disjoint schema picking out single grains, OR overlapping schema amassing aggregates of grains (e.g., spoonfuls, bowlfuls)	Common
Substances	Usually, an overlapping schema that groups frequently encountered amassments of stuff. Sometimes a schema that identifies contextually provided bounded amassments	Rare

Table 1:
Summary of
noun classes and
their properties

Prototypical objects: The types that pick out the individuable entities in the denotations of prototypical object nouns are also highly consistent indicators of when to apply the nouns. The pressures of individuation and reliability work in the same direction, i.e., they converge on the count encoding. We, therefore, have no reason to expect much variation from the count encoding, cross- and intralinguistically.

Collective artifacts and non-bounded objects: The context-sensitivity of nouns in these classes affects the reliability with which any single individual predicate type applies. For example, across contexts, a sum of fence pieces can count as one fence, or two fences; and a pestle and mortar can count as one item of kitchenware or two items of kitchenware. This means that any particular individuation schema will unreliably determine the extension. To prioritise individuation, multiple individuation schemas, each indexed to a context, can be used. This yields count nouns such as *fence*, and *Küchengeräte* ('kitchenware' German). Alternatively, to prioritise reliability, all individuation schemas can be merged together. This yields a non-disjoint schema, and so motivates the encoding of nouns, such as *fencing* and *kitchenware*, as mass nouns.

Granulars: Context-sensitivity with granular noun denotations has an effect on what quantities of the relevant stuff are needed to qualify for that stuff to fall under a given noun denotation. Granular nouns tend to be easily perceptually individuable (in terms of salient indi-

vidual grains), but given that single grains are not always enough to qualify as falling under a given noun denotation across all contexts, the type for single grains, that prioritises individuation, is inconsistent as a basis for applying a noun. Prioritising individuation yields a count noun encoding, which is commonly presupposed by pluralisation, e.g. *lentils*, *oats*, *kaurahiutale-et* ('oatmeal', Finnish). On the other hand, prioritising reliability yields a non-disjoint individuation schema, and so leads to a mass noun encoding, as in *oatmeal*, *kaura* ('oats', Finnish), *čočka* ('lentils', Czech).

Substances: Context-sensitivity also has an effect on amounts of quantities (e.g., of substances, liquids, and gases) reaching a certain threshold to qualify as falling under a given noun (e.g., *mud*, *blood*, and *air*). However, the perceptual qualities of the denotations of these nouns does not easily enable the prioritisation of individuation that could be achieved for count granular nouns. If individuation cannot easily be prioritised, then we should expect to find more cases where reliability will be. Therefore, we expect a heavy tendency towards mass encoding for these nouns.

REFERENCES

- J. L. AUSTIN (1950/1979), Truth, in J. O. URMSON and G. J. WARNOCK, editors, *Philosophical Papers, Third Edition*, pp. 117–133, Oxford University Press, Oxford, Originally in: *Symposium: Truth, Proceedings of the Aristotelian Society*, Vol. 24 (1950).
- Jon BARWISE and John ETCHEMENDY (1987), *The Liar: An Essay on Truth and Circularity*, Oxford University Press USA.
- Gennaro CHIERCHIA (2010), Mass Nouns, Vagueness and Semantic Variation, *Synthese*, 174:99–149.
- Robin COOPER (2012), Type Theory and Semantics in Flux, in R. KEMPSON, T. FERNANDO, and N. ASHER, editors, *Philosophy of Linguistics, Handbook of the Philosophy of Science*, pp. 271–323, Elsevier.
- Robin COOPER, Simon DOBNIK, Shalom LAPPIN, and Staffan LARSSON (2014), A Probabilistic Rich Type Theory for Semantic Interpretation, *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics*.
- Robin COOPER, Simon DOBNIK, Staffan LARSSON, and Shalom LAPPIN (2015), Probabilistic Type Theory and Natural Language Semantics, *LILT*, 10(4).

- Simon DOBNIK, Robin COOPER, and Staffan LARSSON (2012), Modelling language, action, and perception in type theory with records, in *Constraint Solving and Language Processing*, pp. 70–91, Springer Berlin Heidelberg.
- Jan van EIJCK and Shalom LAPPIN (2012), Probabilistic Semantics for Natural Language, in P. CHRISTOFF, N. GIERASIMSZUK, A. MARCOCI, and S. SMETS, editors, *Logic and Interactive Rationality Volume 2*, pp. 17–35, University of Amsterdam: ILLC.
- Hana FILIP and Peter SUTTON (2017), Singular Count NPs in Measure Constructions, manuscript, to be presented at SALT 2017.
- Charles J. FILLMORE (1975), An Alternative to Checklist Theories of Meaning, *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1:123–131.
- Charles J. FILLMORE (1976), Frame semantics and the nature of language, *Annals of the New York Academy of Sciences*, 280(1):20–32, ISSN 1749-6632.
- Scott GRIMM (2012), *Number and Individuation*, PhD Dissertation, Stanford University.
- Daniel C. HYDE and Elizabeth S. SPELKE (2011), Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition, *Developmental Science*, 14(2):360–371, ISSN 1467-7687, doi:10.1111/j.1467-7687.2010.00987.x, <http://dx.doi.org/10.1111/j.1467-7687.2010.00987.x>.
- Ray JACKENDOFF (1991), Parts and Boundaries, *Cognition*, 41:9–45.
- A. KOLMOGOROV (1950), *Foundations of probability*, Chelsea Publishing, New York.
- Manfred KRIFKA (1989), Nominal Reference, Temporal Constitution and Quantification in Event Semantics, in Renate Bartsch and J. F. A. K. van Benthem and P. van Emde BOAS, editor, *Semantics and Contextual Expression*, pp. 75–115, Foris Publications.
- Fred LANDMAN (2011), Count Nouns – Mass Nouns – Neat Nouns – Mess Nouns, *The Baltic International Yearbook of Cognition*, 6:1–67.
- Fred LANDMAN (2016), Iceberg Semantics for Count Nouns and Mass Nouns: The evidence from portions, *The Baltic International Yearbook of Cognition Logic and Communication*, 11:1–48.
- Daniel LASSITER (2016), Must, knowledge, and (in)directness, *Natural Language Semantics*, 24(2):117–163, ISSN 1572-865X, doi:10.1007/s11050-016-9121-8, <http://dx.doi.org/10.1007/s11050-016-9121-8>.
- Suzi LIMA (2014), All notional mass nouns are count nouns in Yudja, *Proceedings of SALT*, 24:534–554.

- Suzi LIMA (2016), Container constructions in Yudja: locatives, individuation and measure, *The Baltic International Yearbook of Cognition Logic and Communication*, 11:1–40.
- Godehard LINK (1983), The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretic Approach, in P. PORTNER and B. H. PARTEE, editors, *Formal Semantics - the Essential Readings*, pp. 127–147, Blackwell.
- S. PIANTADOSI, H. TILY, and E. GIBSON (2011), The communicative function of ambiguity in language, *PNAS*, 108(9):3526–3529.
- Roberta PIRES DE OLIVEIRA and Susan ROTHSTEIN (2011), Bare singular noun phrases are mass in Brazilian Portuguese, *Lingua*, 121:2153–2175.
- James PUSTEJOVSKY (1995), *The Generative Lexicon*, MIT Press.
- Susan ROTHSTEIN (2010), Counting and the Mass/Count Distinction, *Journal of Semantics*, 27(3):343–397, doi:10.1093/jos/ffq007.
- Peter R. SUTTON and Hana FILIP (2016a), Mass/Count Variation, a Mereological, Two-Dimensional Semantics, *The Baltic International Yearbook of Cognition Logic and Communication*, 11:1–45.
- Peter R. SUTTON and Hana FILIP (2016b), A probabilistic, mereological account of the mass/count distinction, *LNCS 10148, Proceedings of TbiLLC 2015*, p. To appear.
- Leonard TALMY (2000), *Toward a Cognitive Semantics – Vol. 1*, The MIT Press.
- Sandro ZUCCHI and Michael WHITE (1996), Twigs, Sequences and the Temporal Constitution of Predicates, in Teresa GALLOWAY and Justin SPENCE, editors, *Proceedings of SALT 6*, pp. 223–270, Linguistic Society of America.
- Sandro ZUCCHI and Michael WHITE (2001), Twigs, Sequences and the Temporal Constitution of Predicates, *Linguistics and Philosophy*, 24(2):223–270.

This work is licensed under the Creative Commons Attribution 3.0 Unported License.

<http://creativecommons.org/licenses/by/3.0/>

