

A probabilistic, mereological account of the mass/count distinction

Peter R. Sutton and Hana Filip*

Heinrich Heine Universität, Düsseldorf, Germany
peter.r.sutton@icloud.com, hana.filip@gmail.com,
<http://peter-sutton.co.uk>
<https://user.phil-fak.uni-duesseldorf.de/filip/>

Abstract. In this paper, we attempt to answer the vexing question why it should be the case that only certain types of noun meanings exhibit a mass/count variation in the lexicalization of their semantic properties, while others do not. This question has so far remained unanswered, or been set aside. We will do so by focusing on the role of context-sensitivity (already highlighted in recent theories of the mass/count distinction), and argue that it gives rise to a conflict between two pressures that influence the encoding of noun meanings as mass or count, one stemming from learnability constraints (*reliability*) and the other from constraints on informativeness (*individuation*). This will also lead us to identifying four semantic classes of nouns, and to showing why variation in mass/count encoding is, on our account, to be expected to occur widely in just two of them. Context-sensitivity forces a choice between prioritizing individuation, which aligns with count lexicalization, and prioritizing consistency, which aligns with mass lexicalization.

Keywords: count/mass, probabilistic semantics, mereology, vagueness, context-sensitivity.

1 Introduction

The focus of this paper is on some of the most puzzling data in the domain of the mass/count distinction, which have so far seemed intractable or have been

* First, we would like to thank the participants and organizers of the *Eleventh International Tbilisi Symposium on Language, Logic and Computation*. We especially enjoyed the inspirational and welcoming environment provided by our Georgian friends and colleagues. We are very grateful to Robin Cooper for his invaluable comments and suggestions during the preparation of this manuscript. We also wish to thank the audiences at *Type Theory and Lexical Semantics* held at *ESSLLI 2015* (Barcelona, Spain), *Sinn und Bedeutung 20* (Tbingen, Germany) and *Semantics and Philosophy in Europe 8* (Cambridge UK). Finally, we thank our colleagues and collaborators in the Department of Linguistics and the affiliated *Collaborative Research Center 991* at Heinrich Heine University, and in particular the participants of the *Semantics and Pragmatics Exchange (SemPrE)* colloquium. This research was funded by the German Research Foundation (DFG) CRC 991, Project C09.

set aside in current accounts. In a substantial number of cases, we observe cross- and intralinguistic variation in the lexicalization of nouns as mass or count. As is well-known, but puzzling, in languages with a fully-developed grammaticized lexical mass/count distinction, things in the world like furniture, jewelry, hair, lentils fall under a count or a mass description, but cats are uniformly describable by basic lexical count nouns, while air or mud by mass nouns. The questions to ask are: ‘Why do certain types of noun meanings exhibit a mass/count variation in the lexicalization of their semantic properties, while others do not?’ ‘Is this variation ad hoc, arbitrary, or is it due to some general principles that underlie the form-meaning mappings in the noun domain?’ We will address these questions by developing two key ideas from recent work on the mass/count distinction that both highlight the importance of context-sensitivity. First, some nouns are context-sensitive in that what counts as minimal in a number neutral predicate’s denotation varies with context. This form of context sensitivity is also associated with *vagueness* (Chierchia [2]). For example, what counts, minimally, as *rice* or *mud* is context dependent. A single grain of rice or a single fleck of mud is sufficient to count as *rice* or *mud* in some contexts, but not in others. Second, what counts as ‘one’ unit with respect to a given noun varies with context, as well (Rothstein [20], Landman [12]). For example, in some ‘counting contexts’ (Rothstein [20]), a teacup and saucer will count as one item of kitchenware, in other contexts as two items, and in other cases a teacup and saucer may “simultaneously in the same context” count as both one and two items of kitchenware (Landman [12]).

We argue that a more general level of explanation underlies the impact of context on countability, namely, that context-sensitivity of either of the two varieties just mentioned can be understood as giving rise to a conflict between two pressures on how languages encode the meanings of nouns, and which lead to predictions about when exactly variation in the mass/count lexicalization patterns is to be expected. One pressure, *reliability*, is derived from learnability constraints, and has to do with consistent criteria guiding the acquisition of noun meanings and their felicitous use in a variety of contexts. The second pressure, *individuation*, is derived from constraints on informativeness, which, for nouns, as we argue, amounts to the pressure to encode what counts as ‘one’ entity in their denotation. In other words, and put in the simplest terms, what is understood as individuation, as a prerequisite for counting, is here recast partly in information-theoretic terms.

The two varieties of context-sensitivity (one related to ‘quantity vagueness’ and the other to what counts as ‘one’) and the conflict they generate between the pressures coming from learnability constraints and constraints on informativeness, which impact on how languages encode the meanings of nouns, leads us to identifying four semantic classes of nouns. Most importantly, we motivate why only two of these, *granulars* and *collective artifacts and homogeneous objects* are systematically subject to the striking variation in the mass and count lexicalization, while the other two, *prototypical objects* and *substances, liquids, gasses*, are not. In brief, context-sensitivity forces a choice between prioritizing individua-

tion, which aligns with count lexicalization, and prioritizing consistency, which aligns with mass lexicalization. As far as we know, no motivation of this kind has yet been provided.

We formally represent these ideas in a probabilistic mereological theory, *probabilistic mereological Type Theory with Records* (probM-TTR) which is an enrichment of probabilistic Type Theory with Records (Cooper et al. [5]). This theory has the advantage that it provides us with rich representational means to model the key ideas and processes that, as we argue, underly the mass/count distinction: namely, vagueness, counting-context sensitivity, overlap between entities that count as one, the impact of semantic learning on meaning representations, reliability of application criteria for nouns, and why, in some cases, multiple individuation criteria are licensed.

In section 2, we summarize some of the leading recent contributions to the semantics of the mass/count distinction and we highlight and connect the role of context in each of them. In section 3, we argue that the two notions of context sensitivity identified in section 2 can be used to demarcate four semantic classes of nouns. We then argue for the presence of two competing pressures on natural language predicates that we call *reliability* and *individuation*. In section 4, we briefly introduce our formal framework, probM-TTR. In sections 5 and 6, we show how the two types of context sensitivity from section 2 give rise to conflicts between individuation and reliability, and so also give rise to the licensing of either mass or count encoding. We summarize these findings in section 7.

2 Context-Sensitivity and the Mass/Count Distinction

2.1 Vagueness as a variation in extensions across contexts: Chierchia (2010)

Chierchia’s [2] main claim is that mass nouns are vague in a way that count nouns are not. While count nouns have *stable atoms* in their denotation, that is, they have entities in their denotation that are atoms in every context, mass noun denotations lack *stable atoms*. If a noun lacks stable atoms, there is no entity that is an atom in the denotation of the predicate at all contexts. In this sense then, mass nouns have only *unstable individuals* in their denotation. But counting is counting of stable atoms only. Therefore, mass nouns are uncountable.

Chierchia enriches mereological semantics with a form of supervaluationism wherein vague nouns interpreted at ground contexts have extension gaps (vagueness bands). Contexts then play the role of classical completions of a partial model in other supervaluationist formalisms such that at every (total) context, a nominal predicate is a total function on the domain.

Contexts stand in a partial order to one another such that if c' precisifies c ($c \propto c'$), then the denotation of a predicate P at c and at a world w is a (possibly not proper) subset of P at c' and w . For an interpretation function F : $F(P)(c)(w) \subseteq F(P)(c')(w)$.

On Chierchia’s supervaluationist account, mass nouns such as *rice* are vague in the following way. It is not the case that across all contexts, for example, a few

grains, single grains, half grains, and rice dust always count as rice. Thus these quantities of rice are in the vagueness band of *rice*. There may be some total precisifications of the ground context c , in which single grains are rice atoms. There may also be some c' such that $c \propto c'$, where half grains are rice atoms. There may also be some c'' such that $c' \propto c''$, in which rice dust particles are rice atoms. Most importantly, 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.

2.2 Disjointness in Context

Rothstein [20] argues that neither formal atomicity (defined in mereological terms with reference to a Boolean lattice structure, presupposed by Chierchia 1998), nor natural atomicity (understood in terms of a “natural unit” in the sense of Krifka [10]) are sufficient or necessary to account for the differences between mass nouns and count nouns. A major contribution of Rothstein’s work is to provide a formal model of how nouns such as *fence*, *wall*, which are not naturally atomic, nonetheless exhibit the hallmark grammatical properties of count nouns.

In contrast to Chierchia’s use of context, Rothstein [20] 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 are indexed via entity–context pairs and so are of type $\langle e \times k, t \rangle$. The following is Rothstein’s example. Suppose that a square field is encircled by fencing. The answer to the question *How many fences encircle the field?* is wholly dependent on context. In some contexts, it would be natural to answer four (one for each side of the field). In other contexts, it would be more natural to answer one (one fence encircling the whole field). By indexing count nouns to contexts, Rothstein is able to capture how there can be one answer to the above question in any particular context (either one or four), despite fence lacking natural atoms, atoms that are independent of counting-context.

Rothstein’s and Chierchia’s context differ in their formal properties. On the assumption that we restrict our discussion to what Rothstein refers to as “default contexts” (relative to which the denotations of predicates are disjoint), Rothstein’s contexts are not precisifications definable as a partial order. For example, let us again consider the field surrounded by fences a, b, c, d . Then at the context, k , at which a, b, c, d each individually counts as a single fence, their sum $a \cup b \cup c \cup d$ is excluded from the denotation of *fence*, while at the context k' at which $a \cup b \cup c \cup d$ jointly count as a single fence, a, b, c, d each taken individually are excluded from the denotation of *fence*. Clearly, therefore, one context does not precisify another.

There is, however, arguably a formal connection between the use of ‘context’ in Rothstein [20] and Chierchia [2]. Take the following quote from Chierchia, where, for ease of comparison, we added Rothstein’s fence example to his mountain(s) example.

“We must independently require (on anyone’s theory) that for a concrete sortal noun N , its atoms are chosen so as not to overlap spatiotemporally.

To put it differently, a disagreement over whether what you see in (43) is one or two mountains [ONE OR FOUR FENCES, IN OUR FIELD EXAMPLE ABOVE, Sutton&Filip] is, in the first place, a disagreement on how to resolve the contextual parameters. The key difference between nouns like *heap* or *mountain* [OR FENCE, Sutton&Filip] and mass nouns like *rice* is that minimal rice amounts, once contextually set, can still be viewed as units or aggregates without re-negotiating the ground rules.” Chierchia [2] p. 123.

From this point of view, we could therefore, tentatively, associate the role of ground contexts (the contexts that set the “ground rules”) in Chierchia [2] with the role of counting contexts in Rothstein [20]. Ground contexts, for Chierchia, set upper bounds on precisifications. This means that they set the positive extension for predicates. Formally speaking, ground contexts set the precisification g such that for all precisifications c , if $c \propto g$, then $c = g$.¹ In this sense we have two distinct notions of context. For counting one must, as per Rothstein’s account, set a schema of individuation (via a counting context). However, as per Chierchia’s account, there may still be ways to resolve the extension of a predicate across contexts of use that can undermine countability by obscuring what the individuals for counting are. In sections 5 and 6, we make these two types of contexts explicit in our formalism and analyze how they interact.²

2.3 Overlap in context

In Landman [12] the set of generators, $\mathbf{gen}(X)$, of the regular set X is the set of semantic building blocks, which are either “the things that we would want to count as one” Landman [12, p. 26], relative to a context, or are contextually determined minimal parts. If the elements in the generator set are non-overlapping, as in the case of count nouns, then counting is sanctioned: Counting is counting of generators and there is only one way to count. However, if generators overlap, as in the case of mass nouns, counting goes wrong. One of Landman’s innovations is to provide a new delimitation of the two cases when this happens, and hence two subcategories of mass nouns: mess mass nouns like mud, and neat mass nouns like furniture. A mass noun is *neat* if its intension at every world specifies a regular set whose set of minimal elements is non-overlapping. A noun is a mess mass noun if its intension at every world specifies a regular set whose set of minimal elements is overlapping.

For Landman, counting goes wrong when the variants of the generator set have different cardinalities simultaneously, but under different “perspectives”

¹ Thanks to a reviewer who pointed out this formulation.

² Our association of ground context and counting context is only tentative, however. They may, formally, operate in a similar manner, but there are clear informal differences. For example, Rothstein’s counting contexts are meant to set the ground rules in the sense of determining what count as ‘one’. Chierchia’s ground contexts set the ground rules more in the sense of determining the boundary between the positive extension and the vagueness band.

on one and the same set of entities. Variants of a set are maximally disjoint subsets of a set. For example, for the set $X = \{a, b, c, d, a \cup b, c \cup d\}$, there are four variants of X : $v_1 = \{a, b, c, d\}$, $v_2 = \{a, b, c \cup d\}$, $v_3 = \{a \cup b, c, d\}$, $v_4 = \{a \cup b, c \cup d\}$. Clearly, therefore, the effect of deriving variants of a set can be associated with the effect of applying a default counting context (from Rothstein [20]) to a predicate: every variant marks one way that an overlapping denotation could be made disjoint.

Context, although not a prominent part of Landman’s account, is mentioned in relation to neat mass nouns. His paradigm example of a neat mass noun is *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 [12] pp. 34-35.

A striking 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, that the individuation schema one applies fails to resolve overlap. The former possibility is in effect a description of Rothstein’s typical distinction between mass and count nouns wherein mass nouns are not indexed to counting contexts. However, equally possible is that, for some reason, one may choose a schema of individuation that fails to remove overlap. We will motivate this latter option in section 3.

3 Count/Mass Variation, Reliability, and Individuation

3.1 Four Semantic Classes of Nouns and the Variation in Mass/Count Encoding

Considering just concrete nouns, as most do, we observe a considerable amount of puzzling data with respect to the variation in mass/count encoding between and within languages. This variation is not random, however. We may distinguish five classes of nouns depending on their main lexicalization patterns. They are summarized in Table 1 where the ‘Noun Class’ is a cover term for the descriptive labels below it. We then argue that these may be grouped into four classes in terms of the semantic properties given in Table 2.

The first striking pattern that we observe is markedly little variation in the mass/count encoding in two groups: namely, first, there is a strong tendency for *substances, gasses, liquids* to be encoded as mass (*mud, blood, air*), and second, a strong tendency for both animate and inanimate prototypical individuals (*prototypical objects*) to be encoded as count (*cat, boy, chair*). The second striking pattern is a substantial amount of variation in the encoding of *collective artifacts*

Table 1. Classes of Nouns and Mass/Count Variation

Noun Class	Examples
proto-typical objects	<i>chair</i> _{+c} ; <i>tuoli</i> _{+c} ('chair' Finnish); <i>Stuhl</i> _{+c} ('chair' German) <i>dog</i> _{+c} ; <i>koira</i> _{+c} ('dog' Finnish); <i>Hund</i> _{+c} ('dog' German) <i>boy</i> _{+c} ; <i>poika</i> _{+c} ('boy' Finnish); <i>Junge</i> _{+c} ('boy' German)
collective artifacts	<i>furniture</i> _{-c} ; <i>huonekalu</i> - <i>t</i> _{+c,PL} ('furniture' Finnish) <i>meubel</i> - <i>s</i> _{+c,PL} , <i>meubilair</i> _{-c} ('furniture' Dutch) <i>kitchenware</i> _{-c} ; <i>Küchengerät</i> - <i>e</i> _{+c,PL} (German, lit. kitchen device-s) <i>footwear</i> _{-c} ; <i>jalkineet</i> - <i>t</i> _{+c,PL} ('footwear' Finnish)
homogeneous objects	<i>fence</i> _{+c} , <i>fencing</i> _{-c} ; <i>hedge</i> _{+c} , <i>hedging</i> _{-c} <i>wall</i> _{+c} , <i>walling</i> _{-c} ; <i>shrub</i> _{+c} , <i>shrubbery</i> _{-c}
granulars	<i>lentil</i> - <i>s</i> _{+c,PL} ; <i>linse</i> - <i>n</i> _{+c,PL} ('lentils' German) <i>lešta</i> _{-c} ('lentils' Bulgarian); <i>čočka</i> _{-c} ('lentils' Czech) <i>oat</i> - <i>s</i> _{+c,PL} ; <i>oatmeal</i> _{-c} ; <i>kaura</i> _{-c} ('oats' Finnish); <i>kaurahiutale</i> - <i>et</i> _{+c,PL} (Finnish, lit. oat.flake-s)
substances, liquids, gases	<i>mud</i> _{-c} ; <i>muta</i> _{-c} ('mud' Finnish); <i>Schlamm</i> _{-c} ('mud' German) <i>blood</i> _{-c} ; <i>veri</i> _{-c} ('blood' Finnish); <i>Blut</i> _{-c} ('blood' German) <i>air</i> _{-c} ; <i>lenta</i> _{-c} ('air' Finnish); <i>Luft</i> _{-c} ('air' German)

as mass/count *furniture*, *footwear*, *kitchenware*; *homogeneous objects* ('homogeneous' in the sense of Rothstein [20]) like *fence*, *wall*, *hedge*; *granulars* like *rice*, *lentils*. Such observations immediately prompt the question what is the reason why the mass/count variation is rife for some nouns, but scarce for others? What semantic facts or constraints allow us to make predictions when the mass/count variation is expected?

Let us first consider in more detail the groupings which display much mass/count variation (table 2) and their attribution of properties, which are based on Rothstein [20], Chierchia [2], and Landman [12].

Prototypical objects: These nouns are not vague in the sense of Chierchia [2]. In Landman's [12] terms, they are count and so have non-overlapping minimal generators and non-overlapping generators. In Rothstein's [20] terms, these count nouns and as such indexed to counting contexts, and they have atoms in their denotations that do not vary across counting contexts. A dog, a chair, or a boy will count as one *dog*, *chair*, or *boy* by any reasonable schema of individuation.

Collective artifacts: These nouns contain typical cases of what Chierchia [2] calls "fake mass" nouns (following a long-standing tradition), and for which Landman [12] coins the term "neat mass" nouns: e.g., *furniture*, *footwear*, *kitchenware*. Chierchia takes the mass encoding of these nouns to be independent of vagueness, because they have stable atoms. Landman takes these nouns to have overlapping generators (but their minimal generators are non-overlapping). If it is the case that, from counting context to counting context, what counts as 'one *P*' varies, then, these nouns are counting context sensitive. Most importantly, they also have count counterparts, cross- and intralinguistically. Take *footwear* versus *jalkineet*_{+c,PL} ('footwear', Finnish). On Landman's account, a shoe, and a pair of shoes can count as single items of footwear simultaneously in the same

Table 2. Interpretation of theories of the mass/count distinction

Noun Class	Variation	Chierchia [2]	Landman [12]	Rothstein [20]
prototypical objects	Little	Not Vague	Not Overlapping generators	Not context sensitive
collective artifacts & homogeneous objects	Much	Not Vague	Overlapping generators	Context sensitive
granulars	Much	Vague	Not Overlapping generators	Not context sensitive
substances, liquids, gases	Little	Vague	Overlapping generators	(Context sensitive)

context. On Rothstein’s account, being indexed to a (default) counting context would prohibit this, but it is not the case that Finnish must pick a single counting context. In some contexts a pair of shoes will count as one item of footwear. In another context, a pair of shoes will count as two items of footwear.

Homogeneous objects: Following Rothstein [20], we use “homogeneous” as a description of nouns such as *fence*, *wall*, *hedge*. The homogeneity is meant to capture that, at least for relatively large samples, a single stretch of fence or wall could be viewed, in another context, as two or more stretches of fence or wall. According to Chierchia (see 2.2), nouns such as *fence* and *wall* do not denote unstable entities relative to a ground context (e.g. relative to a counting context). On Rothstein’s account, these are central cases of context-indexed nouns that are counting context sensitive. Most significantly, notice that these count nouns have mass counterparts (*fencing*, *walling*, *hedging*). As mass nouns, they presumably have, on Landman’s account, overlapping generators. It seems reasonable to conclude that, for example, *fencing* denotes overlapping entities that can, simultaneously and in the same context, count as single Landman categorizes *fencing* as neat mass (p.c.). This would lead one to expect a felicitous cardinality comparison with, for example, *more than* constructions. However, native speakers are divided on the felicity of this reading. If this is an accurate description, then homogeneous objects pattern along with collective artifacts as not vague, overlapping, and counting context sensitive, hence the grouping of the two in table 2.

Granulars: The denotations of granular nouns (*rice*, *lentils*) contain small grains. On Chierchia’s [2] account, these nouns are vague, since no quantities of grains or parts of grains are stable atoms (in some contexts parts of grains would suffice, in other contexts, more than one grain may be required). Notably, those mass nouns in these categories often have cross-linguistic count-counterparts. On Landman’s account, these count-counterparts should have non-overlapping generators. For example, the generators of *lentil* are presumably the individual

lentils (they count as one).³ However, it is hard to see how less than or more than a single lentil could equally count as one lentil, thus these granular count nouns arguably have non-overlapping minimal generators (they are *neat*). Similarly, for nouns such as *lentil*, it is hard to see how counting context could affect this individuation criteria. If single lentils count as one on one counting context for *lentil*, then, like nouns such as *cat*, they should count as one across all counting contexts. Although nouns such as *lentil* should be indexed to counting contexts on Rothstein’s account, they are not counting context sensitive. Despite the mass encoding of granular nouns such as *rice*, we take similar considerations to apply.⁴ Furthermore, reasons for thinking that nouns such as *rice* are neat, not mess, are given in [19].

Substances, gasses, liquids: These nouns are also vague on Chierchia’s [2] account. On Landman’s [12] account, such nouns are mess mass (because they have overlapping *minimal* generators). Insofar as these nouns are rarely encoded as count, it is hard to say whether or not they are counting context sensitive. However, in Yudja (Lima [14]), at least for nouns such as *mud* which do display count noun behavior, it seems that the quantities of mud that can count as one could vary from context to context (a pile in one context, a bucketful in another). Hence, we may tentatively conclude that *mud* is counting context sensitive (hence the parenthesis in table 2).⁵

3.2 Two Competing Pressures: Reliability and Individuation

In the formal framework we propose in sections 4–6, we will investigate a hypothesis that could account for much of the cross- and intralinguistic mass/count data. Our hypothesis supposes that there are (at least) two competing pressures on natural languages, one derived from learnability, the other from being a tool for effective communication. We take a cue for this proposal from work on information theoretic models of communication. For an example of how this type of approach of balancing learning and communicative pressures can be used to derive a theory of vagueness in information theoretic terms, see Sutton [17]. For a comparable approach applied to ambiguity, see [15].

³ This is a vexed issue, however. Prima facie, *rice* and *lentil-s* should be treated similarly, however the mass noun *rice* should have overlapping generators, but the count noun *lentil* should have non-overlapping generators.

⁴ Actually, this issue is also somewhat vexed. Nouns such as *lentil* cause problems for Landman [12] since, if subparts of lentils are not in the generator set and constitute proper parts of elements of the generator set, then they should not be in the denotation of *lentil(s)*, but this prediction is not accurate. This problem is remedied in Landman [13], where generators are replaced by “bases”.

⁵ There are also nouns which denote *fibrous* entities like *hair(s)*, *string(s)* which, on the one hand seem to pattern with granulars like *rice* insofar as they denote saliently perceptually distinguishable entities and are lexicalized as mass, but on the other hand, they also pattern with context-sensitive count nouns like *fence* insofar as what counts as one is contextually determined.

Generally, there is an informational trade-off between being more informative and being learnable. For example, in the extreme case, a language could have one and only one predicate to describe all entities. This would be easily learnable, but maximally underdetermined, and so be a highly inefficient means of communication. At the other extreme, one could have a lexicalized classifier for every discernible property (e.g. a different lexical items for one N, two Ns, two big Ns etc.). Each of these classifiers would be highly informative, but would make languages unstable and unlearnable, since a learner would not receive sufficient instances of all classifiers to be able to infer their denotations (this is a form of the ‘bottleneck’ problem as discussed in the iterative learning paradigm [9]). Typically, classifiers convey an amount of information that in some way balances these pressures.

These general pressures are instantiated in the learning of concrete nominal predicates. On the one hand, there is a pressure for nominal predicate to be informative. In these cases, the amount of information conveyed is linked to how much of the domain is excluded by a classifier. Intuitively, a predicate which allows one to *individuate*, to pick out individual entities, is more informative than one which conveys no individuation schema, hence there is a general pressure towards establishing an individuation schema if this is possible given other perceptual and/or functional properties of the entities in the denotation of the relevant predicate. Individuation can be set in information theoretic terms. If the meaning of a noun (the signal) determines a specific criteria for counting, as opposed to something more ambiguous or vague, then the message will be more informative (carry a higher informational value). For example, if N_1 specifies $\{a, b, c\}$ as countable entities with some high probability (and so excludes sums thereof), but N_2 is has a level distribution between the sets $\{a, b, c\}$ and $\{a \cup b, a \cup c, b \cup c\}$, then in information theoretic terms, N_1 carries more information than N_2 .

On the other hand, there is a pressure for learnability. One’s criteria for classifying should be a reliable indicator of the correct way to apply the predicate, and consistently across various contexts. Call this pressure *reliability*. In the context of countability, if the individuation criteria sometimes correctly but sometimes wrongly excludes entities from the denotation of a noun, then it is unreliable. This very simple pressure, in effect requires that the probability of correctly applying a predicate, given the individuation schema is high.

As we will discuss in Sections 5 and 6, these pressures may sometimes push in opposing directions. However, in the case of prototypical count nouns, *reliability* pushes in the same direction as *individuation*. There is a single and specific individuation schema for cat, namely being a cat individual (a single cat). Furthermore, being a cat individual (or a sum thereof) is a very good indicator of being in the denotation of *cat(s)*.

4 Formal Framework

4.1 Type Theory with Records (TTR)

Type Theory with Records (Cooper [6], and references therein) is a richly typed formalism with a wide number of possible applications. In the following, we discuss only its application to natural language semantics and the representation of semantic structures as a form of compositional frame semantics (for discussion see Cooper [3], [6]). In its application to natural language semantics, TTR is a system that combines insights from Fillmore’s frame semantics [7] and situation theory, but also from formal semantics in the Montague tradition. In this section, we briefly introduce readers to the aspects of TTR that we will use in this article. Full formal details can be found in Cooper [6].

Two formal structures that are central to TTR are *records* and *record types*. Records are approximately *situations* from situation theoretic semantics, and record types are *situation types* from the same tradition, *frames* in the sense of Fillmore, but also what act as the TTR equivalent of *propositions*, namely, intensional structures that are made true by parts of the world, i.e. records/situations.

Record Types are represented as Field-Type matrices such as the one in (1) which details a highly simplified *cat*-frame.

$$\left[\begin{array}{l} x : Ind \\ s_{cat} : \langle \lambda v:Ind(\text{cat}(v)), \langle x \rangle \rangle \end{array} \right] \quad (1)$$

The fields (to the left of the colons) contain the labels x , s_{cat} will determine what values are provided by the record (situation) to which this frame is applied. For those more familiar with frameworks such as DRT, labels can also be thought of as approximating discourse referents. To the right of the colons are types. *Ind* is the basic type for individuals. In the spirit of semantics in the Frege-Montague tradition, predicates are functions. $\langle \lambda v.cat(x), \langle x \rangle \rangle$ is a predicate which is a function from entities of type *Ind* to a type of situation. It is important to note that predicates apply to values for labels, not labels themselves. For example, if the value for x is *felix*, then this will yield a type of situation, $\text{cat}(\text{felix})$ in which Felix is a cat.

To form *properties* (the equivalent of expression of expressions of type $\langle e, t \rangle$), frames can be abstracted over to take a record as an argument. This is shown in (2) and provides a highly simplified representation of the English *cat*. What (2) requires is an application to a situation (record) which contains an individual. Now the type is restricted to take the value for the label x in the record ($r.x$), and apply it to the type statements in the record type/cat frame.

$$\lambda r : [x : Ind]. [s_{cat} : \langle \lambda v:Ind(\text{cat}(v)), \langle r.x \rangle \rangle] \quad (2)$$

Such a record could be the one given in (3). Records are finite sets of ordered pairs of labels and values. This is shown in matrix format in (3) where the label is x and the value is *felix*.

$$[x = \text{felix}] \quad (3)$$

For our purposes, *felix* could be thought of as the actual cat, and the label is just a way of tracking and accessing this object. Applying the record in (3) to the function in (2) yields a proposition: $[s_{cat} : \text{cat}(\text{felix})]$ which will be true just in case there is a situation in which Felix is a cat. In other words, propositions are the equivalent of $\langle s, t \rangle$ expressions, except that TTR propositions are true of situations, which are partial and more cognitively plausible as truth makers than non-partial worlds (usually understood as sets of propositions).⁶

For this very brief introduction to TTR, another important point to note is the role of agents in the formalism. Agents can make a judgement that some object or situation a is of some type T (A judges that $a : T$). In an Austinian spirit, type judgements of this kind can be true or false. In section 4.2 we will expand on how the notion of an agent’s judgement set is linked to a probabilistic learning model (fully detailed in Cooper et al. [4], [5]).

Finally, with respect to notation, we henceforth follow the standard brevity convention in TTR by simplifying how predicates are represented. Instead of $\langle \lambda v(P(v)), \langle x \rangle \rangle$ we will use just $P(x)$. For example, the frame in (1) will, under the convention, be represented as in (4).

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

4.2 Probabilistic Type Theory with Records (prob-TTR)

A full outline of prob-TTR may be found in Cooper et al. [4], [5], we again introduce only that which will be necessary for our purposes. The central enrichment of TTR made in prob-TTR is to replace truth/falsity conditions of judgements with probability conditions. In later work, Cooper et al. [5] say that this is the probability of a judgement being the case, however, more in the spirit of the learning centric approach detailed in prob-TTR, we find that a more informative gloss on probability value for a judgement is the probability an agent ascribes to a competent speaker making that judgement (estimated with respect to her linguistic experiences and learning data).⁷ Once integrated with a Bayesian learning model, probabilistic judgements are symbolized $p_{A, \mathfrak{J}}(a : T) = k$ or the probability that agent A judges that a is of type T with respect to her judgement set \mathfrak{J} is $k \in [0, 1]$. Judgement sets record type judgements made of particular situations along with a probability value. Judgement sets are updated and form the basis for novel type judgements by the agent. The value k in (5) will represent the prior probability an agent A has for some individual being a cat, given her judgement set \mathfrak{J} . Conditional probabilities are then computed as in (6) using a type theoretic version of Bayes’ Rule where $\|T\|_{\mathfrak{J}}$ is the sum of all probabilities associated with T in \mathfrak{J} .

⁶ Another feature of TTR is that types are inherently intensional. This is because types are themselves viewed as objects to which other objects/situations belong, not merely as sets of objects/situations. As such, two distinct types may be coextensional.

⁷ This formulation is due to Shalom Lappin p.c.

$$p_{A,\mathfrak{J}}(s : \left[\begin{array}{l} x : Ind \\ s_{cat} : cat(x) \end{array} \right]) = k \quad (5) \quad p_{A,\mathfrak{J}}(s : T_1 | s : T_2) = \frac{\|T_1 \wedge T_2\|_{\mathfrak{J}}}{\|T_2\|_{\mathfrak{J}}} \quad (6)$$

4.3 Probabilistic, Mereological Type Theory with Records (probM-TTR)

The simple enrichment we make to prob-TTR is to expand the domain of the basic type *Ind* from individuals to individuals and mereological sums thereof.⁸ That is to say that we replace the basic type for individuals with the type of ‘stuff’ which we express as the basic type **Ind*. A learner’s task will be to establish what, if anything, the individuals denoted by a particular predicate are. For example, given a world full of stuff, a learner of the predicate *cat* must learn which portions of stuff are individual cats. The type of individual for a predicate *P* will be represented *Ind_P*, so the type of single cat individuals will be *Ind_{cat}*.

Following Krifka [10,11], we distinguish between a qualitative and a quantitative criterion for applying nominal predicates.⁹ Qualitative criteria may include perceptual properties such as color, shape, size and perceptual individuability (for example, grains of sand are harder to perceive and differentiate than grains of rice), but also functional properties. For simplicity, here we simply refer to this cluster of properties for a predicate *P* as the predicate *P_{Qual}*. This simple looking predicate should actually represent an entire frame that details, for example, functional and perceptual aspects of denotations relevant for forming predicate judgements. We will elaborate on the details of these frames in further work. This qualitative frame then acts as an argument for a ‘quantitative’ function *f_{P_{quant}}* : (*RecType* → *NatNum*). This is a function which outputs a natural number as a quantity value for some stuff with some combination of the relevant *P* qualities.

$$\left[\begin{array}{l} s_{pstuff} : \left[\begin{array}{l} x : *Ind \\ s_{pqual} : P_{Qual}(x) \end{array} \right] \\ f_{pquant} : \left(\left[\begin{array}{l} x : *Ind \\ s_{pqual} : P_{Qual}(x) \end{array} \right] \rightarrow \mathbb{N} \right) \\ i : \mathbb{N} \\ s_{pquant} : f_{pquant}(s_{pstuff}) = i \end{array} \right] \quad (7) \quad \left[\begin{array}{l} s_{ricestuff} : \left[\begin{array}{l} x : *Ind \\ s_{ricequal} : rice_{Qual}(x) \end{array} \right] \\ f_{ricequant} : \left(\left[\begin{array}{l} x : *Ind \\ s_{ricequal} : rice_{Qual}(x) \end{array} \right] \rightarrow \mathbb{N} \right) \\ s_{ricequant} : f_{ricequant}(s_{ricestuff}) = 1 \end{array} \right] \quad (8)$$

Examples of how we represent the qualitative frame and the quantitative function are given as a schema in (7) and for the predicate *rice(x)* in (8). In both, the first field labels a type of situation in which some stuff has the relevant

⁸ This could equally be achieved using sets. For a set of formal atoms $\{a, b, c\}$, the domain of *Ind* entities would be $\{a, b, c, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$

⁹ As pointed out by a reviewer, a related concept is discussed by Geach [8]. However, Geach’s criteria of identity is not identical with, for example, Krifka’s *Natural Unit* function.

P-qualities/*rice*-qualities. The second field specifies a function from the quality record type to a natural number. The fourth field in (7) and the third field in (8) show the output to this function. In (8), this has been specified as 1. For this special case, this will be the type for single rice grains since the perceptually salient partition of rice is into grains. In this special case, we adopt an abbreviation convention in which (8) is rewritten as $[x : Ind_{rice}]$.

4.4 Prototypical count nouns

We can now specify the lexical entry for a concrete noun. Landman [12] specifies lexical entries as pairs of sets \langle denotation, counting base \rangle . We emulate this idea with frames and also adopt the terminology of Landman [13] of *body* for the regular denotation of a predicate, and *base* for the counting base. It should be emphasised, however, that the precise meaning of body and base for us differs from Landman’s proposal. For a predicate such as $cat(x)$, we get:

$$\lambda r : [x : *Ind]. \left[\begin{array}{l} s_{body} : [s_{cat} : cat(r.x)] \\ s_{base} : [r.x : Ind_{cat}] \end{array} \right] \quad (9)$$

Entities of the type for the label s_{body} are in the denotation of the number neutral cat-property. Entities of the type for the label s_{base} provide the potentially countable entities for the number neutral property (the single cats).

This pair of types balances the pressures of individuation and reliability. Picking out single cats from the type of stuff is highly informative, since there is very little uncertainty as to which set of entities should be judged as cat individuals. The individuation schema provided by Ind_{cat} is also a highly reliable indication that one may apply the predicate $cat(x)$. If something is a cat individual in one context, it will rarely if ever be the case that one cannot apply the predicate $cat(x)$ to this individual across contexts. To see why the two pressure of individuation and reliability are both satisfied in this case, consider an alternative individuation schema that would be roughly as informative, for example, one which selected with a high probability, all cat pairs (every sum of two single cats). Unlike the good case, this schema would not be reliable, since it would, for example, incorrectly exclude single cats from being judged as cats.

In sections 5 and 6 we will consider two reasons when or why the type labelled s_{base} (the Ind_P type) is unavailable as a counting base for other nouns.

5 Counting-Context Sensitivity, Overlap, and Disjointness

In standard mereological approaches, overlap (not-disjoint) is a higher-order property of sets. Within our type theoretic paradigm, we will define it as a higher order type (a type of types). In the case of concrete nouns this will be a type of type of individuals. Other than this difference in approach, disjointness may be defined in a relatively standard way. However, one further added complexity is how the probabilistic aspect of our formalism interacts with the

mereology. We introduce a (possibly context sensitive) probability threshold θ above which agents make judgements. A type is disjoint if all entities judged with sufficient certainty to be of that type are disjoint. For those types which have no clear instances, disjointness is undefined (one should not make a judgement either way with respect to disjointness). The intuitive idea here is that one cannot judge something to be disjoint or overlapping with respect to, say, a predicate, if one is not at all certain what falls under the predicate.

Definition 1. *A type T is disjoint relative to a probability threshold θ ($Disj_\theta$):*

IF *there is at least some a such that $p(a : T) \geq \theta$,*
 THEN *$T : Disj_\theta$ iff, for all a, b such that $p(a : T) \geq \theta$ and $p(b : T) \geq \theta$,*
if $a \neq b$, then $a \cap b = \emptyset$,
 ELSE *Undefined.*

We follow Landman [12] in making the grammatical counting function sensitive to disjointness. We also assume that the function applies to the type in a lexical entry labelled s_{base} (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:

$$f_{count, \theta} : (RecType : Disj_\theta \rightarrow NatNum) \quad (10)$$

This type restriction means that the counting function is only defined for types that are disjoint (relative to some probability threshold).

For prototypical count nouns such as *cat*, *woman*, and *chair*, the types for the counting base are Ind_{cat} , Ind_{woman} , and Ind_{chair} , respectively. These are not overlapping. Thus they are defined for grammatical counting.

There are two classes of data that we need to explain, namely, the mass/count variation in *collective artifacts* and in *homogenous objects*. We do this by showing how context sensitivity with respect to individuation schemas results in a tension between the pressures of individuation and reliability.

5.1 Collective Artifacts

For mass nouns such as *furniture*, *kitchenware*, *fencing*, and count nouns such as *huonekalu-t* ('furniture', Finnish), *Küchengerät-e* ('kitchenware', German), and *fence*, the story is a little more complex. In Sutton and Filip [19] we suggested a treatment for neat mass nouns (*furniture*, *kitchenware*) and their count-counterparts. Here, using our more developed formal apparatus, we extend this analysis to context-sensitive semantically atomic nouns analyzed in Rothstein [20] (*fence*, *hedge*), and their mass-counterparts (*fencing*, *hedging*).

As we argued in sections 2.2 and 2.3, for both of these groups of nouns, the difference between mass and count encoding can be seen as involving either the non-resolution of overlap at a general context ('counting as one simultaneously and in the same context'), or as the resolution of overlap at a specific context. One aspect of Rothstein's [20] and Landman's [12] work that we suggested

could be further developed is an account of what counting contexts are. Here we further develop the inchoate suggestion made in Sutton and Filip [19] that counting contexts can be modeled as schemas of individuation (formally modeled as quantitative functions). Furthermore, that, under pressure from individuation, variation in how we interact with the denotations of such nouns leads us to develop distinct individuation schemas (quantitative functions) and thereby distinct Ind_P types. We will give two examples: *furniture*_{-C} vs. *huonekalu*-t_{+C} (‘furniture’, Finnish), and *fencing*_{-C} vs. *fence*_{+C}.

*furniture*_{-C} vs. *huonekalu*-t_{+C}: Informally speaking, when learning what counts as ‘one’ with respect to *furniture* (or what counts as ‘one’ with respect to *huonekalu*), one is faced with inconsistent evidence. For example, vanity tables seem to be single items of furniture, but so do the framed mirrors that can be part of them. This creates a categorization problem, since BOTH the part and the whole should not be counted as one (even if both seemingly do count as one). This variation creates a conflict. A single individuation schema, represented as one quantitative function, would not be a reliable indicator of what counts as one item of furniture across contexts, since, for example, a single schema might correctly exclude counting the mirror in the vanity context, but incorrectly exclude counting such mirrors in other contexts. To remedy this, one must adopt different schemas in different contexts meaning that no one schema is wholly reliable. Hence, prioritizing the pressure towards individuation gives rise to unreliability.

To accommodate the pressure towards reliability, one could form a generalized individuation schema (formed from all admissible quantitative functions). This generalized schema would be a reliable indicator, since at every context, what counts as one would be included by at least one of the individuation schemas. In terms of the probabilistic semantics, this would mean that the conditional probability of correctly applying *furniture*, given the individuation schema would be very high. However, the generalized schema would no longer individuate since it would include as in counting as ‘one’ all entities that could count as one irrespective of whether they overlap (it would include the vanity table and the mirror that is a part of it). No longer individuating, in information theoretic terms, means carrying a lower informational value than a expression that transmits a single individuation schema, since the more general schema is equivocal between all admissible specific schemas. Hence, prioritizing the pressure towards reliability gives rise to less individuation.

For lexical items in this class, languages may, seemingly as a matter of convention, take one of two paths: prioritize individuation (at the expense of reliability), but allow the individuation schema to vary across situations; or prioritize reliability (at the expense of individuation), and form a generalized schema to cover all situations. We now formally outline how these two paths may be represented, then we show how the choice of path leads to a difference in mass/count encoding.

Formally speaking, for each noun where a clash of pressures arises, multiple quantitative functions are inferred by a learner. For example, with *furniture*, one function will map the type of situation which includes a vanity to the value 1 (the

vanity as a whole counts as one). A different function which will map this same type of situation onto the value 2 (for the table and the mirror to be counted separately). In the later case, the same function would map the type of situation containing just the table (without mirror), or just the mirror (without table) to the value 1. Since our terminology $Ind_{furniture}$ is just shorthand for the type of situation where some entity receives a quantitative function value of 1, we can describe there being *two* functions in terms of an agent tracking two $Ind_{furniture}$ types. Call these $Ind_{furniture,1}$ and $Ind_{furniture,2}$. When more than one Ind_P type is being tracked, there are two strategies available for classifying individual P -items:

1. Prioritize individuation. For the case in hand, *furniture*, one could either apply only one type in any given instance. However, as noted above neither $Ind_{furniture,1}$ nor $Ind_{furniture,2}$ is reliable. To remedy this, one could make the choice of individuation schema context sensitive, namely to sometimes apply $Ind_{furniture,1}$ and sometimes applying $Ind_{furniture,2}$.
2. Prioritize reliability. To do this one need merely form a more generalized type to cover all cases. This would obviate the need to add in context sensitivity. In TTR, a more generalized type can be formed via a disjunction (or join) between types as shown in (11).

$$Ind_{P,gen} = Ind_{P,1} \vee Ind_{P,2} \vee \dots \vee Ind_{P,n} \quad (11)$$

However, now the generalized schema does not fully individuate since it equivocates between whether a sum counts as one or more than one item of furniture.

The availability of a ‘choice’ of which pressure to prioritize explains mass/count variation via a difference in lexical entries for mass nouns such as *furniture* (12) as opposed to cross linguistic count-counterparts such as the Finnish *huonekalu* (‘item of furniture’) (13).

$$\llbracket \text{furniture} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{\text{body}} : [s_{furn} : furn(r.x)] \\ s_{\text{base}} : [r.x : Ind_{furn,gen}] \end{array} \right] \quad (12)$$

$$\llbracket \text{huonekalu} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{\text{body}} : [s_{furn} : furn(r.x)] \\ s_{\text{base}} : [r.x : Ind_{furn,i}] \end{array} \right] \quad (13)$$

The reason these entries lead to the mass encoding of *furniture*, but the count encoding of *huonekalu* is due to the semantic qualities of the type for the label s_{base} in each case. In (12), the type $Ind_{furn,gen}$ is not disjoint. This is because, for example, both a dressing table (including mirror) and a dressing table (excluding mirror) will be of this type. Other examples of overlap include tables that are pushed together (are they one or many tables?), and chairs with cushions (should the chairs be counted separately from the cushions or together?). Non-disjoint types are not defined for the counting function (10), and so *furniture* is mass. In contrast, because, in (13), *huonekalu* is encoded to select a specific quantitative function (determined, for example, by the context of use), each type $Ind_{furn,i}$ is

disjoint. As such, *huonekalu* will be defined for counting. That said, from context to context, the counting result may vary. In some contexts, the dressing table (including mirror) will count as one *huonekalu*, in others it may count as two.

This pattern in which counting results may differ from context to context should sound familiar from the case of *fence*. Recall Rothstein’s example of a square field enclosed by fencing. Whether we count this as one fence around the field, or two, three or four may depend on the context. We are able to use exactly the same tools as we use for *furniture* versus *huonekalu* to model this. The entry for *fence* is given in (14) and the entry for *fencing* is given in (15).

$$\llbracket \text{fence} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{\text{body}} : [s_{\text{fence}} : \text{fence}(r.x)] \\ s_{\text{base}} : [r.x : Ind_{\text{fence},i}] \end{array} \right] \quad (14)$$

$$\llbracket \text{fencing} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{\text{body}} : [s_{\text{fence}} : \text{fence}(r.x)] \\ s_{\text{base}} : [r.x : Ind_{\text{fence},gen}] \end{array} \right] \quad (15)$$

The reason these entries lead to the count encoding of *fence*, and the mass encoding of *fencing* parallels that of the previous case. Given that, at any context, the entry for *fence* selects a single quantitative function, the type $Ind_{\text{fence},i}$ is disjoint, and so defined for counting, even if the exact result of counting the same portion of fencing may result in different answers across contexts. In contrast, *fencing* does not distinguish between contexts and is defined in terms of more generalized join type $Ind_{\text{fence},gen}$ that is not disjoint. The reason that it is not disjoint is that, for example, in Rothstein’s square field example, the sum of four fence sides is of type $Ind_{\text{fence},gen}$, but so are the four fence-sides taken individually. Non-disjoint types are undefined for countability, and so *fencing* is mass.

Furthermore, these different conceptions are driven by which pressure is prioritized. If one prioritizes individuation, then the pressure is to find a single counting schema (at least in a context) from the possible schemas. However, in order to be reliable, the schema one uses must be context sensitive. This means that at each context, one has a non-equivocating individuation schema from the set of possible schemas. Choosing a single one (at a context) is maximally informative, thus the pressure of individuation is satisfied. On the other hand, one can prioritize reliability and adopt a generalized schema that ($Ind_{\text{fence},gen}$) that is a reliable indicator of when to apply the number neutral predicate *fence*. However, this generalized schema does not fully satisfy the pressure of individuation since it equivocates between specific schemas.

In this section we have argued that counting-context sensitivity gives rise to a competition between the pressures of individuation and reliability. Prioritizing one of these pressures over the other seems to be a matter of convention. Prioritizing individuation yields count encoding. Prioritizing reliability yields mass encoding. With this form of context-sensitivity, we cannot yet explain count/mass variation in granular nouns such as *lentil*, *rice* which we have assumed have disjoint Ind_P types (the types for single rice grains and single lentils). Nor can we, at this point, say anything about substance mass nouns such as *mud* and *air*. For this, we will need to appeal to another form of context-sensitivity, one related

to vagueness. In section 6, we will argue that vagueness can also lead to a clash between the pressures of individuation and reliability and so also to variation in mass/count encoding.

6 Contextual Variation and Vagueness

The conception of vagueness we adopt is based loosely on Sutton ([17], [18]). On this conception, vagueness is represented as a form of metalinguistic uncertainty that arises, in part, from inconsistent learning data. For example, for color predicates, we have good evidence for judging canonical cases of green as ‘green’, and likewise for blue. Towards the blurred boundary between green and blue, we either have a dearth of evidence for making ‘blue’/‘green’ judgements, or we have conflicting information (sometimes a shade will be described as ‘blue’, sometimes not). Either way, we infer a distribution that describes a gradual trailing off of the probability of a competent speaker making a ‘blue’ judgement as the shade of the object in question becomes ever greener, *mutatis mutandis* for ‘green’.

Following Chierchia ([2]) we argue that a similar mechanism affects the semantic representations of some nouns, however, that the graded increase in uncertainty varies with the output of the quantitative function. This mechanism is again a form of context sensitivity. The variation in what counts as, for example, *rice*, across contexts yields metalinguistic uncertainty (vagueness) with respect to what quantity of rice-stuff is sufficient to classify that stuff as *rice*.

6.1 Granular Nouns

The context-sensitivity of granular and substance nouns differs from that of collective nouns such as *kitchenware* and *furniture*. As Chierchia [2] observes, our judgements about whether granular and other substances are in the denotation of a given predicate vary depending on their amount in a given context. For example, whether we are willing to accept that we have mud on our shoes varies with context. In clean-room manufacturing or scientific contexts, even small specks of mud count as *mud*, because the tolerance for even tiny quantities of mud is near zero. In contexts like entering the apartment after a walk, our tolerance for mud is much higher, and in contexts like entering the garden shed it is even higher. For nouns such as *rice* or *lentils* one could truly say that we do not have any rice/lentils for dinner when only a few grains/lentils remain in the packet, but equally truly say that some rice/lentils fell on the floor during a meal even though the number of grains/lentils may be identical in both cases. Context matters. However, from a probabilistic learning perspective, these cases provide inconsistent data with respect to the categorical application of classifiers such as *mud*, *rice*, and *lentils*. The rational response for a learner (aside from seeking aspects of the contexts to explain this variation) is to lower the confidence with which she would apply the predicate for the specific amount of mud/rice/lentils in question. We model this as a Bayesian update given the

judgement set. The judgement set consists of situations (which can be understood as contexts from a situation theoretic point of view) and probabilistic type judgements made about those situations (contexts). In other words, the agent calculates the probability of applying e.g. the *rice* conditional with respect to the context with some quantity of stuff with the appropriate rice qualities. This is represented in (16) for some quantity value of 10. The value 0.5 would reflect the borderline case where the agent has as much reason to classify some quantity (of grains) of rice as *rice* as she has reason to judge them not to be rice.

$$p_{A,\bar{3}}(r : \left[\begin{array}{l} x \quad : *Ind \\ s_{rice} : rice(x) \end{array} \right] \mid r : \left[\begin{array}{l} s_{rice_stuff} : \left[\begin{array}{l} x \quad : *Ind \\ s_{rice_qual} : riceQual(x) \end{array} \right] \\ f_{rice_quant} : \left(\left[\begin{array}{l} x \quad : *Ind \\ s_{rice_qual} : riceQual(x) \end{array} \right] \rightarrow \mathbb{N} \right) \\ i \quad : \mathbb{N} \\ s_{rice_quant} : f_{rice_quant}(s_{rice_stuff}) = 10 \end{array} \right]) = 0.5 \quad (16)$$

For nouns such as *rice*, numerical values need not be taken to align perfectly with numbers of grains. For higher values, the output of the function could just as easily indicate some range of numbers of grains as some specific number. Either way, uncertainty about whether to apply the rice predicate will increase with smaller quantitative function values. This means a gradual increase of uncertainty about applying the predicate as quantities of rice get smaller. The idea that this represents is simply that one is safer, across contexts, using *rice* to describe larger quantities (a bowlful, a whole packet) than much smaller quantities (a grain, a few grains). The uncertainty involved in using the predicate across these cases reflects this.

Unlike with nouns such as *furniture* and *kitchenware* as well as with *fence* and *fencing*, this uncertainty is not about what counts as one (leading to a proliferation in individuation functions), but uncertainty about how much rice is enough to safely form a *rice* judgement. Yet, similarly to the *furniture*, *kitchenware*, *fence* and *fencing* cases, mass/count encoding of granular nouns can be seen as arising from the competition between the pressures of reliability and individuation.

The pressure of individuation pushes in one direction, namely that, for nouns such as *rice* and *lentils*, the types for the counting bases of the nouns should be the types Ind_{rice} and Ind_{lentil} , respectively. For *furniture*- and *fence*-like nouns, there were multiple competing equally informative individuation schemas (e.g. one which counts the table and mirror as two and another schema that counts the table and mirror as one, a vanity). However, for granular-like nouns, there is really only one plausible individuation schema, namely, that which counts grains, flakes etc.¹⁰ However, the gradation in probability values in the representation of nouns such as *rice*, and *lentils* means that, types for lower quantity values such as 1 (represented as types Ind_{rice} , Ind_{lentil}) are not *reliable* indicators of

¹⁰ This may not be universally true. For example, grains that come in easily separable halves might have two viable schemas, one which counts halves and one which counts wholes.

when to apply *rice* or *lentils*. In other words, prioritizing individuation leads to a fall in reliability. This is because single grains of rice or single lentils will not qualify as *rice* or *lentils*, respectively, *reliably* in all contexts. A strategy of prioritizing individuation will simply enter the $Ind_{\mathcal{P}}$ type as the counting base. The lexical entry for granular nouns could resemble far more closely the one for *cat* in (9). This is what we suggest occurs for nouns such as the English *lentil* as in (19). Individuation is prioritized since types such as Ind_{lentil} are disjoint, but reliability is forfeit since this type is not a wholly reliable indication of when one may apply the predicate $lentil(x)$.

The pressure of reliability pushes in the opposite direction to the pressure of individuation for granular nouns. Prioritizing reliability militates against taking, for example, the type for single grains of rice (Ind_{rice}) or single lentils ($Ind_{lentils}$) as a counting base. Recall that reliability entails finding a counting base such that the probability of (correctly) applying a predicate is high given that some entity is of that type specified in the base. One way to boost this probability and so prioritize reliability, as we find with the English *rice*, is to lexically encoding the counting base not with the type Ind_{rice} , but with the less specific predicate *rice* as in (17). Reliability is maximized here since, trivially, $p_{A,\exists}(a : T|a : T) = 1$, and so the type labelled s_{base} in (17) is a perfect predictor of the type labelled s_{body} . On this strategy, individuation is forfeit, since those entities which perceptually saliently count as one (such as individual rice grains), are not clear cases of the predicate *rice* across contexts.

In summary, for nouns such as *rice* and *lentils*, the context sensitivity that gives rise to graded probability judgements for entities in terms of applying a predicate, given some qualitative properties *and* a quantitative function value, in turn, creates a conflict between the pressures of individuation and consistency. The result is to prioritize one pressure. If one prioritizes reliability, the base does not individuate. Examples are given in (17) for the English *rice* and in (18) for the Bulgarian mass noun *lešta* ('lentil'). If one prioritizes individuation, the base is simply the relevant $Ind_{\mathcal{P}}$ type. An example of this is given in (eq:lentilpair) for the English 'lentil'.

$$\llbracket \text{rice} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{body} : [s_{rice} : rice(r.x)] \\ s_{base} : [s_{rice} : rice(r.x)] \end{array} \right] \quad (17)$$

$$\llbracket \text{lešta} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{body} : [s_{lentil} : lentil(r.x)] \\ s_{base} : [r.x : lentil(r.x)] \end{array} \right] \quad (18)$$

$$\llbracket \text{lentil} \rrbracket = \lambda r : [x : *Ind] . \left[\begin{array}{l} s_{body} : [s_{lentil} : lentil(r.x)] \\ s_{base} : [r.x : Ind_{lentil}] \end{array} \right] \quad (19)$$

The difference between (17) and (18) on the one hand and (19) on the other is in the type for the label s_{base} . In (19), the type Ind_{lentil} is a disjoint type and so is suitable for counting. Hence *lentil* is count. In (17), the type for the labels s_{body} and s_{base} are the same. Depending on the probability threshold, this type contains parts of grains, grains, or collections of grains of rice and sums thereof. As such, the type labelled s_{base} is not disjoint, and so is not defined for grammatical counting.

6.2 Substance Nouns

As we stated above, substance nouns like *mud* are vague in the same way as granular nouns in that what counts as *mud* varies from context to context, thus generating an inconsistent set of evidence for what counts as *mud*. We may assume, therefore, that the same ways of balancing the pressures of reliability and individuation that we employed for vague granular nouns like *rice* and *lentil* could be adopted for substance nouns, namely one of the two entries (20) or (21).

$$\llbracket \text{mud} \rrbracket = \lambda r : [x : *Ind]. \left[\begin{array}{l} s_{\text{body}} : [s_{\text{mud}} : \text{mud}(r.x)] \\ s_{\text{base}} : [r.x : Ind_{\text{mud}}] \end{array} \right] \quad (20)$$

$$\llbracket \text{mud} \rrbracket = \lambda r : [x : *Ind]. \left[\begin{array}{l} s_{\text{body}} : [s_{\text{mud}} : \text{mud}(r.x)] \\ s_{\text{base}} : [s_{\text{mud}} : \text{mud}(r.x)] \end{array} \right] \quad (21)$$

Prioritizing reliability yields the entry in (21) which would lead to the mass encoding of *mud* for the same reason as we got a mass encoding for *rice* in English. The type for the label s_{base} is not disjoint.

In contrast to *lentil*, however, the entry in (20) will not yield count encoding. For object count nouns, collective artifacts, and granular nouns (where the granules are not too small) there is relatively clear perceptual and/or functional based evidence for establishing what counts as ‘one’ item in the denotation of the relevant noun. In probM-TTR terms that means that for such a predicate P , there are at least some objects a , such that an agent is able to judge that $a : Ind_P$ with a reasonably high probability. This is not the case for substance, liquid and gas nouns. Unlike nouns like *cat* and *rice*, the denotations of these nouns are such that there is little, perceptually speaking, to aid in the identification of salient individuated units. Unlike nouns such as *chair* and *furniture*, nor do the denotations of substance nouns typically get partitioned in terms of function. This distinction in itself can be viewed as a further form of vagueness: what the perceptually/functionally salient entities of substance noun denotations are is highly uncertain.

In terms of reliability and individuation, this, in contrast to the granular case, means that types such as Ind_{mud} fail to carry a sufficiently high informational content (fail to specify a sufficiently specific portion of mud such that that portion would count as *one unit of mud*). Furthermore, unless a language imports a significant amount of context-sensitivity in what counts as an individuated mud unit (as could be argued is the case in languages such as Yudja), the pressure of individuation cannot be satisfied. We therefore would expect (21) and not (20) to be the lexical entry for *mud*. Put another way, unless made radically dependent on the context of application, the type Ind_{mud} is simply not useful since it is neither a good indicator for the applicability of *mud* (not consistent), nor does it convey a high enough informational content (does not individuate).

In probM-TTR terms that means that for such a substance/liquid predicate P , there are no objects a , such that an agent is able to judge that $a : Ind_P$ with a high probability. With respect to the disjointness (Definition 1), this means that types such as Ind_{mud} are undefined for disjointness. Since the counting function

requires a disjoint type as input, this means that substance nouns such as *mud* will be encoded as mass, even if their lexical entries are of a similar form to (20).

7 Conclusions and Summary

We hypothesized that there are two competing pressures on natural language predicates: (i) to individuate (recast partly in information-theoretic terms as being informationally rich); (ii) to find a reliable criterion for counting (a criterion which reliably predicts the type for the whole extension of P , modelled as a high conditional probability that something is of the body type, given that it is of the base type).

Inductive evidence for this hypothesis is provided by the predictions it makes with respect to the variation in the mass/count encoding. We show that the ways in which these two pressures can (or cannot) be satisfied in dependence on the different types of context-sensitivity represented in our formal model, predict the expected range of constraints on the variation in the mass/count encoding. In addition, this allows us to cover a broader range of data than other leading accounts.

Prototypical object nouns: 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 on individuation and reliability work in the same direction, i.e., they converge on the count encoding. We, therefore, have no reason to expect much, if any, mass encoding, cross- and intralinguistically.¹¹

Collective and homogeneous object nouns: Context-sensitivity with these nouns affects the reliability with which individual types apply. For example, across contexts, a sum can count as one fence, one item of kitchenware or two fences, two items of kitchenware. This means that any particular individuation schema will inconsistently determine the extension. To prioritize 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 prioritize reliability, all individuation schemas can be merged together. This yields a non-disjoint schema and so mass nouns such as *fencing* and *kitchenware*.

Granular nouns: 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 individual 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 prioritizes individuation, is inconsistent as a basis for applying a noun. Prioritizing individuation yields a count noun encoding, which is commonly presupposed by pluralization, e.g.

¹¹ One possible counter example to this is Brazilian Portuguese which seems to encode mass readings of most or even all object count nouns when used in the bare singular. For example, the bare singular 'How much book...?' can get a non-coerced measure (weight) reading. See [16].

lentils, *kaurahiutale-et* ('oatmeal' Finnish), *oats*. On the other hand, prioritizing reliability yields a non-disjoint individuation schema, and so leads to a mass noun encoding, as in *oatmeal*, *kaura* ('oats', Finnish), *čočka* ('lentils', Czech).

Substance nouns: Similarly as with granular noun denotations, context-sensitivity 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 do not easily enable the prioritization of individuation that could be achieved for count granular nouns.¹² If individuation cannot be prioritized, then reliability will be prioritized, therefore, we expect a heavy tendency towards mass encoding for these nouns.

Our formal account can capture these competing pressures either in terms of how sharply and specifically (as opposed to generally and vaguely) types relate to entities in the world. Our link to learning models also allows us to describe how (un)reliability can arise out of a process of classifier learning. Together, this means that we are not only able to formally represent noun meanings and countability, but we have also outlined the general mechanisms that give rise to the variation in the mass/count encoding.

References

1. Chierchia, G.: Plurality of nouns and the notion of semantic parameter In: Rothstein, S. (ed.) *Events and Grammar*, pp. 53–103. Springer, Berlin Heidelberg (1998)
2. Chierchia, G.: Mass nouns, vagueness and semantic variation. *Synthese*. 174, 99–149 (2010)
3. Cooper, R.: Frames in Formal Semantics. In: Loftsson, H., and Rögnvaldsson, E., and Helgadóttir, S., (eds.) *Advances in Natural Language Processing LNCS vol. 6233*, pp. 103–114. Springer, Berlin Heidelberg (2010)
4. Cooper, R., Dobnik, S., Lappin, S., and Larsson, S. A probabilistic rich type theory for semantic interpretation. *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics*, pp. 72–79 (2014)
5. Cooper, R., Dobnik, S., Larsson, S. and Lappin, S. Probabilistic type theory and natural language semantics. *LILT* 10(4), 1–43 (2015)
6. Cooper, R.: Type theory and semantics in flux. In: Kempson, R., Fernando, T., Asher, N., (eds.) *Philosophy of Linguistics*, pp. 271–323. Elsevier, Amsterdam (2012)
7. Fillmore, C. J.: Frame semantics and the nature of language. *Annals of the New York Academy of Sciences* 280(1), 20–32 (1976)
8. Geach, P. T.: *Reference and Generality: An examination of some medieval and modern theories*. Cornell University Press (1962)
9. Kirby, S. and Hurford, J.: The Emergence of Linguistic Structure: An overview of the Iterated Learning Model. In: Cangelosi, A. and Parisi, D. (eds.) *Simulating the Evolution of Language*, p 121–148. Springer Verlag, London (2002)
10. Krifka, M.: Nominal Reference, Temporal Constitution and Quantification in Event Semantics In Bartsch, R, van Benthem, J. F. A. K. and van Emde Boas, P. (eds.), *Semantics and Contextual Expression*. pp. 75–115. Foris Publications (1989)

¹² However, see the caveat about Yudja in section 6.2.

11. Krifka, M.: Common nouns: A contrastive analysis of English and Chinese. In: Carlson, G. and Pelletier, F. J. (eds.) *The Generic Book*, pp. 398–411. Chicago University Press, Chicago (1995)
12. Landman, F.: Count nouns – mass nouns – neat nouns – mess nouns. *The Baltic International Yearbook of Cognition* 6, 1–67 (2011)
13. Landman, F.: Iceberg Semantics for Count Nouns and Mass Nouns: the evidence from portions. Unpublished Handout (2015)
14. Lima, S.: All notional mass nouns are count nouns in Yudja. *Proceedings of SALT 24*, 534–554 (2014)
15. Piantadosi, S. T., Tily, H., and Gibson, E.: The communicative function of ambiguity in language. *Cognition* 122, 280–291 (2011)
16. Pires de Oliveira, R. and Rothstein S.: Bare singular noun phrases are mass in Brazilian Portuguese *Lingua* 121, 2153–2175 (2011)
17. Sutton, P. R.: Vagueness, communication, and semantic information. PhD Thesis, King’s College London (2013)
18. Sutton, P. R.: Towards a semantics for vague adjectives. In: Zeevat, H., and Schmitz, H-C. (eds.) *Bayesian Natural Language Semantics and Pragmatics*, pp. 221–246. Springer, Berlin Heidelberg (2015)
19. Sutton, P. R., and Filip, H.: Vagueness, Overlap, and Countability. Forthcoming in: *Proceedings of Sinn und Bedeutung* 20 (2016)
20. Rothstein, S.: Counting and the mass/count distinction. *Journal of Semantics*. 27, 343–397 (2010)