Chapter 7

Critical typicality: truth judgements and compositionality with plurals and other gradable concepts

Yoad Winter

Abstract

Compositional semantic frameworks often compute the extension of a complex expression directly from the extensions of its parts. However, much work in cognitive psychology has shown important challenges for compositional methods. For instance, Hampton (1988*b*) showed that speakers may let the complex nominal *sports that are games* include *chess* as one of its instances, without admitting chess in the extension of *sports*. Similarly, Lee (this volume) experimentally supports the common intuition that instances of *red hair* are not necessarily categorized as *red*. This paper reviews further results about *plural quantifiers*, showing similar challenges for compositionality. It is proposed that typicality effects play a systematic role in compositional interpretation and the determination of truth-values. For instance, the "overextension" effect in the *red hair* example is predicted by the fact that focal red is an atypical hair color. Similarly, in the plural sentence *the men are walking and writing*, the availability of the split reading ("some men are walking and some men are writing") increases due to the atypicality of doing both activities at the same time (Poortman, this volume). Further, in reciprocal sentences like *the three men are pinching each other*, the number of pinching acts may be three. This is related to the atypicality of situations where every man pinches two other men at the same time, as required by a strong interpretation of *each other*. The paper gives a uniform account of truth-value judgements on these different constructions, based on the identification of conflicts between typical preferences.

1 Introduction

One of the biggest challenges for semantic theory is the tension between contextuality and compositionality. To take a simple example, suppose that someone tells us that *A and B cost the same*. We may reasonably conclude that if A is expensive, then B is expensive as well. However, as a general inference rule, this conclusion is too hasty. We realize that when we consider the possibility that A is a laptop and B is a car. At some level, our initial conclusion was justified: without doubt, price is the most salient parameter in our interpretation of the concept EXPENSIVE. However, as soon as we become aware of further information about the objects A and B, we may easily discharge conclusions that were drawn on the basis of price alone. A theory of reasoning that ignores such contextual information, e.g. by relegating it to some underspecified pragmatic variables, runs the risk of becoming empirically empty.

It is important to note that the same risk threatens formal semantic theories of meaning composition. When analyzing the compositional interpretation of phrases like *my expensive laptop* and *my expensive car*, we again have to address problems about contextuality.

1

Here, the interpretation of the adjective *expensive* is most directly affected by its immediate linguistic context: the head noun – *laptop* or *car*. The effect that these nouns have on the meaning of *expensive* is similar to the contextual effect that the identity of A and B have in the example above. Therefore, a theory of semantic composition must pay close attention to contextual effects, and not only consider the formal description of language structures. In view of this point, some researchers in cognitive psychology doubt that formal semantic theories, with their arcane symbolic methods, have a prominent role to play in the analysis of concept composition (Barsalou, this volume).

However, there is another side to the compositionally problem, which indicates that formal theories cannot so easily be dismissed. Consider the distinction between the phrase *expensive laptops and cars* and the phrase *laptops and expensive cars*. The former expression is ambiguous (it may or may not be about expensive cars) whereas the latter is unambiguous (it must be about expensive cars). The ambiguity in the former phrase is not simply a matter of the larger prosodic distance between the words *expensive* and *cars*. This can be seen by considering the phrase *expensive laptops in cars*, which unambiguously refers to expensive laptops, not to expensive cars. Clearly, the syntactic nature of the construction – a coordinate structure or a subordinated prepositional phrase – has an important impact on potential ambiguities when concepts evoked by language are put together. To understand such effects we need to rely on syntactic theory. This is where notions of compositionality from formal semantics become important. As we move on to more complex quantificational structures, temporal and spatial expressions, anaphoric items etc., we increasingly need to rely on structural analyses and symbolic principles to explain the close interactions between form and meaning (Pelletier, this volume).

One of the motivations for the present work is the conviction that both contextual considerations and formal considerations are important for analyzing meaning composition. More specifically, this paper proposes that compositional theories should rely on *typicality effects*. These effects are critical for understanding concepts in general, and they have been playing an important role in the study of concept composition since the early works of Osherson and Smith in the 1980s. Notwithstanding, there is an on-going debate on whether typicality effects are relevant for the study of natural language semantics (Sassoon 2013). Following previous work, especially Kamp & Partee (1995), this paper aims to examine the role that typicality effects play in formal semantics. I propose that using observations about typicality effects, we can systematically account for gradability phenomena in truth-value judgements. I claim that the distribution of judgements about "truth" and "falsity" of complex sentences is at least partially predictable from judgements about typical instances for the concepts that sentences evoke. Thus, although typicality effects are logically separate from truth-value judgements, they affect such judgements in a non-trivial way.

One might suppose that the more typical an object is for a category, the likelier it is to be categorized in that category. However, part of the interest in typicality phenomena is that their interactions with truth-value judgements are much subtler than that. For instance, as Lee (this volume) shows, ginger-like hues are much more often accepted for *red hair* than they are for *red car*. As we will see, similar effects appear in other areas of semantic theory, specifically reciprocal expressions ('each other'), and conjunctive plural predicates ('are big and red', Poortman, this volume). In all three cases, a gradable concept combines with another concept, with possibly conflicting typicality preferences. Capitalizing on this parallelism, the paper employs one unified principle for explaining recent experimental results

in the three domains. Section 2 reviews familiar notions about vagueness of gradable concepts. Section 3 uses these notions for also modeling some vagueness effects with plurals. Section 4 addresses familiar motivations to distinguish vagueness from typicality. Section 5 explains the proposed approach in relation to the familiar problem of typicality effects with gradable adjectives ('expensive car', 'red car'). Section 6 applies the same approach to plurals, specifically reciprocal expressions ('each other') and conjunctive predicates ('big and red'). Finally, section 7 puts the pieces together, by analyzing the effects of typicality on truth-value judgements. Section 8 concludes.

2 Vagueness and acceptability functions

"Truth" and "falsity" are often seen as the pinnacles of natural language semantics. However, truth-conditional semantics must address the fact that speakers' truth-value judgements about sentences are often inherently non-uniform, in a way that cannot be explained by ambiguity or by multiple processing strategies. Consider for example a simple sentence like *this hue is red*. For monochromatic light at 610nm, between typical red and typical orange, speakers may disagree on whether the sentence is 'true' or 'false', or find it hard to decide between the two possibilities (Bonini et al. 1999). The phenomenon is commonly referred to as *vagueness*.[1] We say that vague sentences such as *this hue is red* may have a varying degree of acceptability depending on the situation. Upon being shown a black hue, virtually all speakers (who are not color-blind) would judge the sentence as 'false'. Upon being shown a hue in the focal red area (around 650 nm), virtually all speakers would judge the sentence as 'true'. For the oranges, pinks, magentas etc., we take the acceptability to be a number between 0 and 1, according to the proportion of speakers who consider the sentence true. Proportion-based analyses of acceptability go back to Black's (1937) "consistency profiles". Black lumps between-speaker consistency together with within-speaker consistency. However, there are empirical distinctions between the two (McCloskey & Glucksberg 1978). I consider here acceptability as a between-speaker measure, because it is in line with the way truth-value judgements and typicality judgements are analyzed in the experiments cited throughout this paper.[2] However, since this paper concerns the connections between acceptability and typicality, it is only crucial that they are consistently measured. Other measures of acceptability and typicality may be used, and are expected to show similar behaviors to those studied here.

Since concepts like RED, EXPENSIVE, TALL etc. impose graded acceptability on situations, we refer to them as *gradable concepts* (GCs). Different theories of vagueness have different interpretations of what acceptability scores for GCs mean and how they should be treated. Here we only use them for reflecting experimental measures of truth-value judgements, without a commitment to any specific theory of vagueness. For convenience, we

---

[1] For a reader on vagueness in philosophy, see Keefe & Smith (1996). For linguistic work on vagueness, see Kennedy (2011) for a semantic overview, and Nouwen et al. (2011) for further work in linguistics and philosophy. See Hampton (1998, 2007), Alxatib & Pelletier (2011), Egré et al. (2013), Serchuk et al. (2011) for experimental work.

[2] Measuring speakers' judgements embodies an "internalist" approach to meaning (Pelletier, this volume). James Hampton (p.c.) remarks that speakers' acceptability judgements on expressions like *risk* may diverge quite substantially from "externalist" definitions of their meanings (e.g. based on probability theory). This divergence may reflect two different senses of the word *risk* – an "ordinary" sense and a "scientific" sense. This distinction hardly affects the concepts studied in this paper.

also ignore all sorts of questions about the non-linguistic context. Thus, for a sentence like *this hue is red* we may assume that the 'situation' is a simple color grid presented without any specific context, ignoring complex problems that may appear when the contextual information is more intricate (e.g. as in Egré et al. 2013). These assumptions are useful for addressing our main problem here, and they do not affect too much the proposed analysis.

Summarizing, we define acceptability functions as follows.[3]

(1)   *Given a gradable concept GC, the* **acceptability function** ACC_{GC} *is a function from situations to the real numbers in* $[0, 1]$.

In many cases, the order that GCs induce on situations is *upward-monotone* in the relevant dimension. To see what this means, consider for instance the following sentence.

(2)   This jacket is expensive.

The concept EXPENSIVE satisfies the following condition:

For any two situations S_1 and S_2 with the same jacket x, if the price of x in S_1 is less or equal to its price in S_2, then $\text{ACC}_{\text{EXPENSIVE}}(S_1) \le \text{ACC}_{\text{EXPENSIVE}}(S_2)$.

Monotonicity implies that the acceptability function for the concept EXPENSIVE induces a partial order on situations on the basis of the price. For instance, in sentence (2), suppose that we consider two situations: S_1, where the jacket costs \$50, and S_2, where it costs \$120. We expect sentence (2) to be at least as acceptable in situation S_2 as it is in S_1. Thus, when speakers are asked out-of-the-blue if sentence (2) is "true", we expect the proportion of positive responses for situation S_2 to be greater or equal than for situation S_1. Similarly, for the concept RED, the nearer the hue of an object x is to the definitely red area (say, around 650nm), the higher the acceptability of the sentence *x is red*. Note that not all gradable concepts are upward-monotone in this way. For instance, the concept MIDDLE-AGED is not upward-monotone in age: its acceptability may increase up to certain ages (say around 55), but at later ages it may start to decrease (say around 65). Thus, this acceptability is upward-monotone in some age groups, and downward-monotone in others.

3   Distributivity and reciprocity as gradable concepts

Noun phrases like *the children* and *the townspeople* were not traditionally studied as part of the puzzles surrounding vagueness in natural language. However, in some works on plurals it has been observed that such noun phrases also exhibit considerable vagueness when they combine with many simple predicates.[4] For instance, consider the following example by Lasersohn (1999).

(3)   The townspeople are asleep.

How many of the townspeople need to be asleep in order for sentence (3) to be considered "true"? When trying to answer such questions, we see that sentence (3) has a characteristic

---

[3] In semantics, the numeric values that acceptability functions return are often called "degrees" (Kennedy 2007). In works on concepts, degrees are often described in terms of "graded membership" in a category (Kamp & Partee 1995, Hampton 1998, 2007 among others).

[4] See Dowty (1987), Brisson (1998), Lasersohn (1999), Burnett (2012) for theoretical work, and Schwarz (2013), Križ & Chemla (2015) for recent experimental work.

behavior of vague sentences.[5] To capture such vagueness effects, we associate an accept-ability function ACC with the predicate "are asleep" in (3). To be precise, we associate this acceptability function with a concept DIST, for a *distribution quantifier*. This means that we roughly interpret sentence (3) as claiming that *many*, or *enough*, of the townspeople are asleep. As with these explicitly vague concepts, the acceptability of sentence (3) gets higher as more townspeople are asleep. This monotonicity is determined by the acceptability function for the DIST concept. Explicitly:

- If each townsperson in S is asleep: $\text{ACC}_{\text{DIST}}(S) = 1$

- If no townsperson in S is asleep: $\text{ACC}_{\text{DIST}}(S) = 0$

- ACC_{DIST} is upward-monotone with respect to containment between sets: if the set of townspeople is the same in situations S_1 and S_2, and the set of sleeping townspeople in S_1 is contained in the set of sleeping townspeople in S_2, then $\text{ACC}_{\text{DIST}}(S_1) \leq \text{ACC}_{\text{DIST}}(S_2)$.

Similar vagueness appears with another plurality phenomenon, which is related to dis-tributivity: *reciprocity* (Langendoen 1978, Dalrymple et al. 1998). Consider the following sentence.

(4)   The three girls admire each other.

Intuitively, sentence (4) is perfectly true if each girl admires the other two. The less admi-ration relations there are, the less acceptable the sentence is. In experiments by Kerem et al. (2009) and Poortman et al. (2016), sentences similar to (4) were evaluated by participants in different situations, which are described schematically in Figure 7.1.

XXX Figure 7.1 approximately here XXX

Participants in Poortman et al.'s experiment were presented with schemes as in Figure 7.1 and were asked to judge if a given scheme is a "possible depiction of the sentence". Of the participants, 96%, 48% and 4%, respectively, judged schemes I6, I3 and I2 as possible descriptions for a Dutch translation of sentences like "A, B and C admire each other". We describe these facts using an acceptability function $\text{ACC}_{\text{RECIP}}$ for the reciprocity concept RECIP:

- In situation I6: $\text{ACC}_{\text{RECIP}}(I6) = 0.96$

- In situation I3: $\text{ACC}_{\text{RECIP}}(I3) = 0.48$

- In situation I2: $\text{ACC}_{\text{RECIP}}(I2) = 0.04$

---

[5]E.g. consider the vagueness criteria surveyed by Kennedy (2011). Sentence (3) has borderline cases, e.g. when 75% of the townspeople are asleep. The same sentence also demonstrates the "Sorites Paradox": if you consider (3) true for a situation with 10,000 sleeping townspeople, you'll find it hard to change your mind if one of them suddenly wakes up. Sentences like (3) have also been demonstrated to be context sensitive. E.g. consider Lasersohn's (1999) example "the subjects are asleep", in the context of an experiment where the purpose is that all the subjects fall asleep. In such cases the preference that all subjects are asleep becomes stronger. See Burnett (2012) for further discussion. Some authors refer to vagueness effects as in (3) as 'non-maximality' or 'homogeneity violation' (Križ & Chemla 2015).

- The acceptability function $\mathrm{ACC_{RECIP}}$ is upward-monotone with respect to containment between binary relations. Thus, if the set of three girls is the same in two situations S_1 and S_2, and the set of admiration pairs in S_1 is contained in the set of admiration pairs in S_2, then $\mathrm{ACC_{RECIP}}(S_1) \leq \mathrm{ACC_{RECIP}}(S_2)$.

The measured acceptability values for situations I2, I3 and I6 are consistent with the monotonicity assumption about $\mathrm{ACC_{RECIP}}$, since the set of arrows in I6 contains the set of arrows in I3, and the set of arrows in Figure I3 contains the set of arrows in I2.

We have considered three different kinds of gradable concepts (GCs) with adjectives (EXPENSIVE, RED), quantificational distributivity (DIST), and reciprocal quantifiers (RE-CIP). As we have seen, all three kinds of GCs are vague and display upward-monotonicity. With this background on GCs, we can address some problems surrounding their interpretation and typicality effects.

4 Acceptability vs. typicality

Speakers often judge an object to be an atypical member of a concept, but without denying its membership in the concept's extension, or *category* (Rosch 1973). As Kamp & Partee (1995, p.133) mention, if Robbie is a pelican then his acceptability as an instance of the concept BIRD ought to be close to 1. Notwithstanding, Robbie is clearly an atypical instance of this concept. Since Rosch and her associates' seminal works in the 1970s, this dissociation between acceptability and typicality has been illustrated by a variety of *typicality effects*. For instance, in McCloskey & Glucksberg's (1978) experiments, the following category-exemplar pairs received typicality ratings lower than 7 (on a 1-10 scale), but membership was considered positive by more than 90% of the participants:

ANIMAL: *cobra, lizard, woman*   BIRD: *penguin*   CARPENTER'S TOOL: *crowbar*   DISEASE: *alcoholism, schizophrenia*   INSECT: *louse, silkworm*   NATURAL EARTH FORMATION: *sinkhole* SCIENCE: *agriculture*   WEATHER PHENOMENON: *dew*   VEGETABLE: *rhubarb, soy bean*

This kind of empirical dissociation between measures of acceptability (=graded membership) and measures of typicality has been repeatedly demonstrated, and is hardly controversial. A bigger controversy surrounds the question whether typicality effects relate to the same mental faculties that affect vagueness and graded membership. Hampton (2007) describes one position in the debate by stating that "Osherson & Smith and Kamp & Partee contend [that] typicality T has quite different properties as a variable from degree of membership M". By contrast, Hampton himself (a.o.) proposes "that both degree of membership M and typicality T are based on a single underlying metric of similarity [to the prototype]". For more on this debate, see Rips (1989), Hampton (1998).

The present paper shows more evidence that typicality affects category membership.[6] Specifically, typicality for lexical concepts may influence graded membership in categories for complex expressions made of these lexical concepts. For a start, this section considers some examples where typicality is distinguished from acceptability, and which are specifi-

---

[6]Note however, that I remain agnostic regarding Hampton's claim that typicality and acceptability are determined by the same underlying metrics of similarity (to a prototype). Especially, the proposal below is neutral with respect to Hampton's assumption that typicality measures with *complex expressions* reflect the same metrics that affects truth-value judgements.

cally relevant for the rest of this paper. These examples concern the behavior of noun concepts, reciprocity and distributivity, as tested in recent experiments by Kerem et al. (2009), Lee (this volume), Poortman (this volume) and Poortman et al. (2016). These examples help us observe common principles in the complex interplay between acceptability and typicality.

Example 1: HAIR

In one of Lee's experiments (Lee, this volume: Figure 5.4), participants were shown a picture of hair that was atypically dyed focal red. 87% of the participants accepted this picture as an instance of the concept RED HAIR. Therefore, we may reasonably conclude that the stimulus was quite acceptable as an instance of HAIR.[7] However, in a color preference task for the concept HAIR, only 24% of the participants preferred focal red to a more natural hue for hair (RGB value: 201,113,13). In accordance with naive intuition, we may conclude that hair that is dyed focal red is an acceptable instance of HAIR, but is atypical for the category.

Example 2: SHAKE

In one of Kerem et al.'s experiments, participants were shown two pictures of a man shaking infant beds (Figure 7.2a-b), as well as an incomplete transitive sentence "the man is shaking (something)".[8] Although both pictures clearly show the man doing the activity reported in the sentence, 87% of the participants preferred Figure 7.2a to Figure 7.2b as a "better illustration for the sentence". We conclude that a situation with two patients is an atypical instance of the concept SHAKE, despite its high acceptability.

XXX Figure 7.2 approximately here XXX

Example 3: WALK **and** WRITE

In one of the experiments by Poortman (this volume), participants were instructed to rate the oddness of a situation where a person *walks* and *writes* at the same time. The average oddness rate for this pair of verbs was 3.39 on a scale between 1 and 6, where 1 meant "not odd at all" and 6 meant "physically impossible". We conclude that such a situation is atypical for the concepts WALK and WRITE. Nevertheless, the situation is highly acceptable as a legitimate instance for each of those concepts: when Lucy happens to be walking and writing at the same time, we do not expect many speakers to reject the description "Lucy is walking" just because the event is atypical for the concept WALK. Similarly we do not expect the sentence "Lucy is writing" to be rejected if she happens to be walking at the same time.

To describe the dissociation between judgements about vagueness and typicality, it is common to associate concepts with *typicality functions*. Similarly to our treatment of acceptability functions, we will use typicality functions that map situations to a value between

---

[7]Lee did not directly check the degree of acceptability for HAIR. Overextension effects (Hampton 1996) may mean that the actual acceptability for HAIR in the focal red stimulus was slightly below 87%. But this effect cannot be too big, since the pictures that Lee used were quite natural hair.

[8]The Hebrew example in Kerem et al.'s questionnaires was *ha-iŝ mena'ane'a* ("the-man shake"). The Hebrew verb only has the transitive meaning of the English verb *shake* ("cause to move"), not the intransitive meaning ("tremble").

0 and 1.[9] The typicality value assigned to a situation may be different than the value assigned by the acceptability function. Summarizing the judgements in examples 1-3 above, we make the following assumptions on the typicality and acceptability functions.

For the concept HAIR in Example 1, let S_1 be a situation with a focally red hair. Based on the observations above, we denote:

$$\text{ACC}_{\text{HAIR}}(S_1) \approx 1 \quad \text{and} \quad \text{TYP}_{\text{HAIR}}(S_1) \ll 1.$$

For the concept SHAKE in Example 2, let S_2 be the situation in Figure 7.2b. We denote:

$$\text{ACC}_{\text{SHAKE}}(S_2) \approx 1 \quad \text{and} \quad \text{TYP}_{\text{SHAKE}}(S_2) \ll 1.$$

For the concepts WALK and WRITE in Example 3, let S_3 be a situation where Lucy is walking and writing simultaneously. We denote:

$$\text{ACC}_{\text{WALK}}(S_3) \approx 1 \quad \text{and} \quad \text{TYP}_{\text{WALK}}(S_3) \ll 1, \quad \text{and similarly for WRITE.}$$

With this background on concepts, their vagueness and typicality effects, we can move on to the question of concept composition.

5 Guppy effects with gradable adjectives

In formal semantics it is commonly assumed that meanings of complex phrases are determined by a general principle, known as *compositionality* (Janssen 1997, Barker & Jacobson 2007, Werning et al. 2012). According to this principle, the meaning of a complex expression is determined by the meanings of its parts, and the way they combine with each other. For phrases like *a vegetarian student*, the simplest way to follow the compositionality principle is to use logical conjunction: a vegetarian student is someone who is independently categorized as both a student and a vegetarian. While in this example the conjunctive process works quite well, other modification constructions require a more complicated semantic or pragmatic treatment. Osherson & Smith (1981) study the implications of applying the simple conjunctive treatment to compounds like *pet fish*, as in the following analysis.

(5) An entity is categorized as a PET FISH if it is <u>independently</u> categorized as a PET and as a FISH.

According to O&S's intuitions, a guppy should be considered more typical for the category PET FISH than it is for any of the categories PET and FISH in isolation. This means that the typicality of an object for the category PET FISH cannot be easily determined using the object's typicality values for each of the two constituent categories independently. This kind of problem for understanding typicality was demonstrated in many experiments, starting in Smith & Osherson (1984). We collectively refer to such challenges about typicality as "guppy effects".[10] Applying O&S's claims to adjective-noun compounds, Smith & Osher-

---

[9]I here ignore the question of whether there is indeed a bounding value for typicality (Osherson & Smith 1981, Smith & Medin 1981, Smith & Osherson 1984, Hampton 1998, 2007). I also do not rule out the possibility that typicality and acceptability (at least at the lexical level) are determined by the same variable, as Hampton suggests. The data discussed in this paper are consistent with this possibility. The important point is the uncontroversial assumption that typicality rating may be different than acceptability.

[10]Somewhat ironically, Storms et al. (1998) report a failure to experimentally attest a guppy effect with the *pet fish* example of O&S. However, this failure only concerns O&S's specific example, and the so-called guppy effect was attested in many experiments, including Storms et al.'s.

son (1984) showed that objects may be more typical for such constructions than they are for the adjective and the noun in isolation. For instance, in S&O's experiment, brown instances of apples received higher typicality ratings for the expression *brown apple* than for each of the words *brown* and *apple* in isolation.

Despite the relevance of O&S's puzzle for the study of complex concepts, it should be noted that O&S themselves did not consider guppy effects to be a immediate problem for a compositional rule as in (5). O&S's approach is based on a sharp distinction between typicality and vagueness (Osherson & Smith 1997). Thus, according to O&S's view, category membership can be determined using a compositional rule like (5), while typicality judgements are governed by other, possibly non-compositional, mechanisms. However, O&S's sharp dissociation of typicality from vagueness is not easy to maintain. As many works have shown, there are important relations between typicality and category membership. McCloskey & Glucksberg (1978) showed that speakers often disagree about category membership when exemplars have intermediate typicality values, but they are less likely to disagree about membership if typicality is very high or very low. Furthermore, Hampton (1988*b*) showed guppy effects for typicality rating tasks, as well as for membership rating tasks. For instance, Hampton studied membership rating for complex expressions like *sport which is also a game*. The results showed that the item *chess* was more often categorized as a *sport which is also a game*, than it was categorized as both a *sport* and a *game* independently. This kind of *overextension effect* challenges the conjunctive rule in (5), and it was demonstrated with many items in relative clauses. Chater et al. (1990) obtained similar results to Hampton's using a Y/N category membership task, without membership rating. Similar effects were shown with adjective modifiers, noun compounds, disjunction and negation (Storms et al. 1996, Hampton 1988*a*, 1997).

The overextension effects that were studied in the literature normally dealt with conjunctive expressions where there is not necessarily a conflict between the conjoined categories. For instance, *golf* is both a highly typical sport and a highly typical game, and in Hampton's experiments it was commonly judged to belong in both categories (Hampton 1988*b*). Thus, there is no necessary competition between the typicality that exemplars have with respect to the concepts SPORT and GAME.[11] A similar point holds for many of the items studied in the literature on overextension. By contrast, O&S were interested in concept composition where one concept is in a "negatively diagnostic" relation with another concept (Smith & Osherson 1984, p.340). For instance, consider the inherent typicality conflict in the expression *red hair*. Hair instances that are typical of RED are atypical for the concept HAIR. Conversely, typical HAIR instances are of hues that are either not red at all or quite atypical for RED. Lee (this volume) checked the effects of such typicality conflicts on truth-value judgements. His experiments contrast combinations like *red hair*, where there is an intuitive conflict in typicality (S&O's "negatively diagnostic" relation), and cases like *red car*, where there is no typicality conflict (S&O's "nondiagnostic" relation). Accordingly, in Lee's experiment 2, participants were asked to give their judgements on two images:

> *Image 1*: an image of a woman whose hair is the same hue of Example 1 above (RGB value: 201,113,13 – Lee, this volume: Figure 5.4).

---

[11]In fact, many typical sports are also typical games. For this reason, Hampton's (1997) study of constructions like *sports that are not games* is exceptional in that it (indirectly) tests potential typicality clashes, here between *sports* and *non-games*.

Image 2: an image of a car that is painted the same hue as in Image 1.

The majority (92%) of the participants in Lee's experiment accepted the categorization *red hair* for Image 1, but only 17% accepted the categorization *red car* for Image 2. This means that for at least 75% of the participants, a simple conjunctive analysis as in (5) would not work without additional assumptions. We can standardly assume that the likelihood that an object is categorized as *red* is affected by the linguistic and non-linguistic context.[12]

Here we concentrate on the effects of the noun concept on the interpretation of contrastive pairs like *red hair/car*. Such contrasts show a simple enough illustration of the problem for the conjunctive rule. Therefore, from a theoretical perspective, they are useful for analyzing the relations between membership and typicality with complex expressions. Like other gradable concepts, the concept RED imposes a natural ordering on entities. Here the ordering is naturally based on hue, and can be expressed by the comparative statement *x is redder than y*. Let us now consider the concept CAR in the phrase *red car*. The relation between the concepts RED and CAR is what S&O classify as a "nondiagnostic" relation: the typicality of CAR instances is likely to remain by and large unaffected by changes in hue. More formally, suppose that we are given two situations S_1 and S_2 with a car, where the only difference between S_1 and S_2 is in the car's hue. We may reasonably assume that $\mathrm{TYP}_{\mathrm{CAR}}(S_1)$ is close to $\mathrm{TYP}_{\mathrm{CAR}}(S_2)$. Specifically, suppose that S_1 has a car painted focal red, and S_2 has a car painted some other hue, quite distant from focal red. Both cars are expected to be equally typical for the concept CAR.

Unlike the concept CAR, the concept HAIR clearly has more typical and less typical colors. For instance, various shades of black and brown are more typical for human hair than, say, shocking pink. Among the hues between orange and focal red, some hues, at the margins of the concept RED, may be categorized as quite typical for HAIR. These are the hues that are most common for hair that is classified as RED HAIR. Let us informally refer to these hues as 'ginger'. When we consider the complex expression *red hair*, we see a typicality conflict (S&O's "negatively diagnostic" relation), which is due to the effect of hue on the typicality for HAIR. Starting from those hues that we called 'ginger', the redder the hue gets, the lower the typicality is for the concept HAIR. We classify this effect as *downward-monotonicity* of the typicality function for the concept HAIR, and say the function $\mathrm{TYP}_{\mathrm{HAIR}}$ is downward-monotone relative to the order imposed by the gradable concept RED. More formally:

For any two instances x_1 and x_2 of HAIR, where x_1 and x_2's hues are between ginger

---

[12]For linguistic work, see Cresswell (1976), Kamp (1975), Kamp & Partee (1995), Klein (1980), Kennedy (2007). For experimental work, see Hansen et al. (2006), Kubat et al. (2009). Because of contextual effects, it is likely that some of the participants in Lee's experiment would accept the hair in Image 1 as *red* but reject the car in Image 2, even if the nouns would not be mentioned, relativizing redness to the visual instances of the concepts HAIR and CAR. Lee's work did not try to factor out possible effects of the visual stimuli, which are also central for concept composition (see Barsalou, this volume). However, Lee's effects in distinguishing gradable-concept categorization in complex expressions with "neutral" vs. "biased" categorizes are much stronger than in previous work on categorization with simple color terms. Therefore, we can maintain the assumption that the interpretation of an adjective is affected both by its visual context and its linguistic context. Below I ignore effects of the visual context because they are less directly relevant to the compositionality problem in linguistics. Most importantly, effects coming from the non-linguistic context are not immediately testable when analyzing reciprocity and distributivity, because these concepts, unlike adjective concepts, are not easily studied in isolation: unlike category names such as RED, reciprocity and distributivity are bound to appear as sentence parts.

and focal red: if x_1 is redder than x_2, then $\mathrm{TYP}_{\mathrm{HAIR}}(x_1) \leq \mathrm{TYP}_{\mathrm{HAIR}}(x_2)$

Note that this downward-monotonicity is only local: if we look at the hues that lie between the ginger hues and, say, the green hues, the HAIR typicality function is *upward*-monotone, since ginger hues are more typical as hair colors than green hues. Therefore, the ginger hues give a *local maximum* of the typicality for the concept HAIR among the hues that may reasonably be categorized as *red*.[13]

Above I have described the distinction between *red car* and *red hair* in terms of the behavior of the typicality functions for the concepts CAR and HAIR under changes of hue. Let us look at more concrete, hypothetical typicality functions that show this behavior, based on Lee's example. For purposes of illustration, we may suppose that green hues have typicality 0 for the concept RED, ginger hues (around RGB value 201,113,13, as in Lee's experiments) have typicality 0.6 for RED, and focal red has typicality 1 (Figure 7.3 on top). In all our analyses, we concentrate on typicality for the *head concept* (HC): the concept that is being modified by the graded concept in the construction, e.g. the concepts HAIR and CAR in the constructions *red hair* and *red car*. For these two expressions, we look at situations that vary with respect to their typicality for the gradable concept RED. We encode situations according to their typicality for the concept RED. Specifically, we let the situations that we consider be described by numbers between 0 and 1, according to their typicality for RED. For instance, for RED HAIR we are interested in various situations where hair only differs in hue, and the hues range between green and focal red. Now, Figure 7.3a approximares the typicality for the head concept HAIR as a function of the hue typicality for the gradable concept RED.[14] As illustrated in this figure, ginger attains local maximal typicality for the concept HAIR. We look at this local maximum as an optimal "compromise" in typicality between the concepts RED and HAIR, referring to it as a *critical typicality* (CT) point for the complex concept RED HAIR.

XXX Figure 7.3 approximately here XXX

Using the same kind of analysis, let us now consider the concept CAR. We assumed that typicality for CAR is constant in the hue. Thus, all hues obtain local maxima for CAR, as illustrated in Figure 7.3b. Among these hues, we define the CT point for RED CAR as the *reddest* one, i.e. focal red. Thus, the CT point for RED CAR is much redder than ginger,

---

[13]The choice of green hues as the point where typicality starts to increase towards ginger is for presentation sake. Black, brown and blond may be globally more typical than ginger hues. Therefore, other hues besides ginger attain local maxima for HAIR typicality. However, between the black, brown and blond hues that are most typical for HAIR, we assume that typicality for HAIR must decrease somewhere before it raises again towards the ginger hues. Hence, for the sake of illustration, it is safe to assume that ginger attains a local maximum for hair typicality also in other dimensions of color other than the one illustrated here. This simplified presentation would not work for multi-dimensional concepts like PET or FISH, which introduce further complications when dealing with constructions like *pet fish* (Hampton 2007, p.374). I believe that the current proposal can be subsumed by current models of distributional semantics (Baroni & Zamparelli 2010, Mitchell & Lapata 2010 , McNally and Boleda, this volume). However, since *pet fish* examples concern problems with noun modification that are not immediately relevant for plurals, this task is left for further work.

[14]The graphs in Figure 7.3, as well as in Figures 7.4 and 7.5, are only for illustrative purposes. Using exact typicality values here is impossible, since the experiments by Lee, Poortman et al. and Poortman did not measure typicality directly, but only compared various situations as for their typicality. With respect to acceptability (as opposed to typicality), Figures 7.7, 7.8 and 7.9 that follow later better approximate the results in those experiments.

which is the CT point for RED HAIR.[15]   Note that the difference between Figures 7.3a and 7.3b is predicted from the difference between the typicality functions for the concepts HAIR and CAR, and the way they behave under changes in typicality for RED.[16] Thus, our ability to see differences between expressions like *red hair* and *red car* is based on the information we have on their parts, in agreement with compositionality. However, unlike standard conjunctive rules, the information we use is not just the extension of these parts, i.e. it is not the sets of items categorized by them. Rather, as the "meaning" of the parts we use the typicality functions derived from the lexical concepts.

With different CT points for the expressions *red hair* and *red car*, we are half way in the analysis of Lee's results. As will be described more explicitly below, instances at the CT points are expected to be frequently judged as instances of the complex concept. Thus, since focal-red cars are at the CT point, they are likeliest to be classified as a *red car*. However, ginger cars are substantially below the CT point for CAR, hence they are unlikely to be classified as *red car*. By contrast, ginger hair instances are at the CT point for HAIR, hence are likely to be classified as *red hair*. A separate question, which will also be discussed below, is what happens with focal-red instances of HAIR, which are substantially *above* the CT points for HAIR. As we will see, different concepts behave differently at such points.

Before moving on to plural concepts, let us explicitly summarize our analysis of the *red hair* example.

(6)   Let LM-HAIR be the set of local typicality maxima for the concept HAIR. In formula:

$$\text{LM-HAIR} = \arg\max_x \text{TYP}_{\text{HAIR}}(x).$$

The *critical typicality* point(s) for RED HAIR is defined by:

$$\arg\max_{x \in \text{LM-HAIR}} \text{TYP}_{\text{RED}}(x).$$

In words: the critical typicality points for RED HAIR are those points that attain maximal typicality for RED among the points that attain maximal typicality for HAIR. Suppose that the local maxima in LM-HAIR are four points that represent the focal hues for brown, black, blond and ginger. Of these focal hues, ginger attains maximal typicality for RED, hence it is classified as the CT point for RED HAIR. A similar procedure yields a focal red as the critical point for RED CAR: supposing, as we did, that all hues are equally typical for CAR, we get the set of local maxima for CAR contain the whole spectrum. Of these hues, focal red is most typical for RED, hence classified as the CT point for RED CAR.

In more general terms, let us define the critical typicality point(s) for *gradable concepts* (e.g. RED) when they operate with *head concepts* like CAR or HAIR.

---

[15]This idea is inspired by the following intuition in Krifka (1999): *Context-sensitive interpretation of* RED *can be given as: "When combined with a noun meaning N, it singles out those objects in N that appear closest to the color of blood for the human eye."*

[16]The process can be extended for complex phrases like *giant midget/midget giant* (Kamp & Partee 1995), where both concepts are gradable (and unidimensional). However, the focus here is on those cases where one of the concepts is gradable and the other one is not, since in those cases the generalization to plural concepts is most direct.

(7) **Critical Typicality** (CT): *Let GC be a gradable concept and HC be a head concept. Let LM-HC be the set of local typicality maxima for HC, defined by:*

$$\textit{LM-HC} = \arg\max_x \mathrm{TYP}_{\mathrm{HC}}(x).$$

The set $\mathrm{CT}_{\mathrm{GC\text{-}HC}}$ *of* **critical typicality** *(CT) points for the complex concept GC-HC is defined by:*

$$\arg\max_{x \in \text{LM-HC}} \mathrm{TYP}_{\mathrm{GC}}(x).$$

In words: the CT points for the complex concept GC-HC are those points that attain maximal typicality for the gradable concept GC among the points that attain maximal typicality for the head concept HC. This is precisely the same definition as in (6), but using more general notation.[17] This definition will now help us analyze various guppy effects where the gradable concept is a *plurality operator*: distributivity or reciprocity.

6   Guppy effects with plurals

While guppy effects in concept composition have been widely studied in the literature on vagueness and adjectives, they have less often been addressed in relation to the interpretation of plurals. The main proposal of this paper is that the contextual mechanisms affecting meaning composition with vague adjectives are also operational with vague plural quantifiers. To see the parallelism, consider the following two sentences (Philip 2000, Winter 2001).

(8)   a.   Mary, Lucy and Candy *are pinching each other*.
      b.   Mary, Lucy and Candy *know each other*.

As we may naturally expect, the concept PINCH shows preference for situations where an agent only pinches one patient at a time (this was experimentally verified by Kerem et al. and Poortman et al.). However, the reciprocity concept for *each other* prefers as many relations as possible (Dalrymple et al. 1998, Sabato & Winter 2012). In sentence (8a) this leads to a conflict between the preferences of the two concepts. By contrast, in sentence (8b) there is no conflict in preferences between the verb and the reciprocal expression: a person may know many people, without clear typicality preferences between different numbers of acquaintances. Intuitively, this contrast points to a possible guppy effect in sentence (8a) when compared to (8b). Indeed, Kerem et al. and Poortman et al. experimentally showed that there is a substantial difference in the interpretations of sentences like (8a) and (8b). Poortman et al. tested truth-value judgements on Dutch versions of these sentences (and similar ones) for I3 situations as in Figure 7.1 above: three agents, each of them acting

---

[17] James Hampton (p.c.) points out that definition (7) is not fully general. For instance, for GREEN HAIR, we would like the CT point to be around focal green, and not, say, ginger or focal black. However, since ginger and focal black are local maxima for HAIR typicality, definition (7) might classify them as CT points, despite their zero typicality for GREEN. To address this problem, we would need definition (7) to only consider local maxima of HC-typicality which are close enough to the maximal typicality points for GC. E.g., for GREEN HAIR, we would only consider hair instances at a certain distance d from focal green. Reasonably, all these points would be equally (a)typical for HAIR, hence classified as local maxima. Accordingly, focal green would be classified as the CT point for GREEN HAIR. By contrast, if we look at hair instances at distance d from focal *red*, we expect to see some increase in HAIR typicality at different points closer to the browns, blonds or gingers. This would only allow such points be the local maxima, in agreement with the analysis above. For sake of presentation, we avoid this complication of (7), which is unnecessary for the analysis of plural concepts.

on another agent. While 88% of the participants accepted (8a) as true in this I3 situation, only 36% accepted (8b) in the same situation. Kerem et al. and Poortman et al. explain such differences as a guppy effect, using a principle that they call the *Maximal Typicality Hypothesis* (MTH). As an instance of the more general principle in (7), the MTH is analyzed as responsible for the guppy effect. For example, with the concept PINCH in (8a), the I3 situation is the most typical situation for the reciprocity concept that is consistent with the preference that one person does not pinch more than two people at the same time. If we try to add more relations to I3 in order to satisfy better the preferences of the reciprocal concept, we get atypical situations for the PINCH concept, As a result, I3 is the critical typicality point for the verb phrase *pinch each other* in sentence (8a). By contrast, in sentence (8b), with the concept KNOW, there is no substantial typicality difference between I3 and configurations containing more relations. Consequently, the preferences of the reciprocal concept are free to take over, and I3 is not a CT point for the expression *know each other* in (8b). Rather, in this case the CT point is the I6 situation, where every one of the three people knows every other person. Figures 7.4a-b summarize the typicality considerations for sentences (8a) and (8b) with I3 and I6. For contrast, these figures also include the I2 situation, with only two pinching relations between the three people. In Kerem et al.'s and Poortman et al.'s experiments, I2 situations consistently showed very low acceptability for reciprocal sentences with all tested verbs.

XXX Figure 7.4 approximately here XXX

In (9) below we apply principle (7) to the case of the complex concept for *pinch each other*, where PINCH is the head concept HC and RECIP is the gradable concept GC.

(9) Let LM-P be the set of local typicality maxima for the concept PINCH. In formula:

$$\text{LM-P} = \arg\max_x \text{TYP}_{\text{PINCH}}(x).$$

The CT point(s) for *pinch each other* is defined by:

$$\arg\max_{x \in \text{LM-P}} \text{TYP}_{\text{RECIP}}(x).$$

In words: the critical typicality points for *pinch each other* are the situations that attain maximal typicality for the gradable concept RECIP among the situations that attain maximal typicality for PINCH. As we saw, this entails that I3 is a CT point for *pinch each other*. This analysis is analogous to the analysis of the RED HAIR example. A similar procedure gives us I6 as the CT point for *know each other*, in analogy to the RED CAR example.

Poortman (this volume) studies another guppy-like effect with plurals. Consider the following sentences.

(10) a. Dan, Bill, John and George *are walking and writing*.
b. Dan, Bill, John and George *are walking and singing*.

In these constructions, we view 'bi-concepts' like WALK&WRITE and WALK&SING as the head concepts. The gradable concept of "distributivity" DIST is not overtly present in the sentence, but is standardly introduced in the analysis, as discussed above. We analyze the distributivity concept DIST as a gradable vague concept similar to *many* or *enough*. The complex concept for the plural expression *are walking and writing* is denoted DIST WALK&WRITE (as opposed to the conjunctive concept WALK&WRITE for the uninflected

conjunction). Similarly, the complex concept for the expression *are walking and singing* is denoted DIST WALK&SING.

Now let us consider three types of situations for sentences (10a-b), illustrated in Figure 7.5:

- A *joint* situation, where each of the four people is doing both activities, i.e. walking and writing/singing.

- A *full-split* situation, where two people are performing one activity, and the other two are performing the other activity.

- A *partial-split* situation, where one of the persons is performing one activity, and another person is performing the other activity. The other two people are not engaged in any of the activities.

XXX Figure 7.5 approximately here XXX

Partial-split situations are quite unacceptable for both sentences (10a) and (10b). Thus, when Dan walks, George writes and Bill and John are doing neither activity, sentence (10a) is quite odd, and similarly for (10b). By contrast, in full-split situations, Poortman's experiments show a difference between Dutch sentences similar (10a) and (10b). While 81% of the participants accepted sentences like (10a) in a full-split situation, only 24% accepted (10b) (in Poortman's experiment the subjects were definite descriptions, e.g. *the boys*). Once more, when accounting for this effect we rely on critical typicality points. In (11) below we apply principle (7) to the case of the expression *are walking and writing*. In this example, we denote WALK&WRITE for the head concept HC, where DIST, the distributivity concept, is the gradable concept GC.

(11)   Let LM-WW be the set of local typicality maxima for the concept WALK&WRITE. In formula:

LM-WW $= \arg\max_x \mathrm{TYP}_{\text{WALK\&WRITE}}(x).$

The CT point(s) for the complex concept DIST WALK&WRITE is defined by:

$\arg\max_{x \in \text{LM-WW}} \mathrm{TYP}_{\text{DIST}}(x).$

In words: the CT points for the expression *are walking and writing* are the situations that attain maximal typicality for the gradable concept DIST among the situations that attain maximal typicality for WALK&WRITE. Among the three situations we consider, the full-split and partial-split situations are more typical for WALK&WRITE than the joint situation. Between these two situations, the full-split situation is substantially more typical for the concept DIST, as it contains more people who are engaged in the relevant activities. This means that for sentence (10a), full-split is the CT point among the three situations. By contrast, for sentence (10a), the three situations are equally typical for the head concept WALK&SING. This is because the concepts WALK and SING, unlike WALK and WRITE, are not in any conflict (this kind of contrast is shown in another experiment by Poortman). In this case, for sentence (10b), the critical typicality point is the "joint situation", which is the most typical for the distributivity concept alone. This analysis is summarized in Figure 7.6.

XXX Figure 7.6 approximately here XXX

7   The effect of critical typicality on acceptability

We have seen that for three different cases, typicality considerations allow us to specify points of *critical typicality*. Such CT points hit the optimal equilibrium between the (possibly conflicting) typicality preferences of concepts appearing together in a complex construction. We would like to use CT points in order to explain facts like the following:

(12)   a.   Ginger-hued hair is more acceptable for *red hair* than car instances of the same hue are for *red car* (0.92 vs. 0.17 acceptability rates in Lee's experiments).

   b.   I3 situations are more acceptable for *pinch each other* than for *know each other* (0.88 vs. 0.36 acceptability rates in Poortman et al.'s experiments).

   c.   Full-split situations are more acceptable for *walking and writing* than for *walking and singing* (0.81 vs. 0.24 acceptability rates in Poortman's experiments).

The three experiments compared acceptability proportions when participants were asked to categorize instances of two complex expressions. From the perspective developed here, which follows Black (1937), a central goal of semantic theory is to account for such results on acceptability. It is important to note that *a priori*, it is not self-evident that acceptability proportions reflect typicality preferences for the complex concept. Typicality effects are only studied here insofar as they help to predict acceptability, as measure in truth-value judgement tasks. The typicality data that were tested in Lee's, Poortman et al.'s and Poortman's experiments concern the head concept, and not any complex concepts. As Fodor (1981) claimed, it becomes increasingly harder to study typicality information when it comes to complex expressions.[18] An example for the difficulty was illustrated in one of Kerem et al.'s (2009) experiments. In a forced choice experiment, only 26% of the participants preferred I3 to I6 as a better illustration for sentences of the form *A, B and C are pinching each other*. Thus, if we want to explain why I3 situations are more acceptable for *pinch each other* than for *know each other*, it is not clear how typicality effects might help us: when participants were asked about their preferences (as opposed to their acceptability judgements), many of them preferred I6 to I3 situations for *pinch each other*, and ignored the atypical status of such scenarios for the concept PINCH. This effect nicely demonstrates Fodor's point: as expressions get longer and more complicated, typicality judgements do not seem very informative.

Because of this point, we concentrate on acceptability judgements for complex expressions, rather than on typicality judgements. Typicality judgements are only collected for lexical expressions, which are then used for predicting the CT point(s). These CT points, in turn, are proposed to be an important factor that boosts acceptability judgements for complex expressions. The facts in (12) are all explained using the following generalization.

(13)   Let GC-HC$_1$ and GC-HC$_2$ be two complex concepts, with the same graded concept GC, but with different head concepts HC$_1$ and HC$_2$. Let two situations, S_1 and S_2 be exemplars for HC$_1$ and HC$_2$, respectively, where S_1 and S_2 are of equal typicality for the graded concept GC. Suppose that S_1 is a CT point for the complex concept GC-HC$_1$, but *below* the CT point for the complex concept GC-HC$_2$. Then

---

[18]Fodor used the complex phrase "American cities situated on the East Coast just a little south of Tennessee" to argue against the plausibility of prototype theory. According to Fodor, it is unlikely that such expressions have any prototype.

the acceptability of S_1 for GC-HC$_1$ is substantially higher than the acceptability of S_2 for GC-HC$_2$.

For instance, since ginger hues are CT points for the concept HAIR, but not for the concept CAR, their acceptability for *red hair* is substantially higher than for *red car*. Similarly, since I3 situations are CT points for the concept PINCH, but not for the concept KNOW, their acceptability for *pinch each other* is substantially higher than for *know each other*. Further, since full-split situations are CT points for the concept WALK&WRITE, but not for WALK&SING, their acceptability for *are walking and writing* is substantially higher than for *are walking and singing*.

Note that the CT point is proposed to be a factor that boosts acceptability, but does not necessarily maximize it. Points above the CT point may have acceptability lower than or equal to the acceptability at the CT point, but they may also have higher acceptability. Here are two examples for this variability:

- In Lee's experiments, focal red hair was slightly less acceptable for *red hair* than ginger hair (0.87 vs. 0.92). By contrast, a bike made completely of wood was more acceptable for *wooden bike* than a more typical bike with some wood parts (0.96 vs. 0.71).

- In Poortman et al.'s experiments, the I6 situation was as acceptable for *pinch each other* as the I3 situation (0.88 in both cases). By contrast, the I6 situation was substantially less acceptable for *bite each other* than I3 (0.44 vs. 0.84).

Thus, above the CT point (I3, ginger hues, partially wooden bike), acceptability may increase, decrease, or remain the same. I propose that this happens because acceptability for a complex concept above the CT point is compositionally affected by the acceptability values for the parts. For instance, participants may fail to accept I6 situations as *biting* situations, simply because they reject the possibility that one person can be biting two people simultaneously. By contrast, a bike made completely of wood is still a bike, hence acceptability for *wooden bike* may continue to increase beyond the CT point.

Based on this account, we use CT points for specifying acceptability functions for complex concepts on the basis of the acceptability functions of their parts. Consider again our running example, the concept RED HAIR. Let us assume that x_0 is the least red hue that is still categorized as RED by some speakers.[19] For instance, x_0 may be a slightly redder hue than focal green (where the latter has zero acceptability for RED). Based on the minimal acceptability point x_0 and the critical typicality point CT, we can describe the acceptability function for RED HAIR as consisting of three parts:

(i) For every hue x between focal green and x_0, the acceptability $\text{ACC}_{\text{RED HAIR}}(x)$ is constantly zero (as x is definitely not red).

(ii) For every hue x between the CT point (ginger) and focal red: $\text{ACC}_{\text{RED HAIR}}(x)$ is compositionally determined by $\text{ACC}_{\text{RED}}(x)$ and $\text{ACC}_{\text{HAIR}}(x)$, e.g. by multiplication.[20] In-

---

[19]It is not guaranteed that such a point exists: acceptability for RED may asymptotically reach zero. The assumption of x_0 is only a convenience here.

[20]The exact way to compose acceptability functions must be studied empirically, and is likely to be quite different from multiplication, which is only assumed here for illustration. The important assumption is that the acceptability of for the complex concepts grows monotonically in the acceptability of its parts.

tuitively, hair hues that are redder than the CT point(s) inherit their acceptability compositionally, from the acceptability for RED and for HAIR independently.

(iii) For every hue x between x_0 and the CT point (ginger): $\text{ACC}_{\text{RED HAIR}}(x)$ grows monotonically – e.g. linearly.[21] This penalizes acceptability of hues below the CT point, while allowing them to be non-zero.

This captures three facts about RED HAIR. Point (i) explains why instances of HAIR that are close enough to focal green, i.e. less red than x_0, are fully unacceptable for RED HAIR, as they are for RED. Point (ii) explains how hair instances that are redder than ginger (the CT point) may have lower, equal, or even greater acceptability: the only restriction that this assumption puts on the acceptability function for the complex expression is that it is determined compositionally from the acceptability of its parts. Part (iii) of the acceptability function describes how points between x_0 and ginger are only mildly acceptable for RED HAIR, while letting acceptability grow monotonically as hues move towards ginger.

In more general terms, we adopt the following rule, which defines how CT points interact with acceptability for complex concepts.

(14) **CT-induced Acceptability**: Let GC be a gradable concept and let HC be a head concept. Suppose that in the interval $[0, 1]$, the CT point(s) for the concept GC-HC are precisely the segment/point $[x_1, x_2]$ where $x_1 \leq x_2$. Suppose further that x_0 is a maximal point s.t. $\text{ACC}_{\text{GC}}(x_0)$ attains a ceratin minimum, denoted *min*. Thus, for every $x \leq x_0$, $\text{ACC}_{\text{GC}}(x) = min$, and for every $x > x_0$: $\text{ACC}_{\text{GC}}(x) > min$. We define the acceptability function for the concept GC-HC as follows:

$$\text{ACC}_{\text{GC-HC}}(x) = \begin{cases} min & x \leq x_0 \\ \textbf{linear} & x_0 \leq x \leq x_1 \\ \textbf{constant} & x_1 \leq x \leq x_2 \\ x \cdot \text{ACC}_{\text{HC}}(x) & x \geq x_2 \end{cases}$$

The linear part is only assumed for concreteness, and may be replaced by any other function that grows monotonically in the acceptability for GC. The same remark holds for the last part, where multiplication may be replaced by any other two-place function that is upward-monotonic on both its arguments. The acceptability function $\text{ACC}_{\text{RED HAIR}}(x)$ of the concept RED HAIR is plotted in Figure 7.7a, under the linearity assumption between x_0 and the CT point, and a slight decline in acceptability (in Lee's experiment – from 92% to 87%) in the points higher than the CT point. In Figure 7.7b, a similar graph describes the acceptability function $\text{ACC}_{\text{RED CAR}}(x)$ of the concept RED CAR. Here the CT point is the point for focal red, hence the acceptability function raises linearly from the point x_0 of minimal acceptability to that CT point of maximal acceptability. The difference between the graphs in Figures 7.7a-b illustrates the effect of the CT point on the acceptability for complex concepts.[22]

XXX Figure 7.7 approximately here XXX

Let us consider another example for the application of CT-induced acceptability: the acceptability function for *pinch each other* in sentence (8a). We assume that x_0=I2 is a

---

[21]As with the case of multiplication above, linearity is only assumed here for illustration. The important assumption is that the function grows monotonically.

[22]The linearity approximation in Figure 7.7b grossly underplays the effect of the CT point in Lee's experiment: in fact, only 17% of the participants accepted ginger hues for RED CAR.

minimum point with zero acceptability for *each other*, and that the critical typicality point CT is I3.[23] As a result, we get the following acceptability function, with two parts in its definition (since I2 and I3 are discrete points).

$$\text{ACC}_{\text{PINCH EACH OTHER}}(x) = \begin{cases} 0 & x \text{ is a situation with 2 or less } \text{PINCH relations} \\ x \cdot \text{ACC}_{\text{PINCH}}(x) & x \text{ is a situation with between 3 and 6 } \text{PINCH relations} \end{cases}$$

Here the acceptability of any lower point than I3 (=CT point), with 2 or less relations, is zero (or otherwise very low, to the extent reciprocity tolerates such situations, see Beck & von Stechow 2007, Mari 2014). Higher points than the CT, between I3 and I6, inherit their acceptability from the acceptability of reciprocity (x) and the acceptability for the concept PINCH (which decreases with number of relations). In Poortman et al.'s experiments this leads to equal acceptability of I3 and I6 for *pinch each other*. The acceptability function for *pinch each other* is described in Figure 7.8a.

XXX Figure 7.8 approximately here XXX

When applying principle (14) to *know each other* in sentence (8b), we again use the assumption that x_0=I2 is a minimum point. However, as we saw, the CT point for *know each other* is I6, hence the acceptability function that (14) generates is different than for *pinch each other*. This function is given below.

$$\text{ACC}_{\text{KNOW EACH OTHER}}(x) = \begin{cases} 0 & x \text{ is a situation with 2 or less } \text{KNOW relations} \\ \textbf{linear} & x \text{ is a situation with between 3 and 6 } \text{KNOW relations} \\ 1 & x = \text{I6 (i.e. 6 } \text{KNOW relations)} \end{cases}$$

Here I2 and "smaller" situations are of low acceptability as with *pinch each other*. However, the CT point for *know each other* is I6. As a result, I6 is the only point where sentence (8b) attains maximal acceptability. In the points between I2 and I6, the acceptability function increases monotonically. This acceptability function for *know each other* is described in Figure 7.8b.

The analysis of the acceptability functions for plural predicates like *are walking and writing/singing* in (10a-b) is similar. For sentence (10a), we get the following acceptability function.

$$\text{ACC}_{\text{DIST WALK\&WRITE}}(x) = \begin{cases} 0 & x \text{ is any situation between partial to full-split} \\ & \text{(=CT point)} \\ x \cdot \text{ACC}_{\text{WALK\&WRITE}}(x) & x \text{ is any situation between full-split and joint} \end{cases}$$

Here the CT point is the full-split situation, where two people are walking and two are writing. The acceptability of any lower point, with three or less active agents, is zero or very low (to the extent that distributivity tolerates such non-maximal situations). Higher points than the CT, between full-split and joint, inherit their acceptability from the acceptability of distributivity (x) and the acceptability for the head concept WALK&WRITE, which may stay constant or decrease a little with the size of the intersection, but (probably) not increase.

---

[23]The first assumption is only for the sake of illustration: actually 16% of the participants in Poortman et al.'s experiments accepted I2 for sentences like (8a). It is likely that situations like I1 and I0 (one and zero relations) would get even lower acceptability, hence the actual x_0 point is probably I1 or I0. The second assumption on typicality is based on the clear preference in Poortman et al.'s experiments for situations in which each agent only pinches one patient, rather than more than one.

The acceptability function for the complex concept DIST WALK&WRITE (for *are walking and writing*) is described in Figure 7.9a.

XXX Figure 7.9 approximately here XXX

For sentence (10a) the CT point is the joint situation, hence we get the following acceptability function.

$$
\text{ACC}_{\text{DIST WALK\&SING}}(x) = \begin{cases} 0 & x \text{ is any situation between partial to full-split} \\ & (=\text{CT point}) \\ \textbf{linear} & x \text{ is any situation between full-split and joint} \\ x \cdot \text{ACC}_{\text{WALK\&SING}}(x) & x \text{ is the joint situation} \end{cases}
$$

Here the partial-split situation and "smaller" situations are of low acceptability as with *walk and write*. However, the CT point is the joint situation, where plausibly, both distributivity and the WALK&SING concept attain maximal acceptability. In between these two points, the function increases monotonically. This function for the complex concept DIST WALK&SING is described in Figure 7.9b.

8   Conclusion

This paper aimed to point out some general principles, which, if correct, should have desired implications for theories of concept composition in linguistics and psychology. It is well-accepted that truth-value judgements are systematically related to typicality, and that this relation is important for the compositional analysis of vagueness with nominal expressions. However, the general implications of this point for the analysis of compositionality are still understudied. Recent experimental work has shown that typicality effects influence the interpretation of plurals in a rather similar way to their relations with other vague concepts. Hopefully, the explorative work in this paper may help in improving existing theories of context and concepts, and the way they interact in the construction of meaning for complex expressions.

References

Alxatib, S. & Pelletier, F. J. (2011), 'The psychology of vagueness: Borderline cases and contradictions', *Mind & Language* **26**(3), 287–326.

Barker, C. & Jacobson, P. (2007), Introduction: Direct compositionality, *in* C. Barker & P. Jacobson, eds, 'Direct Compositionality', Oxford University Press, Oxford, pp. 1–22.

Baroni, M. & Zamparelli, R. (2010), Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space, *in* 'Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing', Association for Computational Linguistics, pp. 1183–1193.

Barsalou, L. W. (2017), Issues for psychologically plausible theories of conceptual combination. this volume.

Beck, S. & von Stechow, A. (2007), 'Pluractional adverbials', *Journal of Semantics* **24**(3), 215 –254.

Black, M. (1937), 'Vagueness. an exercise in logical analysis', *Philosophy of science* **4**(4), 427–455. Reprinted in Keefe & Smith (1996).

Bonini, N., Osherson, D. N., Viale, R. & Williamson, T. (1999), 'On the psychology of vague predicates', *Mind & Language* **14**(4), 377–393.

Brisson, C. (1998), Distributivity, Maximality and Floating Quantifiers, PhD thesis, Rutgers University.

Burnett, H. (2012), Vague determiner phrases and distributive predication, *in* M. Slavkovik & D. Lassiter, eds, 'New Directions in Logic, Language and Computation', Springer, Dordrecht, pp. 175–194.

Chater, N., Lyon, K. & Myers, T. (1990), 'Why are conjunctive categories overextended?', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **16**(3), 497–508.

Cresswell, M. J. (1976), The semantics of degree, *in* B. Partee, ed., 'Montague Grammar', Academic Press, New York.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S. & Peters, S. (1998), 'Reciprocal expressions and the concept of reciprocity', *Linguistics and Philosophy* **21**, 159–210.

Dowty, D. (1987), Collective predicates, distributive predicates and *all*, *in* 'Proceedings of the Eastern States Conference on Linguistics, ESCOL3 (1986)', Cascadilla Press, pp. 97–115.

Egré, P., de Gardelle, V. & Ripley, D. (2013), 'Vagueness and order effects in color categorization', *Journal of Logic, Language and Information* **22**(4), 391–420.

Fodor, J. A. (1981), The present status of the innateness controversy, *in* J. A. Fodor, ed., 'Representations: Philosophical Essays on the Foundations of Cognitive Science', MIT Press, pp. 257–316.

Hampton, J. A. (1988*a*), 'Disjunction of natural concepts', *Memory & Cognition* **16**(6), 579–591.

Hampton, J. A. (1988*b*), 'Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**(1), 12–32.

Hampton, J. A. (1997), 'Conceptual combination: Conjunction and negation of natural concepts', *Memory & Cognition* **25**(6), 888–909.

Hampton, J. A. (1998), 'Similarity-based categorization and fuzziness of natural categories', *Cognition* **65**(2), 137–165.

Hampton, J. A. (2007), 'Typicality, graded membership, and vagueness', *Cognitive Science* **31**(3), 355–384.

Hansen, T., Olkkonen, M., Walter, S. & Gegenfurtner, K. R. (2006), 'Memory modulates color appearance', *Nature neuroscience* **9**(11), 1367–1368.

Janssen, T. M. V. (1997), Compositionality, *in* J. Van Benthem & A. ter Meulen, eds, 'Handbook of Logic and Language', Elsevier, Amsterdam.

Kamp, H. & Partee, B. (1995), 'Prototype theory and compositionality', *Cognition* **57**, 129–191.

Kamp, J. A. W. (1975), Two theories about adjectives, *in* E. L. Keenan, ed., 'Formal Semantics of Natural Language', Cambridge University Press, Cambridge.

Keefe, R. & Smith, P., eds (1996), *Vagueness: A reader*, MIT Press, Cambridge, Massachusetts.

Kennedy, C. (2007), 'Vagueness and grammar: the semantics of relative and absolute gradable adjectives', *Linguistics and Philosophy* **30**, 1–45.

Kennedy, C. (2011), Ambiguity and vagueness: An overview, *in* C. Maienborn, K. von Heusinger & P. Portner, eds, 'Semantics: An International Handbook of Natural Language Meaning', Vol. 1, De Gruyter, Berlin, pp. 507–535.

Kerem, N., Friedmann, N. & Winter, Y. (2009), Typicality effects and the logic of reciprocity, *in* E. Cormany, S. Ito & D. Lutz, eds, 'Proceedings of Semantics and Linguistic Theory, SALT19', *eLanguage*, pp. 257–274.

Klein, E. (1980), 'A semantics for positive and comparative adjectives', *Linguistics and Philosophy* **4**, 1–45.

Krifka, M. (1999), Compositionality, *in* R. A. Wilson, F. C. Keil & A. Pierce, eds, 'The MIT Encyclopaedia of the Cognitive Sciences', MIT Press, Cambridge, MA, pp. 152–153.

Križ, M. & Chemla, E. (2015), 'Two methods to find truth-value gaps and their application to the projection problem of homogeneity', *Natural Language Semantics* **23**(3), 205–248.

Kubat, R., Mirman, D. & Roy, D. (2009), Semantic context effects on color categorization, *in* N. Taatgen & H. van Rijn, eds, 'Proceedings of the 31st Annual Meeting of the Cognitive Science Society', Cognitive Science Society, Inc., pp. 491–495.

Langendoen, D. T. (1978), 'The logic of reciprocity', *Linguistic Inquiry* **9**, 177–197.

Lasersohn, P. (1999), 'Pragmatic halos', *Language* **75**, 522–551.

Lee, C. (2017), Color typicality knowledge and the interpretation of color adjectives. this volume.

Mari, A. (2014), '*Each Other*, asymmetry and reasonable futures', *Journal of Semantics* **31**, 209–261.

McCloskey, M. E. & Glucksberg, S. (1978), 'Natural categories: Well defined or fuzzy sets?', *Memory & Cognition* **6**(4), 462–472.

McNally, L. & Boleda, G. (2017), Conceptual vs. referential affordance in concept composition. this volume.

Mitchell, J. & Lapata, M. (2010), 'Composition in distributional models of semantics', *Cognitive Science* **34**(8), 1388–1429.

Nouwen, R., van Rooij, R., Sauerland, U. & Schmitz, H.-C., eds (2011), *Vagueness in Communication*,

Springer, Dordrecht.

Osherson, D. N. & Smith, E. E. (1981), 'On the adequacy of prototype theory as a theory of concepts', *Cognition* **9**, 35–58.

Osherson, D. N. & Smith, E. E. (1997), 'On typicality and vagueness', *Cognition* **64**(2), 189–206.

Pelletier, F. J. (2017), Compositionality and concepts: From a linguistics and philosophy of language perspective. this volume.

Philip, W. (2000), 'Adult and child understanding of simple reciprocal sentences', *Language* **76**, 1–27.

Poortman, E. (2017), Concept typicality and the interpretation of plural predicate conjunction. this volume.

Poortman, E., Struiksma, M., Kerem, N., Friedmann, N. & Winter, Y. (2016), When logic meets a prototype: Relational concepts and reciprocal reasoning. Unpublished ms., Utrecht University.

Rips, L. J. (1989), Similarity, typicality, and categorization, *in* S. Vosniadou & A. Ortony, eds, 'Similarity and analogical reasoning', Cambridge University Press, Cambridge, pp. 21–59.

Rosch, E. (1973), On the internal structure of perceptual and semantic categories, *in* T. Moore, ed., 'Cognitive Development and the Acquisition of Language', Academic Press, New York.

Sabato, S. & Winter, Y. (2012), 'Relational domains and the interpretation of reciprocals', *Linguistics and Philosophy* **35**, 191–241.

Sassoon, G. W. (2013), *Vagueness, gradability and typicality: the interpretation of adjectives and nouns*, Brill, Leiden.

Schwarz, F. (2013), Maximality and definite plurals – experimental evidence, *in* G. W. E. Chemla, V. Homer, ed., 'Proceedings of Sinn und Bedeutung 17', pp. 509–526.

Serchuk, P., Hargreaves, I. & Zach, R. (2011), 'Vagueness, logic and use: Four experimental studies on vagueness', *Mind & Language* **26**(5), 540–573.

Smith, E. E. & Medin, D. L. (1981), *Categories and concepts*, Harvard University Press, Cambridge, MA.

Smith, E. E. & Osherson, D. N. (1984), 'Conceptual combination with prototype concepts', *Cognitive Science* **8**(4), 337–361.

Storms, G., de Boeck, P., van Mechelen, I. & Ruts, W. (1996), 'The dominance effect in concept conjunctions: Generality and interaction aspects', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**(5), 1266–1280.

Storms, G., de Boeck, P., van Mechelen, I. & Ruts, W. (1998), 'Not guppies, nor goldfish, but tumble dryers, Noriega, Jesse Jackson, panties, car crashes, bird books, and Stevie Wonder', *Memory & cognition* **26**(1), 143–145.

Werning, M., Hinzen, W. & Machery, E. (2012), *The Oxford handbook of compositionality*, Oxford University Press, Oxford.

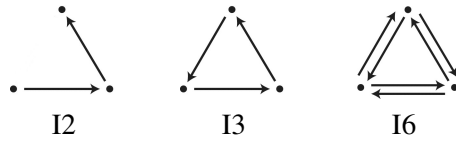Winter, Y. (2001), 'Plural predication and the Strongest Meaning Hypothesis', *Journal of Semantics* **18**, 333–365.

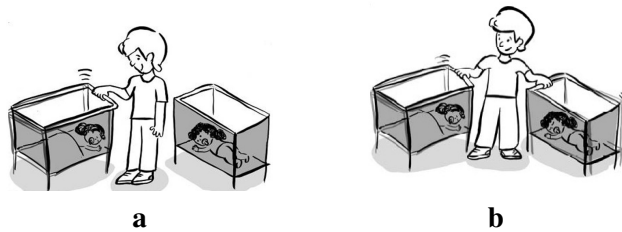**Figure 7.1**: Three situations for the interpretation of the sentence *the three girls admire each other*.

Figure 7.2: Two illustrations of the transitive meaning of the Hebrew verb for "shake"

$$\text{TYP}_{\text{RED}}(\text{green}) = 0 \qquad \text{TYP}_{\text{RED}}(\text{ginger}) = 0.6 \qquad \text{TYP}_{\text{RED}}(\text{focal red}) = 1$$
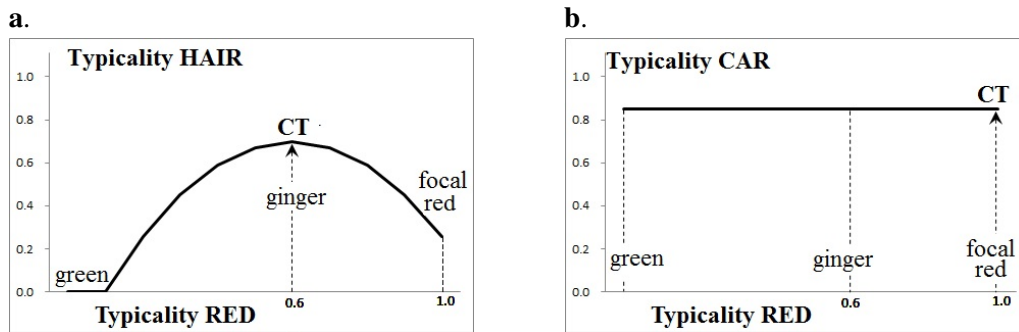
a.

Typicality HAIR

CT

ginger

focal
red

green

0.6   1.0

Typicality RED

b.

Typicality CAR

CT

green   ginger   focal
red

0.6   1.0

Typicality RED

Figure 7.3: Determining the CT point for RED HAIR **and** RED CAR – *Typicality values of different hues for* RED *(see on top) are used to describe typicality for* HAIR *and* CAR *as a function of typicality for* RED *(figures* a *and* b*). The reddest local maximum point(s) of these functions is the critical typicality (CT) point (def. (7), example (6)). The CT point affects the determination of acceptability for the complex concept (section 7).*

1
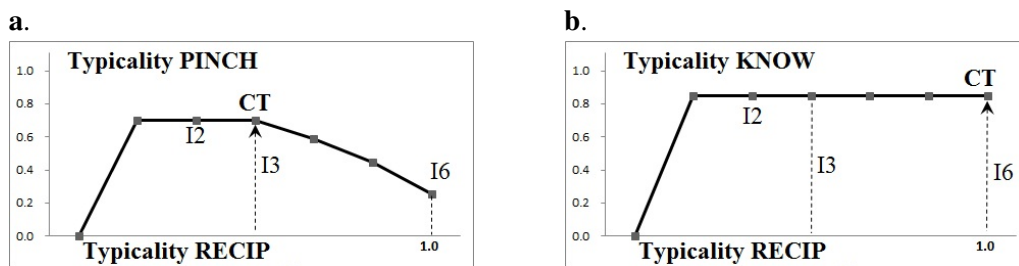
Figure 7.4: Determining the CT point for *pinch each other* and *know each other* – *Typicality values of the I2, I3, and I6 situations are used to describe typicality for* PINCH *and* KNOW *as a function of typicality for* RECIP *(figures* a *and* b*). The "most reciprocal" local maximum point(s) of these functions is the critical typicality (CT) point (def. (7), example (9)). The CT point affects the determination of acceptability for the complex concept (section 7).*

1
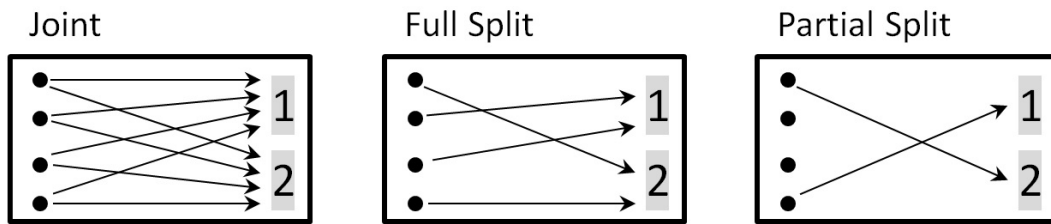
Figure 7.5: Three situations for the interpretation of sentences (10a-b) – The dots represent the agents Dan, Bill, John and George, and the numbers 1 and 2 represent the activities done by each person (*walk*, *write/sing*, or both).
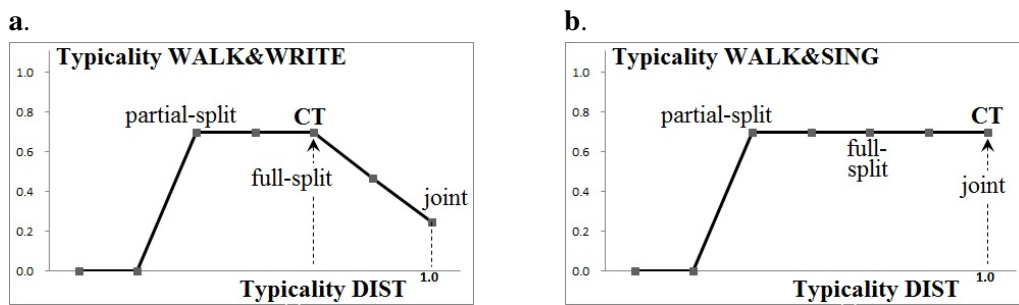
Figure 7.6: Determining the CT point for *walking and writing* **and** *walking and writing* – *Typicality values of different configurations (partial-split, full-split and joint) are used to describe typicality for* WALK&WRITE *and* WALK&SING *as a function of typicality for* DIST *(figures* a *and* b*). The local "most distributive" maximum point(s) of these functions is the critical typicality (CT) point (def. (7), example (11)). The CT point affects the determination of acceptability for the complex concept (section 7).*

1
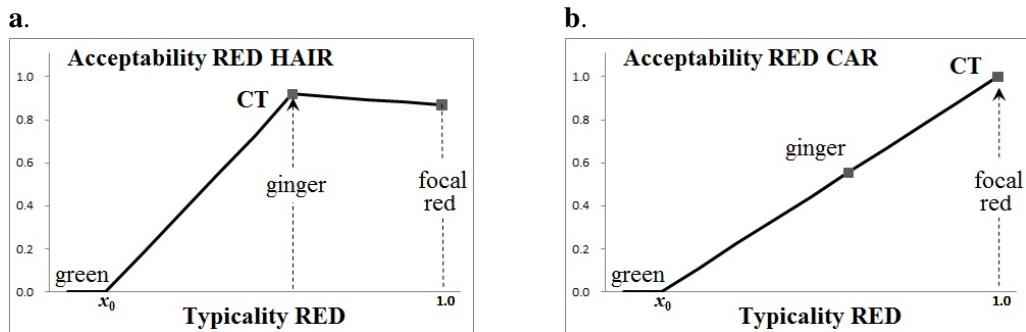
Figure 7.7: Effects of CT point on acceptability of RED HAIR **and** RED CAR – *The critical typicality (CT) point (def. (7)) is used for determining acceptability for complex concepts in two noun-adjective constructions (def. (14)).*
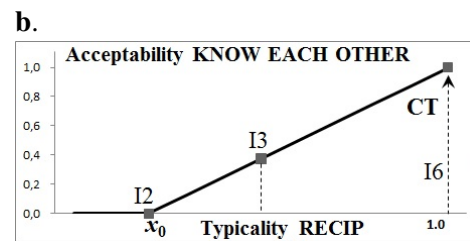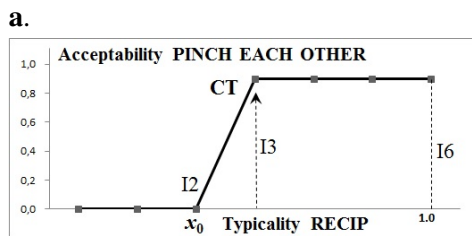
Figure 7.8: Effects of CT point on acceptability of PINCH EACH OTHER **and** KNOW EACH OTHER – *The critical typicality (CT) point (def. (7)) is used for determining acceptability for complex concepts in two reciprocal constructions (def. (14)).*
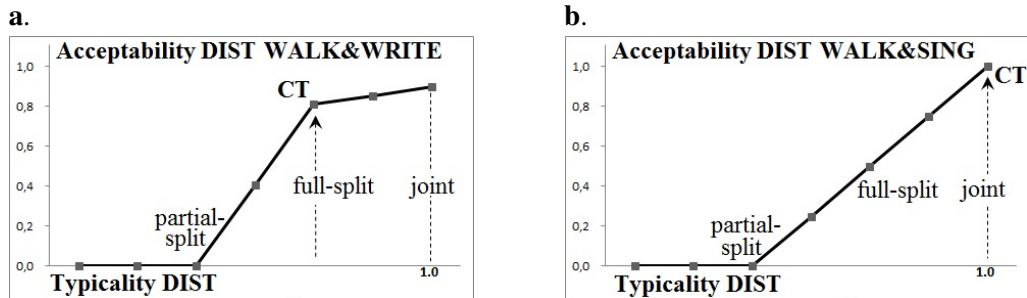
Figure 7.9: **Effects of CT point on acceptability of** *are walking and writing/singing –*
The critical typicality (CT) point (def. (7)) is used for determining acceptability for complex
concepts in two conjunctive plural constructions (def. (14)).