



# Demonstratives as bundlers of conceptual structure

**EVA WITTENBERG** 

**SHOTA MOMMA** 

**ELSI KAISER** 

*\*Author affiliations can be found in the back matter of this article*

**RESEARCH**

**]u[ubiquity press**

## Abstract

Pronoun resolution has long been central to psycholinguistics, but research has mostly focused on personal pronouns (“he”/“she”). However, much of linguistic reference is to events and objects, in English often using demonstrative pronouns, like “that”, and the non-personal pronoun “it”, respectively. Very little is known about potential form-specific preferences of non-personal and demonstrative pronouns and the cognitive mechanisms involved in reference using demonstratives. We present a novel analysis arguing that the bare demonstrative “that” serves a different function by bundling, and making linguistically accessible, complex conceptual structures, while the non-personal pronoun “it” has a form-specific preference to refer to noun phrases mentioned in the previous discourse. In two English self-paced reading studies, each replicated once with slight variations, we show that readers are reading the demonstrative slower throughout, independently of frequency or complexity of the referent, as a reflection of differences in processing demonstratives vs. pronouns. These findings contribute to two distinct but connected research areas: First, they are compatible with an emergent experimental literature showing that pronominal reference to events is preferably done with demonstratives. Second, our model of demonstratives as conceptual bundlers provides a unified framework for future research on demonstratives as operators on the interface between language and broader cognition.

**CORRESPONDING AUTHOR:**

**Eva Wittenberg**

University of California, San Diego, US

[ewittenberg@ucsd.edu](mailto:ewittenberg@ucsd.edu)

---

**KEYWORDS:**

pronoun resolution;  
demonstratives; self-paced reading; events; complexity; reference

**TO CITE THIS ARTICLE:**

Wittenberg, Eva, Shota Momma and Elsi Kaiser. 2021. Demonstratives as bundlers of conceptual structure. *Glossa: a journal of general linguistics* 6(1): 33. 1–30. DOI: <https://doi.org/10.5334/gjgl.917>

## 1. Introduction

This paper investigates how demonstrative pronouns in English are resolved in comparison to non-personal pronouns, what their respective form-specific preferences are, and which mental processes they engage in a comprehender. Pronouns as a whole are crucial discourse-structuring devices: They identify and, as we argue later, make available a mental representation created by linguistic or nonlinguistic context, such that it can be further talked about.

For instance, imagine a comprehender listening to her friends talking about their recent trip to Paris, how they visited museums and Napoleon’s tomb, how they ate macarons, and how they went to see some famous sights. While listening, the comprehender is creating mental representations, sometimes called “situation models” (Radvansky & Zacks, 2014; Zwaan, Langston & Graesser, 1995) of the traveling, walking, eating, and visiting events her friends are talking about. Then the friends end the story with one of the following sentences:

- (1) a. It’s a really beautiful city.
- b. They did not taste as sweet as the American ones.
- c. Did you know that he crowned himself?
- d. Then we walked up to Montmartre. That was a really steep hike.

Once the comprehender hears one of the sentences (1a–d), the pronouns in each sentence trigger a search back in memory to one of the concepts Paris, macarons, Napoleon, or ascending Montmartre, in order to pick the intended (or at least what the comprehender perceives as the most likely) referent, and integrate it with the predicate structure required by one of the sentences in (1).<sup>1</sup>

A prominent account of the representational structure of pronouns has been provided by Nunberg (1993). According to this account, pronouns are comprised of four components: (i) the deictic component, which is an index to a potential referent; (ii) a relational component, which specifies the relation between the index and the interpretation of the pronoun, for instance, the fact that “they” excludes the speaker; (iii) a classificatory component, which for instance specifies that “she” cannot refer to an inanimate object, or “it” cannot refer to a human; and finally, (iv) an interpretation, which one might view as the referent that the index points to.

These components all work together to enable pronoun resolution: Translated into psycholinguistic terms, one could say that the deictic component triggers a search (e.g., Ariel 2001, Gundel, Hedberg & Zacharski, 1993), the relational component and the classificatory component restrict the search space, and the interpretation is a commitment to a referent.

This process of pronoun resolution has long held a fascination for psycholinguists. Research on pronouns is an extremely productive field that has yielded valuable insights on representational properties of the referent and the pronoun alike; for instance, that people’s interpretations of personal pronouns depend on a mix of syntactic biases, semantics, discourse coherence, information structure in discourse, event structure, and non-linguistic cues (Arnold, 1998; Brown-Schmidt, 2009; Chow, Lewis & Phillips, 2014; Felser, Patterson & Cunnings, 2014; Gordon, Grosz, & Gilliom, 1993; Hartshorne & Snedeker, 2013; Kaiser, 2013; Kaiser, Runner, Sussman & Tanenhaus, 2009; Koornneef & Sanders, 2013; Järvikivi, van Gompel, & Hyönä, 2017; Kaiser & Trueswell, 2008, 2011; Kehler & Rohde, 2013; Nappa & Arnold, 2014; Rohde & Kehler, 2014; *inter alia*).

However, almost all psycholinguistic research on pronoun resolution has focused on pronouns that refer to people, although humans are just one of the many kinds of things that pronouns can point to. Consider examples (2a–c), which use the demonstrative pronouns “this” and “that”, and the simple non-personal pronoun “it”, which are the pronouns this paper is concerned with:

---

<sup>1</sup> In this paper, we use the term ‘referent’ to denote the linguistic and conceptual entity a pronoun refers to and we use the term ‘entity’ for concrete or abstract objects, excluding events, facts, situations, or propositions. We use ‘event’ to generally mean ‘things that happen over time,’ as nothing in this paper hinges on the exact definition of event (Casati & Varzi, 2008). We use the terms ‘concept’ and ‘conceptual structure’ to refer to mental representations, broadly construed as mental objects with semantic properties (Jackendoff, 2002). We do not distinguish between concepts and percepts, using the term ‘concept’ to refer to both.

- (2) a. The catacombs hold the remains of more than six million people. This is fascinating!
- b. I like Beaubourg. It always has great exhibitions.
- c. They took a walk from the Quartier Latin to Pigalle, and then had dinner. That took almost five hours.

In example (2a), the proximal demonstrative pronoun “this” makes reference to the fact described in the first sentence; in example (2b), the pronoun “it” refers to an inanimate object, namely a museum; and in example (2c), the distal demonstrative pronoun “that” can refer to the whole event of walking and dinner, or either of the subevents; the ambiguity is likely resolved by a combination of world knowledge and a recency bias (e.g., Gordon & Scearce, 1995, but see Stewart, Holler & Kidd, 2007).

Thus, to use Nunberg’s (1993) terminology, the ‘classificatory component’ of these pronouns is qualitatively different from that of personal pronouns: Whereas personal pronouns restrict the search space to the set of animate entities, non-personal and demonstrative pronouns (in English, at least) restrict the search to its complement set.<sup>2</sup> But the classificatory component is also quantitatively different: The set of non-animate concrete and abstract entities in the world is substantially larger and more heterogeneous than the set of animate entities, let alone events, facts, or situations, which non-personal and demonstrative pronouns can also refer to. We can therefore conclude that in English, neither non-personal nor demonstrative pronouns have as tight a link between their form and the *conceptual category* of their referent as personal pronouns do. Investigating this richer space of referential possibilities allows new insights into reference resolution that go beyond what one can observe with personal pronouns.

### 1.1. Previous research on non-personal pronouns and demonstratives

Understanding the different roles and representations of non-personal and demonstrative pronouns has been notoriously difficult, since their functions and uses overlap considerably. Some researchers claim that “it” and “this”/“that” “are indistinguishable with respect to the description they provide for the intended referent (an inanimate object)” (Ariel, 2001, p.29). Others claim that the interpretation of “it” and “this”/“that” depends largely on discourse status, such that non-personal pronouns are used for topics and/or salient referents, and demonstratives are used for non-topical but activated content (e.g., Gundel et al., 1993; see also Grosz, Weinstein & Joshi, 1995; Grosz & Sidner, 1986; see Bosch, Rozario & Zhao, 2003, for a similar argument for German *d*-pronouns).

Halliday (1985) proposes a system in which demonstrative pronouns establish reference to a specific token, whereas “it” specifies a non-specific token; this model was extended by Strauss (2002), who proposes a gradience in focus from “this” (high focus, important referent) and “that” (medium focus) to “it” (low focus, unimportant referent). According to these systems, the choice of a particular pronoun depends upon how much attention a speaker is asking the interlocutor to pay to the particular referent. In a similar vein, some researchers have simply claimed that bare demonstratives like “this” or “that” refer to anything that best fits all the cues which a “reasonable and attentive addressee will take the speaker to be exploiting” (Wettstein, 1984: 73; cited in Smit, 2012).

Finally, there is also a claim that “it” is sensitive to syntactic prominence, or grammatical category, while demonstrative pronouns like standalone “that” are more likely to refer to complex/composite entities (Brown-Schmidt, Byron, & Tanenhaus, 2005). In their study, Brown-Schmidt et al. (2005) found that when people were told “Move the cup onto the saucer. Now move it onto the table”, they were more likely to move only the cup; but when they were told to “move the cup onto the saucer. Now move that onto the table”, people were more likely to move the composite object cup + saucer, which has no unified linguistic antecedent.

---

<sup>2</sup> In this paper, we do not explicitly discuss discourse deixis; for discussions, see Cornish, 2008; Diessel, 1999; Eckert & Strube, 2000; Webber, 1988, Webber et al., 2003, inter alia. However, as will be evident below, our model of demonstratives is compatible with discourse deictic reference as well. We also do not consider contrastive uses of “this” and “that”, or noun phrases with demonstrative determiners (e.g. “this cat” or “this event”; Scott, 2013); see Culicover & Jackendoff (2012), Elbourne (2008), Grosz (2018), Scott (2013), and Strauss (2002) for discussion.

This finding that comprehenders tend to resolve demonstratives as referring to more complex entities than pronouns is also confirmed by Çokal, Sturt and Ferreira (2016), who found that demonstratives tend to refer to propositions, and simple pronouns to objects linguistically encoded in noun phrases (NPs). In an eye-tracking-while-reading study, they found longer reading times when the non-personal pronoun “it” referred to a proposition than when the demonstrative pronoun “this” referred to a proposition; the pattern was reversed for reference to an object NP.

Recent work employing the sentence continuation method — using both constructed texts and snippets of naturally occurring stories — found that people have a strong tendency to use “this” for event reference, and “it” for object reference, modulated both by verb class and intention to re-mention (Loáiciga, Bevacqua, Rohde & Hardmeier, 2018). The observation that reference to events tends to be accomplished with demonstrative pronouns received further support from recent corpus studies showing that almost three-quarters of demonstratives in dialogues refer to events, while only about 5% of non-personal pronouns do (Evans, 2001; Müller, 2007; Poesio, 2015).

Put together, these studies — in addition to corroborating data from other corpus analyses (e.g. Gundel et al., 1993, 2005) — seem to demonstrate two things: First, simple non-personal pronouns tend to refer to easily accessible referents encoded in prior discourse by noun phrases; and second, there is evidence for form-specific constraints (Kaiser & Trueswell, 2008): Demonstrative pronouns trigger comprehenders to construct an antecedent from previous context, one that does not necessarily have to have a simple NP antecedent. That is, the form itself (simple non-personal pronoun “it” vs. demonstrative pronouns “this” and “that”) provides valuable cues to the comprehender about what kind of referent the speaker intends, even if these cues are probabilistic.

These data fit with a theoretical proposal by Elbourne (2008), who analyzes demonstratives as denoting ‘individual concepts’ packaged as definite descriptions. Crucially, demonstrative pronouns in this model also introduce existence and uniqueness presuppositions, just as strong definite determiners do (Abbott, 2004; Strawson, 1950). That is, “I like this” presupposes that there exists something to like, and that this ‘something’ is uniquely identifiable in the context (just like “I want to eat the raspberry macaron” triggers the presupposition that a particular unique raspberry macaron exists), and can be found in the visual/spatial context (for instance, among the other macarons in the box).<sup>3</sup>

In this theory of demonstratives, Elbourne (2008) argues that “that” and “this” are forms with a relational component, for instance, distal and proximal factors: It has been argued that “this” tends to refer to things closer in context, and “that” to things further away (e.g. Kruisinga, 1925–32; Quirk, Greenbaum, Leech & Svartvik, 1985, and many others; see Scott, 2013, for a detailed analysis distinguishing the form-based preferences of both demonstrative pronouns.) Demonstratives also take a property as argument, and use both the property and the proximal/distal cue to map onto a specific referent (Reuter & Lew-Williams, 2018). For instance, in the sentence “This is the best macaron”, the comprehender will use the proximal cue and combine it with the property of being a macaron; ideally, this way the comprehender finds the referent that the speaker had in mind (presumably a very delicious macaron nearby.) Elbourne (2008) does not explicitly discuss reference to events, and how exactly the comprehender may map a demonstrative pronoun to a chunk of linguistic or non-linguistic conceptual structure is unclear; but we agree with the analysis that the demonstrative pronoun identifies an individual concept by means of reference (Loar, 1976; see also O’Madagain, 2020, for a compatible approach).

We argue furthermore that it is the function and purpose of the demonstrative to bundle a potentially complex, diffuse set of conceptual structure into such an individual concept, which the linguistic discourse can access and use down the road (cf. Grosz, 2018). Thus, in an extension of this model, we argue in the spirit of Wiese and Maling (2005) that demonstratives can serve as ‘universal bundlers’ for complex concepts, such as events.

---

<sup>3</sup> As a reviewer points out, the uniqueness requirement depends on context, and can apply to types as well as tokens. In a restaurant setting, asking for “the raspberry macaron” would be construed as “one of those macarons the chef makes”, whereas asking for a specific token of a macaron would likely be done with a demonstrative pronoun, as below.

## 1.2. A proposal: Demonstratives as ‘universal bundlers’

We propose an approach to demonstratives that views them as potential markers of a conceptual process that bundles a chunk of conceptual structure, and marks that chunk linguistically, such that the bundle can be referred to in the ongoing discourse. In this view, demonstratives tend to accomplish more, and have different goals, than simple pronouns.

Simple pronouns like “he” or “it” are indices that typically link the pronoun to an easily accessible noun phrase in the context, provided that this noun phrase adheres to the constraints posited in the relational and classificatory components of the pronoun (including discourse-level, lexical-level, syntactic, and semantic constraints). This is a complex process, as evidenced by the pronoun resolution literature; yet simple pronouns are still usually properly co-referential with a noun phrase in the discourse, except in the occasional case of an unheralded pronoun (Greene, Gerrig, McKoon & Ratcliff, 1994), or other rare exceptions.

Our account argues that the processes of simple pronoun resolution and demonstrative pronoun resolution, given the same context, are fundamentally different (for precursors of this idea, see e.g., Hankamer & Sag, 1976; Jackendoff, 2002, and probably others): Unlike simple pronouns, demonstratives serve as triggers for bundling chunks of conceptual or linguistic structure into an individual concept, such that this individual concept can serve as a referent for further discourse. This bundling procedure can be purely conceptual in nature, and the demonstrative pronoun is simply the linguistic marker for this ‘universal bundling.’<sup>4</sup>

We propose that non-personal simple pronouns like “it” and demonstrative pronouns like “that” have different form-specific constraints in English (see Kaiser & Trueswell 2008 for the form-specific multiple-constraints framework), and employ different psycholinguistic mechanisms to go from index to interpretation: Whereas “it” has a bias to quickly attach to the first noun phrase that satisfies both its relational and classificatory component, “that” can function as a ‘universal bundler’ that makes a chunk of conceptual structure (whether linguistically encoded or not) accessible to linguistic discourse.<sup>5</sup> In other words, demonstratives can serve as universal feature bundlers that allow the subsequent linguistic structure to further address the content of these bundles (or, in Lewis & Vasishth’s 2005 terminology, the content of these ‘chunks’).

This idea of a linguistic element serving as linguistic marker and trigger for a conceptual operation is obviously not new. Other ‘universal machines’ have been introduced in linguistics and philosophy to capture mapping functions that take conceptual objects as input, and yield continuous substances as output, or vice versa. The examples in (3) illustrate cases of the ‘universal grinder’, ‘universal sorter’, and ‘universal packer,’ all of which are conceptual operations triggered by mass and count syntax, respectively (Bunt, 1985; Pelletier 1975; Pelletier and Schubert, 1989; examples from Wiese & Maling, 2005):

- (3) a. There is chicken in the soup.  
(mass syntax with a typically count noun => ‘universal grinder’: animal to meat)

---

4 One prediction of this account is that once a complex conceptual structure has been bundled by a demonstrative (i-a), further reference to it with a simple non-personal pronoun (i-b) should be preferred over another demonstrative reference (i-c):

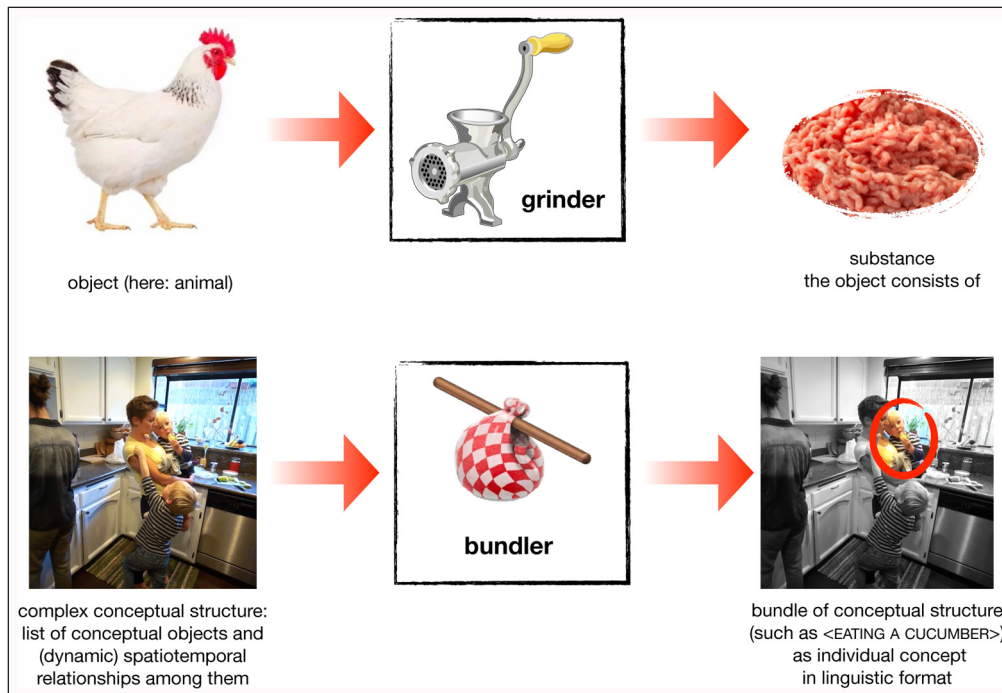
- (i) a. The friends rented an Airbnb near Gare du Nord. That was much cheaper than staying in a hotel.  
b. It was also more convenient than being further south.  
c. ?That was also more convenient than being further south.

This paper does not aim to answer this question, but we hope to test it in further research (see also Loáiciga et al., 2018 for similar findings).

5 Note that under some circumstances and in some contexts, “it” also does not need a linguistic antecedent. To us, these circumstances are limited to expletive “it”, cataphoric usages of the pronoun (e.g., “After it mates, the male bee dies”), or when the Question Under Discussion is abundantly clear (e.g., “Oh no! It’s clogged again”, standing in front of a toilet.) This fact does not change our analysis of the demonstrative’s role in reference. Also, we want to remind the reader that none of the observations on pronoun behavior are absolute; rather, they reflect preferences that have different probabilistic distributions. This is to say, one will likely be able to find a demonstrative pronoun triggering a simple reference process to an NP, and one will likely be able to find a non-personal pronoun triggering a conceptual bundling process, such as is needed in example 3a. Our goal here, as in most psycholinguistic studies, is to describe and explain and experimentally restricted situation that disentangles these referential mechanisms (Mook, 1983).

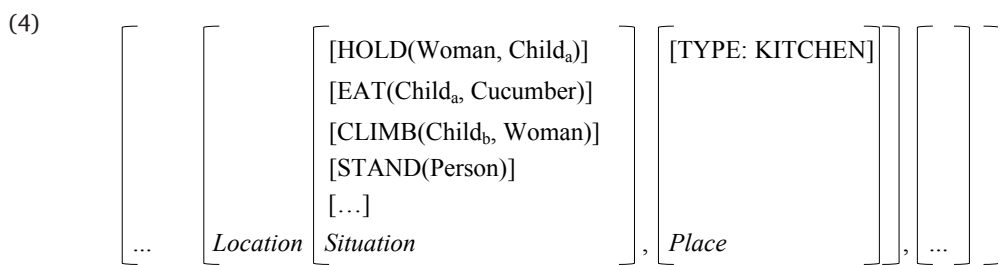
- b. The best wines are from Chile.  
 (count syntax with a typically mass noun => ‘universal sorter’: substances to kinds)
- c. Two beers and a coffee, please.  
 (count syntax with a typically mass noun => ‘universal packer’: substances to portions)

We propose that demonstratives are linguistic markers, analogous to mass or count syntax, of a conceptual operation that takes a conceptual structure as its input. This operation, in English, can only apply to conceptual structures that satisfy the demonstratives’ form-based relational and classificatory components. Its output is a bundle of conceptual structure (an ‘individual concept’) in a linguistic form, which can serve as a definite description of that individual concept in the ongoing linguistic discourse (see [Figure 1](#)).<sup>6</sup>



**Figure 1** Illustration of conceptual operations in two ‘universal machines’, after Wiese & Maling (2005).

For instance, an informal, simplified outline of the conceptual structure for the situation depicted on the lower left in [Figure 1](#) could roughly be sketched as Example (4):



Note that this sketch does not take into account the social, visual and spatiotemporal complexities that the static picture in [Figure 1](#) implies: The children’s and woman’s facial expressions and gaze, inferences about the persons’ ages and relations among each other, the architectural style of the kitchen in the picture, inferences about its location and inhabitants, and so on. All of these additional aspects of the event could be, in principle, included in a finer-grained representation

<sup>6</sup> Note that the bundling process can (potentially vacuously) bundle a set of one NP, such as in Example (ii):  
 (ii) I like Effi Briest. That is one good book!

In this case, the bundle would be identical with the referent of an NP in discourse. It is an empirical question whether the proposed bundling process actually takes place in these cases, one that we do not claim to answer in this paper.

of [Figure 1](#). Important to note is that out of this infinitely rich conceptual structure, a speaker can uniquely target any structure and substructure using a bare demonstrative:

(5) That looks like in my house!

The referent of Example (5) is necessarily vague without more context; “that” can bundle anything from the style of the cabinets to the mess on the counter. However, in the absence of prior discourse, once the speaker includes more lexical content to restrict the classificatory component of the pronoun, identification of the referent is easier:

- (6) a. That's unusual! I thought that children hate vegetables.  
Bundle: [EAT(Child<sub>a</sub>, Cucumber)]
- b. That's yummy, huh?  
Bundle: [Cucumber]
- c. Please don't do that!  
Bundle: [CLIMB(Child<sub>b</sub>, Woman)]

Thus, demonstratives tend to bundle up and refer to eventive or otherwise complex antecedents that are not encoded with a noun phrase, based on these observations and on the evidence from Brown-Schmidt et al. (2005) and Çokal et al. (2016). Hence, we propose the following form-based classificatory and relational properties of English demonstratives:

1. The *classificatory component* of standalone demonstratives specifies that the search for a referent needs to be restricted to a non-animate entity.
2. The *relational component* of standalone demonstratives specifies that the referent is not immediately accessible in context.

The classificatory component's restrictions are intuitively quite straightforward and easily verifiable. For instance, \*“I like Nils<sub>i</sub>. This<sub>i</sub> has such a sunny smile” is ungrammatical in English (even though the use of demonstratives in conjunction with copula verbs may be acceptable for some speakers, i.e., ?“I like Nils<sub>i</sub>. This<sub>i</sub> is such a happy kid”). We do not discuss this component further.

The relational component's specification can surface in varying ways. Linguistically, either the conceptual structure to be bundled up is (i) far away in the discourse (Çokal, Sturt, and Ferreira, 2014), or it is (ii) far away from upper layers in discourse structure, i.e., it is neither the current focus nor the current topic in discourse (Webber, 1989), and is less salient on the level of discourse (Gundel et al., 1993) or the level of argument structure (Chafe, 1976; Brennan, Friedman, & Pollard, 1987).

But the referent of a demonstrative does not have to be linguistically expressed, as evidenced by examples (5) or (6) above, drawn from [Figure 1](#). It can also be a percept in the visual or auditory domain, or another non-linguistic and purely conceptual structure, as (7) shows (e.g., Kaplan, 1989; Jackendoff, 2002):

- (7) a. [*gesture at terrible shirt*] I cannot believe you want to go out dressed like that.  
(= bundle from a visual conceptual structure)
- b. [*screeching sound*] What was that?  
(= bundle from an auditory conceptual structure)
- c. [*smelling cigarette smoke outside*] That must be my mom.  
(= bundle from an olfactory conceptual structure)
- d. [*confronted with a surprise party, friends, and a birthday song*] This is too much!  
(= bundle from a multisensory conceptual structure)

Crucially, demonstratives can not only bundle up non-linguistic conceptual structures, but also conceptual structures that have been encoded linguistically, for instance, when the conceptual structure to be bundled up is not an easily co-referential noun phrase. In the case of Brown-Schmidt et al.'s (2005) study, this bundle contained both the cup and the saucer. The data from their study provides evidence that demonstratives may be used as a cue to gather and bundle up concepts expressed by noun phrases. But how do people refer back to things that are realized linguistically but *not* in the form of noun phrases, such as in (6a) and (6c)? To answer

this question, we turn to event descriptions, such as in “The friends visited Paris”, “The hikers explored the forest”, or “Adam heated the lasagna”.

### 1.3. Current studies and predictions

In this paper, we use reference to objects (linguistically realized as NPs) and events (conveyed by an entire clause) to investigate whether non-personal pronouns like “it” and stand-alone demonstrative pronouns like “this” or “that” access different cognitive mechanisms in reference resolution in English.

We use sentence pairs like in (8), where a context sentence sets up a potential referent for the pronoun that is present in the next (critical) sentence: “it” refers to “lasagna”; “that”, to the act of making the lasagna. We use “it” and “that” because they are distributionally more similar to each other than “it” and “this” (Strauss, 2002).

- (8) Sentence 1: Adam made lasagna for me last night.  
Sentence 2:  
a. It was really amazing.  
b. That was really amazing.

Specifically, we propose that non-personal pronouns are interpreted as coreferential with (salient) lexical item that satisfies the pronoun’s form-specific relational and classificatory constraints: inanimate, grammatically singular noun phrases (“lasagna” in (8)). We propose that, in the context of our experiment, this is a process in which only the linguistic surface needs to be implicated (cf. Hankamer & Sag, 1976): In order to determine the referent of “it”, it is sufficient to access form-based information of the candidate words in context. In languages with grammatical gender, this information may be morphological information (Cacciari, Carreiras & Barbolini Cionini, 1997); but also, lexemic properties such as length and frequency of a word (Simner & Smyth, 1999; Duffy & Rayner, 1990). Thus, we predict that “it” should be sensitive to frequency and length of the preceding linguistic material resolution (although see Egusquiza, Navarrete & Zawiszewski, 2016, and Lago, 2014, for failure to find frequency effects with personal pronouns).

We also predict that the demonstrative “that” should be sensitive to the same surface-form-based features, because these features need to be accessed for any kind of reference. Overall, we predict that after encountering “it” or “that” at the start of the critical sentence, reading times – which reflect processing ease – will reveal effects of the antecedents’ surface properties.

We further predict that “that” will lead to slower reading times than “it.” This is because under our approach, “that” crucially differs from “it” in that only “that” is accessing and bundling up complex conceptual structures – processes which can be assumed to carry a cognitive cost. Thus, we predict that “that” should be read slower than “it” throughout, above and beyond the difference in orthographic length of the pronouns.

Additionally, we predict that “that” should be uniquely sensitive to higher-level conceptual features, such as the complexity of a concept. Thus, we manipulate the conceptual complexity of an event (as discussed below) that a subsequent demonstrative “that” will refer back to.<sup>7</sup> In particular, we predict that (i) more complex events will lead to faster reading times than less complex events for sentences where “that” is used to refer to the event, whereas (ii) we do not predict effects of event complexity for sentences where “it” is used to refer to the event. This prediction originates from literature suggesting that semantically rich representations lead to faster re-access than semantically poor representations (Fisher & Craik, 1980; Craik & Tulving, 1975; Gallo, Meadow, Johnson & Foster, 2008; van Gompel & Majid, 2004; Heine et al., 2006a,b; Hofmeister, 2011; Karimi & Ferreira, 2016).

Finally, we also make an additional prediction that when non-personal pronouns are subsequently specified by event-denoting adjectives (such as “It was very adventurous” or “It was quite laborious”), comprehension should slow down at the adjective, due to an effect of mismatched expectations; likewise, we expect the same when demonstratives are subsequently

---

<sup>7</sup> In this paper, we only talk about eventive referents, since the distinction between events, states, situation, and facts is not crucial for the claims we are making here: Looking at non-eventive states and situations, as well as at different kinds of conceptual features, is an important direction for future work.



specified by object-denoting adjectives (such as “*That* was very small” or “*That* was quite pretty”). We refer to the difference between adjectives like “adventurous” (event-denoting) and “small” (object-denoting) as *adjective bias*. Observing these kinds of slow-down patterns would indicate that readers consider an event reference more when they have read a demonstrative, and an object reference more when they have read a non-personal pronoun. Violations of these expectations would result in a type mismatch, as an interaction of adjective bias with pronoun type. However, we issue this prediction with caution for two reasons: In order to limit the length of the experiment, we introduced such a mismatch only for half of our experimental items, and thus, power is significantly reduced for this analysis. Second, the adjectives were in sentence-final position, which has been associated with complex sentence-wrap up effects (e.g., Warren, White & Reichle, 2009).

In what follows, we present two experiments, each replicated once with a slight difference in stimuli, to test these hypotheses. Both experiments allow us to test the prediction that “that” is read more slowly than “it”, and that adjective bias may interact with pronoun type. Experiments 1a and 1b investigate whether both non-personal and demonstrative pronouns are sensitive to surface features of the linguistic context, including those of the potential referents. Experiments 2a and 2b test the prediction that only demonstrative pronouns are sensitive to higher-level features of the linguistic context.

## 2. Experiment 1a

This study tests the hypothesis that both “it” and “that” access the surface features of a potential referent, the prediction that “that” is read more slowly than “it”, and the interaction of adjective bias with pronoun type. We operationalize surface features as lexical frequency and word length in number of letters, which are inversely correlated: longer words are usually less frequent (e.g., Kliegl, Grabner, Rolfs & Engbert, 2004).

### 2.1. Methods

#### 2.1.1. Participants

200 self-described native English speakers with IP addresses within the United States, recruited from Amazon Mechanical Turk, participated in the experiment for monetary compensation. Mechanical Turk is used widely in research because it allows access to a large number of study participants, and most results, although perhaps somewhat noisier, are comparable to results obtained in the lab (e.g., Mason & Suri, 2012; Munro et al., 2010; Sprouse, 2011).

#### 2.1.2. Materials

We created 40 sets of stimuli, consisting of a sequence of two sentences, as shown in (9) and (10):

- (9) Context Sentence:  
 a. The hikers explored the forest. [higher frequency noun]  
 b. The hikers explored the jungle. [lower frequency noun]
- (10) Critical Sentence:  
 a. That was really adventurous.  
 b. It was really adventurous.

The context sentence in each set (9a or b) always contained an animate subject, and a singular non-personal object, that is, a potential referent for “it”. The critical sentence started with a pronoun (“it” or “that”), continued with a copula verb, and ended with an intensifier and an adjective (10a or b). Half of the adjectives were compatible with an event reading like in (9) and (10), where “adventurous” refers to the whole exploring event; and half of the adjectives were compatible with the noun phrase (which would have been “forest” or “jungle”).<sup>8</sup> We refer to this as the adjective bias manipulation. This was done to ensure that participants would not be inadvertently learning a pattern of (in)compatibility throughout the course of

<sup>8</sup> An example of a stimulus sentence with an object-compatible adjective:

(iii) The girl baked the bread. It was really tasty.

the experiment (Fine et al., 2013), and allows us to test the prediction that that pronoun type should interact with reference type signaled by the adjective (object vs. event).

The experimental manipulations were (i) pronoun type (“that”/“it”) and (ii) noun frequency (“forest” = high frequency, “jungle” = low frequency) in the context sentence, using synonyms or semantically closely related nouns. As described above, we also manipulated (iii) adjective bias (event-denoting vs. object-denoting) between items. Frequency was determined comparing each noun pair in the Celex corpus (Baayen, Piepenbrock, & Gulikers, 1995): low frequency nouns were always less frequent than high frequency nouns. As expected, noun length was inversely correlated with frequency: Low-frequency nouns were significantly longer on average (average length: 6.1 characters) than high-frequency nouns (average length: 5.4 characters;  $F(1,40) = 4.31$ ;  $p < .05$ ). Since both noun length and noun frequency are surface level factors, in this paper we do not aim to separate these factors from each other.

In order to ensure that the pronoun “it” is indeed biased to refer to the noun phrase, and the demonstrative “that” to the event, we conducted a norming study. We created a forced-choice rating task, in which we presented each scenario but replaced the final adjective (e.g., “adventurous” in (10)) with the nonsense adjective “dax.” This resulted in sentences like “The hikers entered the forest. That was really dax.” The adjective replacement was done in order to prevent semantic interference from the final adjective, thus better reflecting interpretations at our regions of interest, before participants read the whole scenario. Participants (40 native speakers of English from Amazon Mechanical Turk) were asked to indicate whether the object (“the forest that the hikers entered”) or the event (“that the hikers entered the forest”) was “dax.” Confirming our intuition, in sentences that contained the pronoun “it”, people overwhelmingly chose the object meaning (67.9%); in sentences that contained the demonstrative “that”, however, people strongly dispreferred the object meaning (24.2% object meaning;  $\beta = 1.7$ ,  $p < .0001$  in a mixed binomial regression). These norming data show that, in the presence of a neutral adjective, the object interpretation in sentences containing “it” was much more likely than the event interpretation; and the reverse was true for sentences containing “that.”

In the main experiment, we presented trials in random order in a masked self-paced reading paradigm, together with 40 filler items, using Ibx, an experiment software and platform tailored to the self-paced reading paradigm (Drummond, 2014). Each filler was followed by a comprehension question.<sup>9</sup> In addition to the sentences that people read word-by-word, there were forty comprehension questions in total. However, when calculating accuracy statistics, one comprehension question was removed, because the answer to the question was coded wrongly.

We predict that both “it” and “that” access the surface features of the context, and thus, we expect a main effect of noun frequency in the critical sentence, at and after the pronoun: People should read both pronoun types faster after sentences containing high-frequency nouns than after sentences with lower-frequency nouns. We also predict a main effect of pronoun, namely that the demonstrative should lead to slower reading times than a simple pronoun; and an interaction of adjective bias with pronoun type.

### 2.1.3. Data Analysis

For this and all other experiments, we used mixed-effects regression models on log-transformed data (Baayen, Davidson, & Bates, 2008) with R’s *lme4* package (Bates et al., 2014) to analyze the reading times for each word. As justified by the design, we implemented a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013).<sup>10</sup> Where noted, the regression structure was modified to ensure model convergence by eliminating interactions on random effects first, then, if necessary, taking out subject random effects, or item random effects. In Experiments 1a and 1b, our fixed effects were Frequency (high vs. low) and Pronoun (“it” vs. “that”); in Experiments 2a and 2b, the fixed effects were Complexity (high vs. low) and Pronoun (“it” vs. “that”). All of these categorical factors were centered (i.e., coded as  $-0.5$  and  $0.5$ ) in regression analyses. In the adjective region, we also included Adjective Bias as a predictor, i.e., whether the final adjective was more compatible with an event (“adventurous”) or an object (“wild”).

<sup>9</sup> The full list of stimuli can be found under <https://osf.io/59qzj/>.

<sup>10</sup> For each experiment, we also ran separate models residualizing over the length of pronoun in letters, but since these models did not change the results patterns significantly, we report only the standard models. Full scripts and results can be found under <https://osf.io/59qzj/>.

## 2.2. Results

We excluded 34 participants because of low comprehension-question accuracy (<75%), or because their median reading times were below 200 ms or above 1500 ms. The accuracy of comprehension questions in the remaining 166 participants was 92% (range = 79% ~ 100%). Reading times above 2000 ms (<.5% of the data) or below 100 ms (<1% of the data) were also excluded.

Figure 2 shows log reading times for each region in Experiment 1, and Table 1 shows the mean reading times with Standard Errors. A summary of the regression results is shown in Table 2. While our predictions only pertain to the critical sentence, we report reading times in each region (including the context sentence) for completeness's sake.

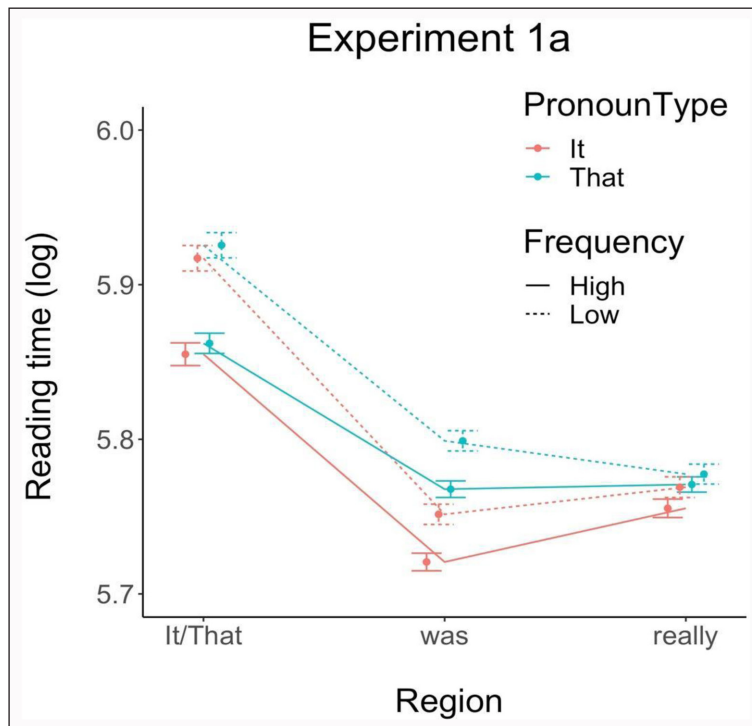


Figure 2 Region-by-region reading times in the critical sentence, on log scale, in Experiment 1a.

In the context sentence, there was a spurious effect of pronoun in the very first region, the subject NP (“the hunters”). Spurious effects are very common in self-paced reading experiments (e.g. Omaki, Lau, Davidson White, Dakan, Apple & Phillips, 2015; Meng & Bader, 2020; among many others). We classify effects as spurious when they fulfil two criteria: When they could not have been introduced by our manipulation, i.e. because they occurred before the manipulation, and when they do not consistently occur between experiments or across subsequent regions.

As predicted, starting in the *object noun* region (“the forest/jungle”), people read faster starting at a high-frequency noun (“forest”), than after a low frequency noun (“jungle”). This *main effect of frequency* remains significant over three contiguous regions: the **object noun** region (the region where frequency was directly manipulated) in the context sentence, and the **pronoun region** and the **copular verb** region in the critical sentence. It is marginal on the subsequent adverb region. Furthermore, we also find a *main effect of pronoun type*, with “it” conditions being read faster than “that” conditions, at the copular verb that immediately follows the pronoun as well as at the immediately following adverb. There are no interactions involving frequency and pronoun type anywhere in either of the two sentences.

We also analyzed the **adjective region** (“adventurous”), with adjective bias as an additional fixed predictor: Here, no main effects were significant. We also did not find any significant interactions of adjective bias with complexity or pronoun type (all *ps* > .15).

## 3. Experiment 1b

This experiment was a replication of Experiment 1a, with one slight change: We included a spillover region after the critical noun (“forest”/“jungle”), to ensure that the object-induced

Region	Object frequency	Pronoun	RT	SE
Subject NP <i>The hikers</i>	High	<i>it</i>	605	7
		<i>that</i>	601	7
	Low	<i>it</i>	603	6
		<i>that</i>	596	6
Verb <i>entered</i>	High	<i>it</i>	386	4
		<i>that</i>	383	4
	Low	<i>it</i>	380	3
		<i>that</i>	387	4
Object NP <i>the forest/jungle.</i>	High	<i>it</i>	747	6
		<i>that</i>	741	6
	Low	<i>it</i>	771	6
		<i>that</i>	758	6
Pronoun <i>It/That</i>	High	<i>it</i>	379	4
		<i>that</i>	384	3
	Low	<i>it</i>	407	4
		<i>that</i>	416	4
Copula verb <i>was</i>	High	<i>it</i>	323	3
		<i>that</i>	340	2
	Low	<i>it</i>	334	3
		<i>that</i>	353	3
adverb <i>very</i>	High	<i>it</i>	341	3
		<i>that</i>	342	2
	Low	<i>it</i>	344	3
		<i>that</i>	346	3
Adjective <i>adventurous</i> (or an object-biased adjective)	High	<i>it</i>	487/485	8/10
		<i>that</i>	489/473	11/8
	Low	<i>it</i>	512/476	11/9
		<i>that</i>	496/484	11/9

**Table 1** Mean region-by-region reading times in milliseconds with Standard Error of the Mean for Experiment 1a. In the last region, mean and SEM was separately reported for each adjective bias condition (event bias/object bias).

frequency effect found on the pronoun in Experiment 1a was not due to a simple spillover effect of people slowing down, in general, after reading low-frequency words.

### 3.1. Methods

The method was the same as Experiment 1a: We used masked self-paced reading on Ibex, hosted on Ibex farm.

#### 3.1.1. Participants

We recruited a different set of 200 native speakers on Amazon Mechanical Turk.

#### 3.1.2. Materials

We used the 40 sets of stimuli used in Experiment 1a, consisting of pairs like in (11) and (12), but added an adjunctive two-word spillover region (“at night”) specifying a location or time:

	The hunters				Entered				The forest/the jungle			
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P
Frequency	0.00	0.01	0.43	0.67	0.00	0.01	-0.26	0.80	0.01	0.01	2.11	<b>0.04</b>
Pronoun type	-0.01	0.01	-2.14	<b>0.03</b>	0.00	0.01	-0.06	0.95	-0.01	0.01	-1.34	0.18
Freq. × Pro. type	0.01	0.01	0.60	0.55	0.02	0.01	1.10	0.27	0.00	0.01	0.03	0.98

	It/That				Was				Very				Adventurous			
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P
Frequency	0.06	0.01	9.30	<b>0.00</b>	0.03	0.01	5.52	<b>0.00</b>	0.01	0.01	1.57	0.08	0.01	0.01	1.43	0.15
Pronoun type	0.01	0.01	1.00	0.31	0.05	0.01	8.32	<b>0.00</b>	0.01	0.01	2.04	<b>0.04</b>	0.01	0.01	-0.79	0.43
Freq. × Pro. type	0.01	0.01	0.46	0.65	0.00	0.01	0.23	0.82	-0.01	0.01	-0.54	0.59	0.00	0.02	-0.23	0.82

- (11) Context sentence:  
 a. The hikers explored the forest at night.  
 b. The hikers explored the jungle at night.
- (12) Critical sentence:  
 a. That was really adventurous.  
 b. It was really adventurous.

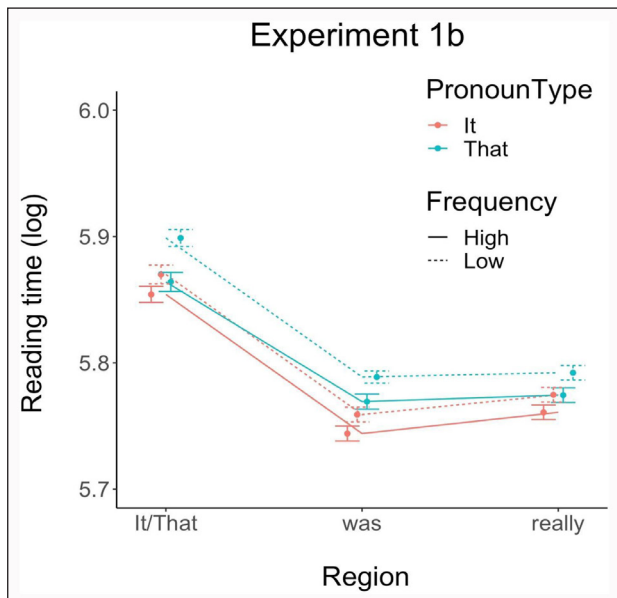
**Table 2** Summary of mixed effects model results by region in Experiment 1a. P-values indicating significant effects are in bold; darker grey shading indicates significance at 5% level, light grey, marginal significance. Effects of adjective bias are only reported in the text.

Our predictions were the same as in Experiment 1a, namely a main effect of frequency, a main effect of pronoun, and an interaction of adjective bias with pronoun type.

### 3.2. Results of Experiment 1b

Of 200 initial participants, we excluded 29 based on the same criteria as in Experiment 1a: Either comprehension question accuracy of less than 75% (N = 11), median reading time (across all regions) of less than 200 ms (N = 19) and more than 1500 ms (N = 0). As before, we also excluded trials with reading times above 2000 ms (2.5% of observations) or less than 100 ms (<1% of observations). The accuracy of comprehension questions in the remaining participants was 94% (range = 77% ~ 100%).

Figure 3 shows log reading times for each region in Experiment 1b; Table 3 shows the mean reading times with Standard Errors, and Table 4 shows regression results on raw reading times over the critical regions.



**Figure 3** Region-by-region reading times in Sentence 2, on log scale, in Experiment 1b.

Region	Object frequency	Pronoun	RT	SE
Subject NP <i>The hikers</i>	High	<i>it</i>	601	6
		<i>that</i>	598	7
	Low	<i>it</i>	601	6
		<i>that</i>	606	6
Verb <i>entered</i>	High	<i>it</i>	383	4
		<i>that</i>	380	4
	Low	<i>it</i>	380	4
		<i>that</i>	381	4
Object NP <i>the forest</i>	High	<i>it</i>	734	5
		<i>that</i>	727	6
	Low	<i>it</i>	733	5
		<i>that</i>	744	5
Spillover I <i>at</i>	High	<i>it</i>	364	3
		<i>that</i>	357	3
	Low	<i>it</i>	387	3
		<i>that</i>	389	4
Spillover II <i>night.</i>	High	<i>it</i>	389	4
		<i>that</i>	380	5
	Low	<i>it</i>	401	4
		<i>that</i>	408	4
Pronoun <i>It/That</i>	High	<i>it</i>	375	3
		<i>that</i>	383	4
	Low	<i>it</i>	381	4
		<i>that</i>	398	3
Copula <i>was</i>	High	<i>it</i>	331	3
		<i>that</i>	340	2
	Low	<i>it</i>	334	2
		<i>that</i>	345	2
Adverb <i>very</i>	High	<i>it</i>	337	2
		<i>that</i>	341	2
	Low	<i>it</i>	343	3
		<i>that</i>	347	2
Adjective <i>adventurous</i> (or an object-biased adjective)	High	<i>it</i>	496/508	9/10
		<i>that</i>	518/512	12/8
	Low	<i>it</i>	510/511	8/9
		<i>that</i>	543/529	11/11

**Table 3** Mean region-by-region reading times in milliseconds and SEM for Experiment 1b. In the last region, mean and SEM was separately reported for each adjective bias condition (event bias/object bias).

We find a spurious effect at the Context Sentence's subject as a main effect of frequency. Predicted effects of our frequency manipulation started at the object noun: High frequency nouns were read faster than low-frequency nouns. This effect of frequency continued and did not subside until the end of the second sentence; specifically, pertaining to our predictions,

**Table 4** Summary of mixed effects model results by region in Experiment 1b. P-values indicating significant effects are in bold; darker grey shading indicates significance at 5% level, light grey, marginal significance. Effects of adjective bias are only reported in the text.

	The hunters				Entered				The forest				AT				Night					
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P		
Frequency	0.05	0.01	6.58	0.00	0.01	0.01	1.37	0.17	0.01	0.01	0.01	0.02	0.01	0.01	2.27	0.02	0.06	0.01	0.01	0.02	8.78	0.00
Pronoun type	0.01	0.01	-1.01	0.31	0.00	0.01	-0.16	0.88	0.01	0.01	0.01	0.80	0.00	0.01	-0.26	0.80	0.01	0.01	0.01	0.80	-1.36	0.17
Freq. x Pro. type	0.02	0.01	1.66	0.10	0.00	0.01	-0.07	0.95	0.02	0.01	1.63	0.10	0.02	0.01	1.63	0.10	0.02	0.01	0.01	0.10	1.46	0.14
<b>It/That</b>																						
<b>Was</b>																						
<b>Very</b>																						
<b>Adventurous</b>																						
Frequency	0.05	0.01	6.70	<b>0.00</b>	0.03	0.01	4.03	<b>0.00</b>	0.02	0.01	3.18	<b>0.00</b>	0.02	0.01	2.64	<b>0.01</b>						
Pronoun type	0.01	0.01	-1.01				0.32	0.02	0.01	2.85	<b>0.00</b>	0.03	0.01	4.90		<b>0.00</b>	0.01	0.01	1.81	<b>0.07</b>		
Freq. x Pro. type	0.02	0.01	1.68	<b>0.09</b>	0.02	0.01	1.30	0.18	0.00	0.01	0.40	0.69	0.01	0.02	0.67	0.50						

starting in the **pronoun region** (“that/it”), people read faster after a high-frequency noun (“forest”), compared after a low frequency noun (“jungle”), with a significant main effect of frequency. Importantly, at the following **copular** (“was”), at the **adverb** (“very”), and, marginally, in the sentence-final **adjective**, we also found a main effect of pronoun type, with people reading faster after “it” than “that.”

We also found a marginal interaction of pronoun type with frequency, at the pronoun region, with an unpredicted slightly larger effect of frequency for demonstratives than for pronouns. Again, we did not find any significant interactions of **adjective bias** with complexity or pronoun type (all  $ps > .34$ ).<sup>11</sup>

## 4. Discussion of Experiments 1a and 1b

Experiments 1a and 1b both show that surface properties, operationalized here as frequency (which is also correlated inversely with word length) of the antecedent object noun, influence the speed with which people process both non-personal “it” and demonstrative “that.” This confirms our prediction that *both pronoun types* are sensitive to surface features in the linguistic context. We did not find a mismatch effect for adjective bias, contrary to what we expected based on our norming data. This may be due to reduced sample size for that region, since the adjective-bias manipulation was between-items, leading to a loss of power; and also due to their position as sentence-final words, which has been shown to introduce complex wrap-up effects (e.g., Warren et al., 2009).

In addition, we had predicted that people would read faster after they resolve the non-personal pronoun “it” compared to the demonstrative “that”. Our data confirmed this prediction. We argue that this effect is based on our model of conceptual bundling: When resolving the demonstrative “that”, readers execute a different, perhaps more extensive, search for a referent or a referential structure, than when resolving the non-personal pronoun “it”. These effects are unlikely to be due to short-lived spillover from orthographic differences between the two referring expressions, since these effects remain significant over two regions in Exp. 1a and three regions in Exp. 1b.

In addition to showing that demonstratives and non-personal pronouns lead to subsequent differences in reading behavior, and thus providing evidence for our main hypothesis that demonstratives accomplish a fundamentally different operation than pronouns, these results also provide a crucial foundation for Experiments 2a and 2b, which test the prediction that only “that”, and not “it”, is uniquely sensitive to higher-level conceptual features of its referent. This prediction about the asymmetrical sensitivity of the two pronoun types to conceptual features is derived from our claim that demonstratives are universal bundlers that make a chunk of conceptual structure available to the discourse, and thus have to access the conceptual, not only surface, features of the referent. Non-personal pronouns, in contrast, do not act as bundlers and thus are not expected to show the same level of sensitivity to conceptual properties of referents. To test this claim, in Experiments 2a and 2b we manipulate the conceptual complexity of an event that a subsequent demonstrative “that” (or non-personal pronoun “it”) will refer back to.

Based on previous studies arguing for semantically richer representations leading to faster re-access (e.g., van Gompel & Majid, 2004; Heine et al., 2006a,b; Hofmeister, 2011; Karimi & Ferreira, 2016), we predict an interaction: reading “that” will be faster after complex events, but reading times for “it” will not be affected, since “it” tends to refer only to the object, not to the whole event.

## 5. Experiment 2a

In the following, we test our prediction that only “that”, and not “it”, is uniquely sensitive to higher-level conceptual features of its referent, following the claim that demonstratives tend to be universal bundlers that make a chunk of conceptual structure available to the discourse, and thus have to access the conceptual, not only surface, features of the referent. In addition,

---

<sup>11</sup> Supplemental analyses as well as raw (anonymized) data for all experiments, including the norming studies, can be found under <https://osf.io/59qzj/>.



we seek to replicate our findings from Experiments 1a and 1b, namely that demonstratives lead to slower reading times than non-personal pronouns; and we test our prediction that adjective bias may interact with pronoun type.

## 5.1. Methods

### 5.1.1. Participants

200 new native English speakers from Amazon Mechanical Turk participated in the experiment for monetary compensation.

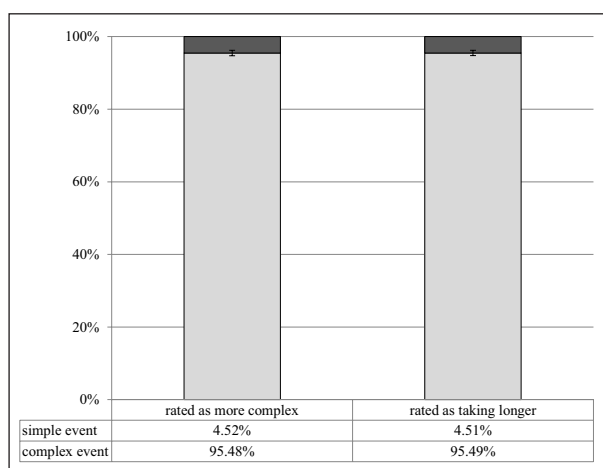
### 5.1.2. Materials

The same experimental item sets as in Experiments 1a and 1b were used, but they were adjusted to manipulate event complexity instead of nouns' linguistic frequency: We used complex events, such as “explore” (13a), and simple events, like “enter” (13b), combined with the high-frequency nouns of Experiment 1 (e.g., “forest”).

- (13) Context Sentence:
- a. The hikers explored the forest.
  - b. The hikers entered the forest.
- (14) Critical Sentence:
- a. That was really adventurous.
  - b. It was really adventurous.

Simple events were created by using presupposed or potential sub-events of complex events. For instance, exploring a place (complex) presupposes entering that place (simple), or cleaning a room (complex) may include sweeping it (simple). Verbs' lexical frequency was matched, using the Celex English wordform database (simple event verbs: 18 per million words, complex event verbs: 17 per million words; no statistically significant difference, as determined by a one-way ANOVA ( $F(1,78) = .01, p > .92$ ). If we find a reading time difference, it should not be due to frequency.

Event complexity was normed in two different ways (see [Figure 4](#)). First, we gave 20 native English speakers on Amazon Mechanical Turk a forced-choice task in which they had to rate which event was conceptually more complex, pitting a simple sub-event (i.e. 13a) against its more complex counterpart (i.e. 13b).<sup>12</sup> Our “explore”-type events were rated as more complex than the “enter”-type events 95.48% of the time. Second, since more complex events often take longer than simple events, we asked 20 different native English speakers on Amazon Mechanical Turk to rate the duration of the events, in another forced-choice test (Wittenberg & Levy, 2017).<sup>13</sup>



**Figure 4** Norming of events used in Experiment 2; complex (e.g., “explored the forest”) is shaded light grey, simple is shaded dark grey (e.g., “entered the forest”).

<sup>12</sup> The exact instructions were: “Your task is simply to imagine the described actions, and tell us which action is more complicated compared to the other. What does it mean to be more complicated? Just imagine what needs to happen in each. For instance, “eating an apple” may be less complicated than “slicing an apple”, because the latter involves more hand movements, with an instrument (a knife), and it may take more time. “Smelling an apple”, on the other hand, may be less complicated than eating it — it only takes a moment, there is no movement involved, and it happens effortlessly and automatically.”

<sup>13</sup> The exact instructions were: “Your task is simply to imagine the described actions, and tell us which action takes longer compared to the other. For instance, “slicing an apple” may take longer than “eating an apple”. “Smelling an apple”, on the other hand, may take even less time.”

Here, 95.49% of the “explore”-type events were rated as taking longer than the “enter”-type events. These results taken together indicate that the items were constructed and classified appropriately into two conceptual, nonlinguistic classes: more complex and less complex events.

Again, the critical items were presented randomly in a self-paced reading paradigm, together with the same 40 filler items as in Experiment 1a and 1b, using Ibex. Each filler was followed by a comprehension question.<sup>14</sup>

## 5.2. Results

We excluded 37 participants because of low question accuracy, or because their median reading times were below 200 ms or above 1500 ms. Reading times slower than 2000 ms (<2% of observations) or faster than 100 ms (<2% of observations) were excluded as well. The accuracy of comprehension questions in the remaining participants was 94% (range = 75% ~ 100%).

Figure 5 shows log reading times for each region in Experiment 2a; Table 5 shows the mean reading times with Standard Errors, and Table 6 shows model results for the critical regions. Before the pronoun region, no significant effects were found ( $ps > .05$ ), except for a main effect of Complexity at the object NP following the manipulated verb, but this effect had subsided by the next region. As predicted, and consistent with results from Experiments 1a and 1b, we found a main effect of pronoun type starting in the **pronoun region** (“that/it”), which remained significant throughout the rest of the trial: “it” conditions were read faster than “that” conditions. No other effects were significant, except for a main effect of complexity at the **sentence-final adjective**; we did not find any significant interactions of adjective bias with complexity or pronoun type (all  $ps > .38$ ).

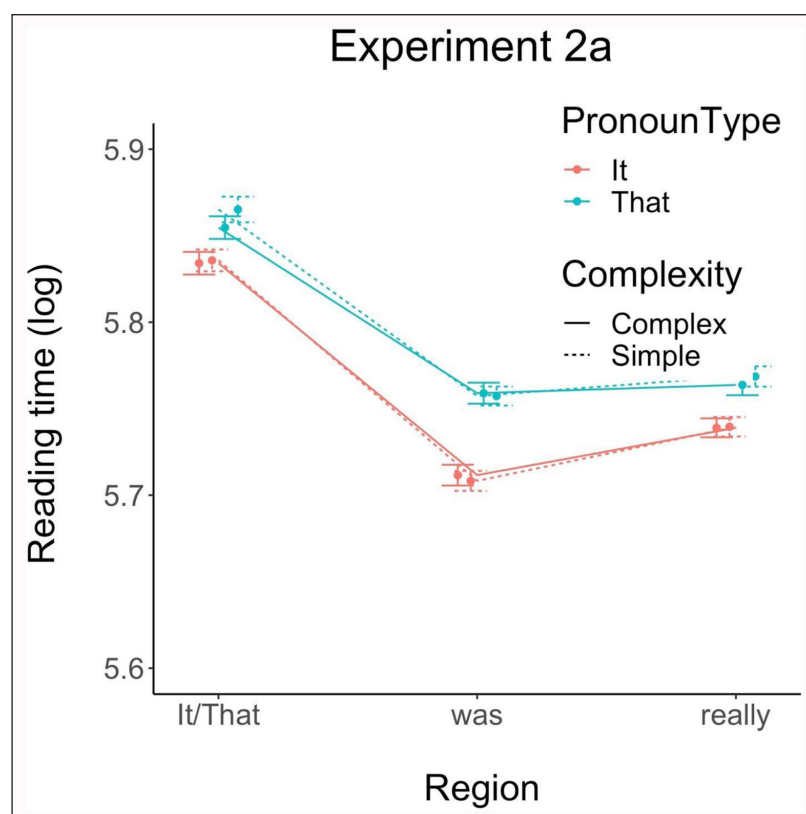


Figure 5 Region-by-region reading times in the Critical Sentence, on log scale, in Experiment 2a.

## 6. Experiment 2b

This experiment was a replication of Experiment 2a, with the same change as from Experiment 1a to Experiment 1b: We included a spillover region after the critical noun (“forest”).

<sup>14</sup> The full list of stimuli can be found under <https://osf.io/59qzj/>.

Region	Complexity	Pronoun	RT	SE
Subject NP <i>The hikers</i>	Complex	<i>it</i>	579	5
		<i>that</i>	585	6
	Simple	<i>it</i>	578	7
		<i>that</i>	578	6
Verb <i>entered/explored</i>	Complex	<i>it</i>	367	3
		<i>that</i>	371	4
	Simple	<i>it</i>	359	3
		<i>that</i>	366	4
Object NP <i>the forest.</i>	Complex	<i>it</i>	735	5
		<i>that</i>	721	5
	Simple	<i>it</i>	722	5
		<i>that</i>	726	5
Pronoun <i>It/That</i>	Complex	<i>it</i>	367	3
		<i>that</i>	380	3
	Simple	<i>it</i>	370	3
		<i>that</i>	384	4
Copula <i>was</i>	Complex	<i>it</i>	317	2
		<i>that</i>	334	2
	Simple	<i>it</i>	317	2
		<i>that</i>	333	2
Adverb <i>very</i>	Complex	<i>it</i>	330	2
		<i>that</i>	341	3
	Simple	<i>it</i>	330	2
		<i>that</i>	339	2
Adjective <i>adventurous</i> (or an object-selecting adjective)	Complex	<i>it</i>	493/472	9/7
		<i>that</i>	526/489	13/8
	Simple	<i>it</i>	500/478	10/7
		<i>that</i>	509/508	10/10

**Table 5** Mean region-by-region reading times and SEM for Experiment 2a. In the last region, mean and SEM was separately reported for each adjective bias condition (event bias/object bias).

## 6.1. Methods

### 6.1.1. Participants

A different set of 200 self-described native English speakers from Amazon Mechanical Turk participated in the experiment for monetary compensation.

### 6.1.2. Materials

We changed the 40 sets of stimuli used in Experiment 2a, consisting of pairs like in (15) and (16), such that they contained a two-word spillover region ('at night'):

- (15) Context Sentence:  
 a. The hikers explored the forest at night.  
 b. The hikers entered the forest at night.
- (16) Critical Sentence  
 a. That was really adventurous.  
 b. It was really adventurous.

	The hunters				Entered/Explored				The forest							
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P				
Complexity	0.00	0.01	0.63	0.53	0.01	0.01	-1.63	0.11	-0.01	0.01	-2.29	<b>0.02</b>				
Pronoun type	0.00	0.01	-0.12	0.90	0.01	0.01	1.44	0.15	-0.01	0.01	-1.01	0.31				
Comp. × Pro. type	0.00	0.01	-0.49	0.63	0.01	0.01	0.68	0.50	0.02	0.01	1.59	0.11				
	It/That				Was				Very				Adventurous			
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P
Complexity	0.01	0.01	0.70	0.48	0.00	0.01	-0.43	0.67	0.00	0.01	0.55	0.58	0.02	0.01	2.11	<b>0.03</b>
Pronoun type	0.03	0.01	3.68	<b>0.00</b>	0.05	0.01	8.84	<b>0.00</b>	0.03	0.01	4.78	<b>0.00</b>	0.03	0.01	3.34	<b>0.00</b>
Comp. × Pro. type	0.01	0.01	0.66	0.51	0.00	0.01	0.19	0.85	0.00	0.01	0.35	0.73	0.01	0.02	0.84	0.40

### 6.1.3. Procedure

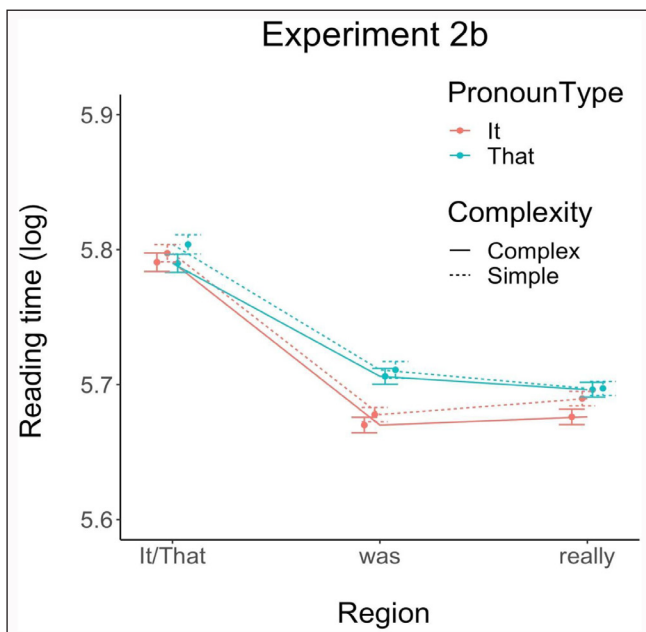
Again, we used self-paced masked reading, collecting data over the internet.

## 6.2. Results

From the initial set of 200 speakers, we excluded 25, because of low accuracy on comprehension questions ( $N = 3$ ), or due to median reading times being too fast ( $N = 20$ ) or too slow ( $N = 2$ ). The accuracy of comprehension questions in the remaining participants was 94% (range = 78% ~ 100%). Again, we also excluded individual trials based on reading times: Longer than 2000 ms (<3% of observations) or shorter than 100 ms (<2% of observations).

Figure 6 shows log reading times for each region in Experiment 2b; Table 7 shows the mean reading times with Standard Errors, and Table 8 shows results of the regression.<sup>15</sup> We find a spurious main effect, and spurious effects of interactions with pronoun type in four regions of the context sentence. However, as expected, and as found in Experiment 2a, we also found a main effect of complexity at the manipulated verb (“entered/explored”) and the object NP. Reassuringly, this effect had subsided in the spillover region, and only reached marginal significance at the pronoun at the start of the critical sentence.

**Table 6** Summary of mixed effects model results by region in Experiment 2a. P-values indicating significant effects are in bold and shaded grey. Effects of adjective bias are only reported in text.



**Figure 6** Region-by-region reading times in the Critical Sentence, on log scale, in Experiment 2b.

<sup>15</sup> Supplemental analyses as well as raw (anonymized) data for all experiments, including the norming studies, can be found under <https://osf.io/59qzj/>.

Region	Descriptive complexity	Pronoun	RT	SE
Subject NP <i>The hikers</i>	Complex	<i>it</i>	733	7
		<i>that</i>	724	8
	Simple	<i>it</i>	722	9
		<i>that</i>	716	8
Verb <i>entered/explored</i>	Complex	<i>it</i>	353	3
		<i>that</i>	356	3
	Simple	<i>it</i>	353	4
		<i>that</i>	343	3
Object NP <i>the forest</i>	Complex	<i>it</i>	858	6
		<i>that</i>	867	6
	Simple	<i>it</i>	859	7
		<i>that</i>	834	8
Spillover I <i>at</i>	Complex	<i>it</i>	335	2
		<i>that</i>	336	3
	Simple	<i>it</i>	337	2
		<i>that</i>	331	3
Spillover II <i>night.</i>	Complex	<i>it</i>	349	4
		<i>that</i>	348	4
	Simple	<i>it</i>	353	5
		<i>that</i>	350	3
Pronoun <i>It/That</i>	Complex	<i>it</i>	352	3
		<i>that</i>	353	3
	Simple	<i>it</i>	356	3
		<i>that</i>	363	3
Copula <i>was</i>	Complex	<i>it</i>	304	2
		<i>that</i>	317	3
	Simple	<i>it</i>	306	2
		<i>that</i>	318	2
Adverb <i>very</i>	Complex	<i>it</i>	308	2
		<i>that</i>	314	2
	Simple	<i>it</i>	313	2
		<i>that</i>	313	2
Adjective <i>adventurous</i> ( <i>or an object-selecting adjective</i> )	Complex	<i>it</i>	462/438	8/7
		<i>that</i>	451/455	10/7
	Simple	<i>it</i>	449/454	7/7
		<i>that</i>	453/467	9/11

**Table 7** Mean region-by-region reading times and SEM for Experiment 2b. In the last region, mean and SEM was separately reported for each adjective bias condition (event bias/object bias).

In the **auxiliary region** (“was”) and the **adverbial region** (“very”) of the critical sentence, we find a main effect of pronoun type, but no effect of complexity, and no interaction, in line with our other experiments. This time, we also find the predicted interaction of adjective bias with pronoun type at the **final adjective** (marginal,  $\beta = .03$ ,  $p < .07$ ), a main effect of

**Table 8** Summary of mixed effects model results by region in Experiment 2b. P-values indicating significant effects are in bold; darker grey shading indicates significance at 5% level, light grey, marginal significance. Effects of adjective bias are only reported in the text.

	The hunters			Entered/Explored			The forest			AT			Night							
	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P	$\beta$	SE	T	P				
Complexity	0.00	0.01	-0.26	0.80	-0.01	0.01	-1.88	<b>0.06</b>	-0.02	0.01	-3.76	<b>0.00</b>	0.00	0.01	-0.77	0.44	0.00	0.01	0.37	0.71
Pronoun type	-0.01	0.01	-2.01	<b>0.04</b>	0.00	0.01	-1.09	0.28	-0.01	0.01	-1.57	0.12	-0.01	0.01	-1.25	0.21	0.00	0.01	-0.08	0.94
Comp. × Pro. type	-0.01	0.01	-0.52	0.60	-0.03	0.01	-2.11	<b>0.03</b>	-0.02	0.01	-1.87	<b>0.06</b>	-0.03	0.01	-2.69	<b>0.01</b>	-0.02	0.01	-1.43	0.15
<b>It/That</b>																				
<b>Was</b>																				
<b>Very</b>																				
<b>Adventurous</b>																				
<b>Very</b>																				
<b>Adventurous</b>																				
Complexity	0.01	0.01	1.63	<b>0.10</b>	0.01	0.00	1.37	0.17	0.01	0.00	1.32	0.19	0.00	0.01	0.25	0.80				
Pronoun type	0.00	0.01	0.47	0.64	0.04	0.00	7.56	<b>0.00</b>	0.02	0.00	3.78	<b>0.00</b>	0.00	0.01	0.74	0.46				
Comp. × Pro. type	0.00	0.01	-0.40	0.69	-0.01	0.01	-0.61	0.54	-0.01	0.01	-1.05	0.30	-0.01	0.01	-0.63	0.53				

complexity ( $\beta = .03, p < .05$ ), and a three-way interaction between pronoun type, complexity, and adjective bias ( $\beta = .07, p < .02$ ).

## 7. Discussion of Experiments 2a and 2b

Experiments 2a and 2b had three main aims: (i) to see whether the main effect of pronoun type that we observed in the first two experiments are replicable in events of differing complexity, (ii) to test whether adjective bias interacts with pronouns, and (iii) to test our prediction that the bare demonstrative “that” – but not the pronoun “it” – is sensitive to higher-level conceptual features, operationalized in this study as the conceptual complexity of an event. These predictions were derived from our model of demonstratives as ‘universal bundlers’ that take chunks of conceptual structure that are not referred to (in our stimuli) with noun phrases. In order to do this bundling, demonstratives must access the conceptual, not only surface, features of the referent. To test this claim, we manipulated the conceptual complexity of the events that subsequent demonstratives referred back to.

Specifically, we predicted that more complex events (as denoted by verbs) would lead to faster reading times for sentences containing “that”, compared to sentences containing “it”. We did not find this predicted interaction, but we did replicate the main effect of pronoun type predicted and found in Experiments 1a and b already: People read slower after a demonstrative than after a personal pronoun, throughout several regions. We discuss these findings in more depth in the general discussion.

We also found the predicted mismatch effect, in the form of an interaction of adjective bias with pronoun type or complexity or both, at the sentence-final adjective in Experiment 2b. We can only speculate as to why this effect only surfaced in one of four experiments: One reason may be that at the sentence-final word, complex wrap-up effects can mask other effects (e.g., Warren et al., 2009).

## 8. General Discussion

This paper set out to test our hypothesis that the English bare demonstrative pronoun “that” tends to refer to bundles of chunks of conceptual structure, whereas the non-personal pronoun “it” can simply refer to a noun phrase that satisfies its classificatory and relational components. We argued that because “that” accesses and bundles up complex conceptual structures, it should induce longer reading times, and perhaps be uniquely sensitive to higher-level conceptual features, such as the complexity of a concept. Thus, we predicted that (i) both “that” and “it” would be sensitive to surface features such as frequency and word length, whereas (ii) only “that” would be sensitive to event complexity, and (iii) “that” would be read more slowly than “it” overall. These predictions were partially supported by two sets of self-paced reading studies:

### 8.1 Sensitivity to surface properties

In the first pair of self-paced reading studies (Experiment 1a and 1b), the data show that both non-personal and demonstrative pronouns were read faster after nouns that were shorter in length and higher in frequency than after nouns that were longer and less frequent. Importantly, these results also surface when there is a delay between the referent and the pronoun, such that the reduction in reading times cannot merely be taken as a spillover effect from the nouns themselves, but rather can be attributed to the reference resolution process itself, supporting our model of demonstratives as bundlers of conceptual structure as opposed to simple linguistic anaphora devices. In addition to this, these findings provided the baseline for Experiments 2a and 2b, which asked whether “that”, and not “it”, is uniquely sensitive to higher-level conceptual features of its referent.

### 8.2 Sensitivity to conceptual properties

The second part of our hypothesis that the bare demonstrative “that” tends to bundle up complex conceptual structures, whereas “it” does not necessarily do so, results in two predictions: First, “that” should result in longer reading times throughout compared to “it”, and “that” should be uniquely sensitive to referential complexity.

The first prediction was confirmed in all four experiments: The demonstrative “that” resulted in longer reading times than “it”, and this main effect was robust, stable, and replicable. This pattern is very likely not due to orthographic differences, since we also analyzed the data residualized over the length of the pronouns, and still found this effect. On the contrary: Since both forms (non-personal pronouns and demonstratives) are on the very high end of the lexical frequency spectrum, finding any effect on the pronoun region is striking, given that due to the word frequency and the word length effect, reaction times to short, high-frequency words tend to stick to floor level (see e.g., Morton, 1970; Kuperman, Drieghe, Keuleers & Brysbaert, 2013; Dirix, Brysbaert & Duyck, 2019, and others for data from many behavioral paradigms). Thus, these results are a powerful demonstration of the different reading behaviors induced by non-personal pronouns and demonstratives.

To test the second prediction, we conducted Experiments 2a and 2b, which manipulated potential referents’ conceptual complexity. Specifically, we manipulated event complexity: We compared sentences with more complex events to sentences with less complex events. We expected the demonstrative “that” to pattern differently from “it” as follows: Based on prior research, and our own previous data, we expected an interaction such “that” would be read faster in conditions with more complex events than in conditions with simpler events. This is because “that” tends to refer to the event, and complex events have been shown to be more easily retrieved the more complex they are (Hofmeister, 2011; see below).<sup>16</sup> While this second prediction was not borne out, the first was, showing that demonstratives are processed differently from personal pronouns, affecting reading times overall.

### 8.3 Our results in light of previous literature

Let us briefly compare our results to other experiments on antecedent frequency effects on pronoun processing, because our results may, at first blush, seem surprising given prior work: Several recent studies have reported that less frequent antecedents lead to *faster* reading times at the pronoun (van Gompel & Majid, 2004), or to *no reliable differences* at the pronoun (Egusquiza et al., 2016; Lago, 2014). However, each of these studies is crucially different from ours in several aspects. First and foremost, van Gompel & Majid (2004) as well as Lago (2014) used animate antecedents (“the arsonist/the criminal”), and likely more importantly, unambiguous possessive determiners (“his bag”) to investigate effects of frequency. Second, they measured effects in eyetracking-while-reading paradigms. And third, they used frequency as a proxy for saliency, on the assumption that all low-frequency items are more salient than high-frequency items.

In contrast, we used a standard masked self-paced reading paradigm, and measured reading times at a bare (non-human-referring) pronoun which was, in principle, compatible with at least two different referent types: the sentential object alone, such as “forest” or “jungle”, or the whole event (“The hikers entered/explored the forest/the jungle.”) We also used items that were close synonyms, such as “forest” and “jungle”, and not as conceptually far apart as many of van Gompel & Majid’s (2004) as well as Lago’s (2014) stimuli (pairs in these papers were, for instance, “student” vs. “vagrant”, or “doctor” vs. “envoy”, which may have entailed not only a differential in frequency, but also in register, pragmatics, and semantic associations and features.) In light of these differences, direct comparisons between experiments are presumably not meaningful.<sup>17</sup>

In contrast, Hofmeister (2011) explicitly manipulated conceptual complexity. He found that people were faster reading “banned” in sentences like (17b) than in sentences like (17a); that is, the more complex NP “alleged Venezuelan communist” in (17b) was easier to retrieve from

---

<sup>16</sup> In previous studies using the same materials, we have found effects exhibiting a pattern as described in the prediction above, but we cannot replicate it with the current data. One reason for this could be that these interaction effects are small, and self-paced reading data is too coarse-grained to detect them (Wittenberg, 2013). We hope to investigate the question further using different techniques.

<sup>17</sup> This, in conjunction with the fact that our frequency manipulation was subtler than in other studies, could explain why we did not find any reading time differences at the antecedent nouns themselves. In a paradigm like ours, which does not allow for backward or forward glances to the context, it is also not surprising to find stable frequency antecedent effects at the time of reading non-personal and demonstrative pronouns referring back to those antecedents. As we argue above, these are very likely to involve different reference resolution mechanisms than possessive determiners (like in “his bag”). As Lago (2014) observed, and we discussed above, “not all types of antecedent lexical information need to be reaccessed during coreference” (Lago, 2014: 110).



working memory and integrate in the argument structure than the less complex NP ‘communist’ in (17a):

- (17) a. It was a communist who the members of the club banned from ever entering the premises.  
b. It was an alleged Venezuelan communist who the members of the club banned from ever entering the premises.

In Hofmeister’s (2011) stimuli, ‘alleged Venezuelan communist’ is indeed semantically richer than ‘communist’, but it is also a longer, more complex noun phrase (and retrieval effects could potentially be due to longer encoding time, Karimi, Diaz & Wittenberg, 2020); whereas in our study, the only variation between the verbs was one of explicitly controlled conceptual complexity. Hofmeister’s Experiment 2, which replaced semantically specific referents like “soldier” with non-specific referents like “person” and is in some sense similar to ours, found only marginally significant retrieval effects.

Even without an effect of conceptual complexity, however, our data can be taken as evidence that “that” and “it” result in significantly different processing. We propose a model of how a bundling process may be triggered by demonstratives like “this” or “that”, and we view this proposal as an extension and unification of approaches to demonstratives that have described their functions in discourse from the perspective of information structure (Ariel 2001, Gundel et al., 1993; Strauss, 2002), or from the perspective of anaphora (Çokal, Sturt, & Ferreira, 2016).

We must stress that the form-specific preferences and mechanisms that we propose here are not deterministic: In English at least, both “it” and “that” can refer to both objects and events. Furthermore, both “it” and “that” can be temporarily ambiguous: “it” could be an expletive (“It is raining”), and “that” could be a demonstrative determiner (“that macaron”; although see Strauss, 2002, for corpus data on how “that” is more than twice as often used on its own as a bare demonstrative pronoun, as opposed to as a modifier to an NP). It seems unlikely that these temporary ambiguities could explain away the results we observed, but we do think that our studies should be extended, possibly into other languages.

Languages other than English also often show a contrast between simple pronouns and demonstrative pronouns, as extensive fieldwork has shown (e.g. Diessel, 1999; Dixon, 2003; Givón, 1978; Himmelmann, 1996); and it must be assumed that they slice the conceptual pie between object and event reference in interestingly different ways than English does (for instance, see Bosch et al., 2003; Grosz, 2018; Kaiser, 2011; and many others for German *d*-pronouns). To take an example that has caught the attention of psycholinguistic research, Kaiser & Trueswell (2008) investigated the processing of simple and demonstrative personal pronouns in Finnish. Their data indicate that these types of pronouns exhibit different form-specific constraints: Whereas the personal pronouns are sensitive to grammatical role, the demonstratives were sensitive to both information structure and grammatical role. Little is known so far about event or object reference in Finnish, however. It will be interesting to see whether all languages have elements that can function as bundlers, and if so, whether demonstratives (or their equivalent) cross-linguistically show this tendency to be able to refer to conceptual structures wider than simple pronouns.<sup>18</sup>

Other open questions concern the limits of bundling. If it is true that bundling is a conceptual process, it should operate under the same working memory constraints as operations on other cognitive units (Baddeley, 2012). In recent years, there have been many promising attempts at integrating research on working memory and language processing. Lewis & Vasishth (2005), a prominent example, proposed a content-addressable memory architecture that is integrated within linguistic theory. Our model of the role of the bare demonstrative “that” as a universal bundler fits squarely within this theory, while also accounting for the non-linguistic, conceptual content that demonstratives bundle up for use by linguistic means.

In sum, in this paper we presented data from two English self-paced reading studies, each replicated once, showing that demonstratives are processed differently from personal pronouns, and that this affects reading patterns throughout the whole sentence. These data can be seen as initial evidence supporting a new, unified model of reference by demonstratives as a process

---

<sup>18</sup> Thanks to an anonymous reviewer for pointing out some of these open questions.

## Acknowledgements

We thank Ray Jackendoff, Cathal O'Madagain, Jeremy Skipper, Joshua Wampler, the audience of AMLaP 2017, as well as three anonymous reviewers for helpful discussion and suggestions.

## Funding information


This research was supported by a UC San Diego Social Sciences Divisional Research Grant to E.W.

## Competing interests

The authors have no competing interests to declare.

## Author affiliations

**Eva Wittenberg**  [orcid.org/0000-0002-3188-6145](https://orcid.org/0000-0002-3188-6145)  
University of California, San Diego, US

**Shota Momma**  [orcid.org/0000-0002-2567-4517](https://orcid.org/0000-0002-2567-4517)  
University of Massachusetts Amherst, US

**Elsi Kaiser**  [orcid.org/0000-0003-3594-4127](https://orcid.org/0000-0003-3594-4127)  
University of Southern California, US

## References

- Abbott, Barbara. 2004. Definiteness and indefiniteness. In Laurence R. Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*, 122–149. Malden, MA: Blackwell. DOI: <https://doi.org/10.1002/9780470756959.ch6>
- Ariel, Mira. 2001. Accessibility theory: An overview. In Ted J. M. Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects* 8. 29–87. DOI: <https://doi.org/10.1075/hcp.8.04ari>
- Arnold, Jennifer E. 1998. *Reference form and discourse patterns*. Stanford, CA: Stanford University dissertation.
- Baayen, Rolf Harald, Doug J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, Rolf Harald, Richard Piepenbrock & Léon Gulikers. 1995. *CELEX2 LDC96L14*. Web Download. Philadelphia: Linguistic Data Consortium.
- Baddeley, Alan. 2012. Working memory: Theories, models, and controversies. *Annual Review of Psychology* 63. 1–29. DOI: <https://doi.org/10.1146/annurev-psych-120710-100422>
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, Douglas, Martin Maechler, Ben Bolker, Steve Walker, R. H. B. Christensen, H. Singmann, et al. 2014. *Package 'lme4'*. Vienna: R foundation for statistical computing. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bosch, Peter, Tom Rozario & Yufan Zhao. 2003. Demonstrative pronouns and personal pronouns: German der vs. er. In *Proceedings of the EACL 2003. Workshop on the Computational Treatment of Anaphora*, 61–68. Budapest: Association for Computational Linguistics.
- Brennan, Susan E., Marilyn W. Friedman & Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th annual meeting of the Association for Computational Linguistics*, 155–162. Stanford: Association for Computational Linguistics. DOI: <https://doi.org/10.3115/981175.981197>
- Brown-Schmidt, Sarah. 2009. Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language* 61(2). 171–190. DOI: <https://doi.org/10.1016/j.jml.2009.04.003>
- Brown-Schmidt, Sarah, Donna K. Byron & Michael K. Tanenhaus. 2005. Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language* 53(2). 292–313. DOI: <https://doi.org/10.1016/j.jml.2005.03.003>

- Bunt, Harry C. 1985. The formal representation of (quasi-)continuous concepts. In Jerry R. Hobbs & Robert C. Moore (eds.), *Formal theories of the commonsense world*, 37–70. Norwood, NJ: Ablex.
- Cacciari, Cristina, Manuel Carreiras & Cristina Barbolini Cionini. 1997. When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language* 37(4). 517–532. DOI: <https://doi.org/10.1006/jmla.1997.2528>
- Casati, Roberto & Achille C. Varzi. 2008. Event concepts. In Thomas F. Shipley & Jeffrey M. Zacks (eds.), *Understanding Events: From Perception to Action*, 31–53. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780195188370.003.0002>
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Charles N. Li (ed.), *Subject & topic*. New York, NY: Academic Press.
- Chow, Wing-Yee, Shevaun Lewis & Colin Phillips. 2014. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology* 5. DOI: <https://doi.org/10.3389/fpsyg.2014.00630>
- Çokal, Derya, Patrick Sturt & Fernanda Ferreira. 2014. Deixis: This and that in written narrative discourse. *Discourse Processes* 51(3). 201–229. DOI: <https://doi.org/10.1080/0163853X.2013.866484>
- Çokal, Derya, Patrick Sturt & Fernanda Ferreira. 2016. Processing of It and This in Written Narrative Discourse. *Discourse Processes*, 1–18. DOI: <https://doi.org/10.1080/0163853X.2016.1236231>
- Cornish, Francis. 2008. How indexicals function in texts: Discourse, text, and one neo-Gricean account of indexical reference. *Journal of Pragmatics* 40(6). 997–1018. DOI: <https://doi.org/10.1016/j.pragma.2008.02.006>
- Craik, Fergus I. & Endel Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104(3). 268. DOI: <https://doi.org/10.1037/0096-3445.104.3.268>
- Culicover, Peter W. & Ray Jackendoff. 2012. Same-except: A domain-general cognitive relation and how language expresses it. *Language* 88(2). 305–340. DOI: <https://doi.org/10.1353/lan.2012.0031>
- Cunnings, Ian, Clare Patterson & Claudia Felser. 2014. Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language* 71(1). 39–56. DOI: <https://doi.org/10.1016/j.jml.2013.10.001>
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins Publishing. DOI: <https://doi.org/10.1075/tsl.42>
- Dirix, Nicolas, Marc Brysbaert & Wouter Duyck. 2019. How well do word recognition measures correlate?: Effects of language context and repeated presentations. *Behavior Research Methods* 51(6). DOI: <https://doi.org/10.3758/s13428-018-1158-9>
- Dixon, Robert M. W. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language* 27(1). 61–112. DOI: <https://doi.org/10.1075/sl.27.1.04dix>
- Drummond, Alex. 2014. Ibx Experimental Platform. <https://github.com/addrummond/ibx>.
- Duffy, Susan A. & Keith Rayner. 1990. Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech* 33(2). 103–119. DOI: <https://doi.org/10.1177/002383099003300201>
- Eckert, Miriam & Michael Strube. 2000. Dialogue acts, synchronizing units, and anaphora resolution. *Journal of Semantics* 17(1). 51–89. DOI: <https://doi.org/10.1093/jos/17.1.51>
- Egusquiza, Nerea, Eduardo Navarrete & Adam Zawiszewski. 2016. Antecedent frequency effects on anaphoric pronoun resolution: Evidence from Spanish. *Journal of Psycholinguistic Research* 45(1). 71–84. DOI: <https://doi.org/10.1007/s10936-014-9325-3>
- Elbourne, Paul. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31(4). 409–466. DOI: <https://doi.org/10.1007/s10988-008-9043-0>
- Evans, Richard. 2001. Applying machine learning toward an automatic classification of it. *Literary and Linguistic Computing* 16(1). 45–58. DOI: <https://doi.org/10.1093/lc/16.1.45>
- Fine, Alex B., T. Florian Jaeger, Thomas A. Farmer & Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS one* 8(10). e77661. DOI: <https://doi.org/10.1371/journal.pone.0077661>
- Fisher, Ronald P. & Fergus I. M. Craik. 1980. The effects of elaboration on recognition memory. *Memory & Cognition* 8(5). 400–404. DOI: <https://doi.org/10.3758/BF03211136>
- Gallo, David A., Nathaniel G. Meadow, Elizabeth L. Johnson & Katherine T. Foster. 2008. Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language* 58. 1095–1111. DOI: <https://doi.org/10.1016/j.jml.2007.12.001>
- Givón, Talmy. 1978. Definiteness and referentiality. *Universals of Human Language* 4. 291–330.
- Gordon, Peter C., Barbara J. Grosz & Laura A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17(3). 311–347. DOI: [https://doi.org/10.1207/s15516709cog1703\\_1](https://doi.org/10.1207/s15516709cog1703_1)
- Gordon, Peter C. & Kimberly A. Scarse. 1995. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition* 23(3). 313–323. DOI: <https://doi.org/10.3758/BF03197233>

- Greene, Steven B., Richard J. Gerrig, Gail McKoon & Roger Ratcliff. 1994. Unheralded pronouns and management by common ground. *Journal of Memory and Language* 33(4). 511–526. DOI: <https://doi.org/10.1006/jmla.1994.1024>
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204.
- Grosz, Barbara J., Scott Weinstein & Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225. DOI: <https://doi.org/10.21236/ADA324949>
- Grosz, Patrick Georg. 2018. Bridging uses of demonstrative pronouns in German. *Linguistics and Philosophy* 41. 367–421. DOI: <https://doi.org/10.1007/s10988-017-9226-7>
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 274–307. DOI: <https://doi.org/10.2307/416535>
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 2005. Pronouns without NP antecedents: How do we know when a pronoun is referential. *Anaphora processing: linguistic, cognitive and computational modelling*, 351–364. DOI: <https://doi.org/10.1075/cilt.263.20gun>
- Halliday, Michael A. K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold Ltd.
- Hankamer, Jorge & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3). 391–428.
- Hartshorne, Joshua K. & Jesse Snedeker. 2013. Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes* 28(10). 1474–1508. DOI: <https://doi.org/10.1080/01690965.2012.689305>
- Heine, Angela, Sascha Tamm, Markus Hofmann, Florian Hutzler & Arthur M. Jacobs. 2006a. Does the frequency of the antecedent noun affect the resolution of pronominal anaphors? An ERP study. *Neuroscience Letters* 400. 7–12. DOI: <https://doi.org/10.1016/j.neulet.2006.02.006>
- Heine, Angela, Sascha Tamm, Markus Hofmann, Rainer M. Bösel & Arthur M. Jacobs. 2006b. Event-related theta activity reflects memory processes in pronoun resolution. *Cognitive Neuroscience and Neuropsychology* 17(18). 1835–1839. DOI: <https://doi.org/10.1097/WNR.0b013e328010a096>
- Himmelman, Nikolaus P. 1996. Demonstratives in narrative discourse: a taxonomy of universal uses. In Fox ed., *Studies in Anaphora*, 205–254. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.33.08him>
- Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes* 26(3). 376–405. DOI: <https://doi.org/10.1080/01690965.2010.492642>
- Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780198270126.001.0001>
- Järvikivi, Juhani, Roger P. G. van Gompel & Jukka Hyönä. 2017. The Interplay of Implicit Causality, Structural Heuristics, and Anaphor Type in Ambiguous Pronoun Resolution. *Journal of Psycholinguistic Research* 46(3). 525–550. DOI: <https://doi.org/10.1007/s10936-016-9451-1>
- Kaiser, Elsi. 2013. Looking beyond personal pronouns and beyond English: Typological and computational complexity in reference resolution. *Theoretical Linguistics* 39(1–2). 109–122. DOI: <https://doi.org/10.1515/tl-2013-0007>
- Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman & Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112(1). 55–80. DOI: <https://doi.org/10.1016/j.cognition.2009.03.010>
- Kaiser, Elsi & John C. Trueswell. 2008. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* 23(5). 709–748. DOI: <https://doi.org/10.1080/01690960701771220>
- Kaiser, Elsi & John C. Trueswell. 2011. Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond saliency. In Edward A. Gibson & Neal J. Perlmutter (eds.), *The Processing and Acquisition of Reference*, 323–353. Cambridge, MA: The MIT Press. DOI: <https://doi.org/10.7551/mitpress/9780262015127.003.0013>
- Kaplan, David. 1989. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. *Themes from Kaplan*, 481–563.
- Karimi, Hossein & Fernanda Ferreira. 2016. Informativity renders a referent more accessible: Evidence from eyetracking. *Psychonomic Bulletin & Review* 23(2). 507–525. DOI: <https://doi.org/10.3758/s13423-015-0917-1>
- Karimi, Hossein, Michele Diaz & Eva Wittenberg. 2020. Sheer time spent expecting or maintaining a representation facilitates subsequent retrieval during sentence processing. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, 2728–2734. Seattle: Cognitive Science Society.
- Kehler, Andrew & Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39(1–2). 1–37. DOI: <https://doi.org/10.1515/tl-2013-0001>
- Kliegl, Reinhold, Ellen Grabner, Martin Rolfs & Ralf Engbert. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16(1–2). 262–284. DOI: <https://doi.org/10.1080/09541440340000213>

- Koornneef, Arnout W. & Ted J. M. Sanders. 2013. Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes* 28(8). 1169–1206. DOI: <https://doi.org/10.1080/01690965.2012.699076>
- Kruisinga, Etsko. 1925–32. *A Handbook of Present Day English*, 4th ed. Utrecht: Kemink. DOI: <https://doi.org/10.1007/BF01521692>
- Kuperman, Victor, Denis Drieghe, Emmanuel Keuleers & Marc Brysbaert. 2013. How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology* 66. 563–580. DOI: <https://doi.org/10.1080/17470218.2012.658820>
- Lago, M. Sol. 2014. *Memory and Prediction in Cross-Linguistic Sentence Comprehension*. College Park, MD: University of Maryland dissertation.
- Lewis, Russell L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. DOI: [https://doi.org/10.1207/s15516709cog0000\\_25](https://doi.org/10.1207/s15516709cog0000_25)
- Loáiciga, Sharid, Luca Bevacqua, Hannah Rohde & Christian Hardmeier. 2018, June. Event versus entity co-reference: Effects of context and form of referring expression. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, 97–103. New Orleans: Association of Computational Linguistics. DOI: <https://doi.org/10.18653/v1/W18-0711>
- Loar, Brian. 1976. The semantics of singular terms. *Philosophical Studies* 30(6). 353–377. DOI: <https://doi.org/10.1007/BF00372537>
- Mason, Winter & Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44(1). 1–23. DOI: <https://doi.org/10.3758/s13428-011-0124-6>
- Meng, Michael & Markus Bader. 2020. Does comprehension (sometimes) go wrong for noncanonical sentences?. *Quarterly Journal of Experimental Psychology* 78(1). 1–28. DOI: <https://doi.org/10.1177/1747021820947940>
- Mook, Douglas G. 1983. In defense of external invalidity. *American Psychologist* 38(4). 379–387. DOI: <https://doi.org/10.1037//0003-066X.38.4.379>
- Morton, John. 1970. A functional model for memory. In *Models of human memory*, 203–254. New York, NY: Academic Press. DOI: <https://doi.org/10.1016/B978-0-12-521350-9.50012-7>
- Müller, Christoph. 2007. Resolving it, this, and that in unrestricted multi-party dialog. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 816–823. Prague: Association of Computational Linguistics.
- Munro, Robert, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen & Harry Tily. 2010, June. Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk*, 122–130. Los Angeles: Association for Computational Linguistics.
- Nappa, Rebecca & Jennifer E. Arnold. 2014. The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology* 70. 58–81. DOI: <https://doi.org/10.1016/j.cogpsych.2013.12.003>
- Nunberg, Geoffrey. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16(1). 1–43. DOI: <https://doi.org/10.1007/BF00984721>
- O'Madagain, Cathal. 2020. This is a Paper about Demonstratives. *Philosophia*. DOI: <https://doi.org/10.1007/s11406-020-00230-5>
- Omaki, Akira, Ellen F. Lau, Imogen Davidson White, Myles L. Dakan, Aaron Apple & Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6. 384. DOI: <https://doi.org/10.3389/fpsyg.2015.00384>
- Pelletier, Francis Jeffrey. 1975. Non-singular reference: some preliminaries. In *Mass terms: Some philosophical problems*, 1–14. Dordrecht: Springer. DOI: [https://doi.org/10.1007/978-1-4020-4110-5\\_1](https://doi.org/10.1007/978-1-4020-4110-5_1)
- Pelletier, Francis Jeffrey & Lenhart K. Schubert. 1989. Mass expressions. In *Handbook of Philosophical Logic*, 327–407. Dordrecht: Springer. DOI: [https://doi.org/10.1007/978-94-009-1171-0\\_4](https://doi.org/10.1007/978-94-009-1171-0_4)
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York, NY: Longman.
- Radvansky, Gabriel A. & Jeffrey M. Zacks. 2014. *Event Cognition*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199898138.001.0001>
- Reuter, Tracy & Casey Lew-Williams. 2018. Look at that: Deixis reveals developmental changes in verbal prediction. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. London: Cognitive Science Society.
- Rohde, Hannah & Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29(8). 912–927. DOI: <https://doi.org/10.1080/01690965.2013.854918>
- Scott, Kate. 2013. This and that: A procedural analysis. *Lingua* 131. 49–65. DOI: <https://doi.org/10.1016/j.lingua.2013.03.008>

- Simner, Julia & Ron Smyth. 1999. Phonological activation in anaphoric lexical access (ALA). *Brain and Language* 68(1). 40–45. DOI: <https://doi.org/10.1006/brln.1999.2112>
- Smit, J. P. 2012. Why bare demonstratives need not semantically refer. *Canadian Journal of Philosophy* 42(1). 43–66. DOI: <https://doi.org/10.1353/cjp.2012.0000>
- Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43(1). 155–167. DOI: <https://doi.org/10.3758/s13428-010-0039-7>
- Stewart, Andrew J., Judith Holler & Evan Kidd. 2007. Shallow processing of ambiguous pronouns: Evidence for delay. *The Quarterly Journal of Experimental Psychology* 60(12). 1680–1696. DOI: <https://doi.org/10.1080/17470210601160807>
- Strauss, Susan. 2002. This, that, and it in spoken American English: a demonstrative system of gradient focus. *Language Sciences* 24(2). 131–152. DOI: [https://doi.org/10.1016/S0388-0001\(01\)00012-2](https://doi.org/10.1016/S0388-0001(01)00012-2)
- Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344. DOI: <https://doi.org/10.1093/mind/LIX.235.320>
- van Gompel, Roger P. G. & Asifa Majid. 2004. Antecedent frequency effects during the processing of pronouns. *Cognition* 90(3). 255–264. DOI: [https://doi.org/10.1016/S0010-0277\(03\)00161-6](https://doi.org/10.1016/S0010-0277(03)00161-6)
- Warren, Tessa, Sarah J. White & Erik D. Reichle. 2009. Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition* 111(1). 132–137. DOI: <https://doi.org/10.1016/j.cognition.2008.12.011>
- Webber, Bonnie L. 1988, June. Discourse deixis: Reference to discourse segments. In *Proceedings of the 26th annual meeting of the Association for Computational Linguistics*, 113–122. Buffalo, NY: Association for Computational Linguistics. DOI: <https://doi.org/10.3115/982023.982037>
- Webber, Bonnie, Matthew Stone, Aravind Joshi & Alistair Knott. 2003. Anaphora and discourse structure. *Computational Linguistics* 29(4). 545–587. DOI: <https://doi.org/10.1162/089120103322753347>
- Wettstein, Howard K. 1984. How to bridge the gap between meaning and reference. *Synthese* 58(1). 63–84. DOI: <https://doi.org/10.1007/BF00485362>
- Wiese, Heike & Joan Maling. 2005. Beers, kaffi, and Schnaps: Different grammatical options for restaurant talk coercions in three Germanic languages. *Journal of Germanic Linguistics* 17(1). 1–38. DOI: <https://doi.org/10.1017/S1470542705000012>
- Wittenberg, Eva & Roger Levy. 2017. If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language* 94. 254–271. DOI: <https://doi.org/10.1016/j.jml.2016.12.001>
- Zwaan, Rolf A., Mark C. Langston & Arthur C. Graesser. 1995. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science* 6(5). 292–297. DOI: <https://doi.org/10.1111/j.1467-9280.1995.tb00513.x>

Wittenberg et al.  
*Glossa: a journal of  
 general linguistics*  
 DOI: 10.5334/gjgl.917

TO CITE THIS ARTICLE:

Wittenberg, Eva, Shota Momma and Elsi Kaiser. 2021. Demonstratives as bundlers of conceptual structure. *Glossa: a journal of general linguistics* 6(1): 33. 1–30. DOI: <https://doi.org/10.5334/gjgl.917>

Submitted: 12 February 2019

Accepted: 20 February 2021

Published: 05 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.